Neuroepigenomics in Alzheimer's disease: The single cell ADds

Renzo J.M. Riemens

Lay-Out: Renzo J.M. Riemens and Ipskamp Maastricht
Cover Design: Renzo J.M. Riemens
Published by: Maastricht University
ISBN: 978-94-6423-524-1
Printed By: Ipskamp Maastricht

Neuroepigenomics in Alzheimer's disease: The single cell ADds

Neuroepigenomik bei der Alzheimer-Krankheit: Die Einzelzell ADds

DISSERTATION

To obtain the degree of
Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans

and

to obtain the degree of Doctor of Philosophy in Natural Sciences
Dr. rer. Nat. at Julius-Maximilians-Universität Würzburg,
on the authority of the President Magnificus,
Prof. dr. Alfred Forchel,
in accordance with the decision of the Graduate School of Life Sciences
Common Graduation Commision,

to be defended in public on Friday 5 November 2021, at 10.00 hours.

by

Renzo J.M. Riemens

This thesis was written under a jointly supervised PhD degree programme of Maastricht University, Maastricht, the Netherlands and Julius-Maximilians-Universität Würzburg, Würzburg, Germany.

Maastricht University
Promotors
Prof. Dr. Daniël L.A. van den Hove
Prof. Dr. Bart P.F. Rutten

Co-Promotor
Dr. Gunter Kenis

Julius-Maximilians-Universität Würzburg
Promotor
Prof. Dr. Thomas Haaf *(Primary supervisor)*

Co-promotors
Prof. Dr. Charlotte Förster *(Second supervisor)*
Prof. Dr. Klaus Peter-Lesch *(Third supervisor)*
Prof. Dr. Daniël L.A. van den Hove *(Fourth supervisor)*

Graduate School of Life Sciences (GSLS) chairperson
Prof. Dr. Christian Wegener

Assessment Committee
Prof. Dr. Jos H.H.J. Prickaerts (Maastricht University, the Netherlands) (Chair)
Dr. Angelika Schmitt-Böhrer (Julius Maximilian University, Germany)
Prof. Dr. Charlotte Förster (Julius Maximilian University, Germany)
Prof. Dr. Geraldine Zimmer-Bensch (Aachen University, Germany)
Prof. Dr. Paul J. Lucassen (University of Amsterdam, the Netherlands)
Dr. Tim Vanmierlo (Maastricht University, the Netherlands)

Table of contents

Chapter 1

General introduction

Alzheimer's disease

Even though the first records of Alzheimer's disease (AD) date back to millennia ago, our knowledge of its underlying pathophysiology is little more than a century old [1]. The German psychiatrist and neuropathologist Alois Alzheimer published his now famous case study of Auguste Deter in 1906 [2]. Since then, we have witnessed an outburst of basic and translational research into the underlying molecular mechanisms, risk factors and potential treatment strategies for AD. Research into AD began with profiling of the disease-associated neuropsychological deficits. Further efforts lead to *e.g.* the identification of cognitive mechanisms affected by the neuropathological hallmarks observed in AD brains, the discovery of genetic, as well as environmental risk factors contributing to the disease, and the development of transgenic animal models harboring AD hallmarks.

Despite the considerable advances in more than a century, gaps of knowledge on the etiopathophysiology of AD continue to exist. In fact, AD nowadays ranks among the most challenging health care problems worldwide, as it remains incurable and impossible to detect in a timely manner that could improve treatment outcome. It has been estimated that AD and related dementias currently affect around 50 million people worldwide, a figure that is projected to double every 20 years to 100 million cases in 2040 [3]. This exponential increase, both in developed and developing countries, possesses a huge social burden on both patients, their direct family members, caregivers and society as a whole. The progressive nature of AD, leading to severe functional deterioration and a need for constant care, is one of the major determinants of this burden. Consequently, the magnitude of this disease has a devastating societal and financial impact on caregivers and the healthcare system resources. Unless an effective treatment will be developed, the medical, as well as non-medical costs, are estimated to surpass one trillion euro by 2050 [3]. As such, the need for a better understanding, more advanced AD models, better diagnostic tools and more effective therapeutics today is more urgent than ever, given the anticipated, alarming increase in incidence in our ever-aging population.

As it is currently understood, AD is as a chronic, progressive neurodegenerative disease associated with memory loss and cognitive decline [4]. Other behavioral and psychological symptoms of AD, which might vary with time and per individual, include agitation, apathy, disturbances in motor behavior, deficiencies in language, elation, irritability, disinhibition, delusions, hallucinations, changes in sleep or appetite, and anxiety and depression [5]. The main pathological hallmarks observed in AD brains are depositions of extracellular amyloid beta (Aβ) plaques and intracellular neurofibrillary tangles consisting of

hyperphosphorylated tau protein [6]. Substantial evidence supports that these depositions are central to the neurobiology of the disease, as they are causally related to the neurodegenerative processes leading to cortical and subcortical atrophy typically observed in AD brains [7]. Other molecular hallmarks of AD include oxidative stress, dysfunctional calcium homeostasis and neuroinflammation, ultimately contributing to the loss of neurons and synapses [8-10]. This loss in neuronal integrity is known to contribute to a disequilibrium in neurotransmitter systems, affecting most, if eventually not all, systems found in the human brain, including glutamate, acetylcholine, dopamine, noradrenaline and serotonin [11]. It has been proposed that a region-specific disbalance in these neurotransmitter systems underlies the behavioral changes observed in AD patients.

It is worth mentioning that aberrant Aβ and tau proteins already start accumulating in the brain long before the manifestation of the first clinical symptoms [12]. The pathological process spans decades and, during this time, the distribution pattern of neurofibrillary tangles appears to develop according to a hierarchical temporospatial pattern [13]. Although Aβ burden has been associated with the neurobiological basis of the disease, only tau pathology has been shown to be directly correlated with progression of cognitive deterioration and clinical symptoms in AD. Braak and Braak therefore introduced a staging system based on aberrant tau, *i.e.* Braak staging, which describes the initial appearance of tauopathy in the transentorhinal regions (stages I and II) that spreads to limbic regions (stages III and IV) and finally to neocortical areas (stages V and VI) [13]. However, more recent studies suggested that tauopathy may begin earlier than previously thought and possibly in the lower brainstem rather than in the transentorhinal regions [14, 15]. Furthermore, the occurrence of various non-cognitive behavioral and neuropsychological symptoms in AD suggests brainstem involvement, specifically that of the dorsal raphe nuclei (DRN) and the locus coeruleus (LC). Even though one cannot ignore the later involvement of entorhinal, limbic and cortical neurodegeneration in relation to the cognitive symptoms, critical questions on the temporal and spatial causality did arise in this respect during the last years. Research on the pathogenesis of AD is therefore currently expanding to areas such as the DRN and the LC.

An individual person's risk of developing AD pathology is determined by genetic, environmental and demographic risk factors, as well as the complex interactions between them. From a genetic point of view, AD can be classified into two forms of dementia: familial AD (fAD) and sporadic AD (sAD). Although both types appear identical in light of clinical features and neuropathological hallmarks, they are distinguishable based on to their etiology and time of onset. Generally, fAD

is referred to as the early onset form of AD, occurring between 30 and 60 years of age, although later development in life has also been observed in patients. The most critical difference when compared to sAD, is that fAD is inherited in a mendelian fashion, resulting from autosomal-dominant mutations in genes involved in Aβ processing that lead to increased neurotoxic Aβ aggregation, including amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*) [16]. Notably, it is estimated that fAD patients only represent less than 5% of all AD cases. sAD, representing the most common variant of AD, is estimated to affect more than 95% of all cases, while having a time of onset starting at around the age of 65 years, with a doubling likelihood of developing the disease every 5 years thereafter [17]. sAD has a less obvious and more diffuse, polygenic, genetic underpinning when compared to fAD and is known to be multifactorial with a life-long etiopathogenesis. However, since theories on contributing factors and the preliminary appearance of pathological hallmarks are constantly changing, a thorough understanding of its etiology remains largely elusive, and a definite, unifying hypothesis on the exact cause-effect relationships of these hallmarks, despite decades of research, remains to be formulated.

Over the last years, large genome wide-association studies (GWAS) have identified genetic risk, as well as protective, loci for the development of sAD, including *ABCA7*, *BIN1*, *TREM2*, *CD33*, *CLU*, *CRI*, *EPHA1*, *MS4A*, *PICALM*, *and SORL1*, among others [18-24]. These genes have been associated with Aβ and tau processing, lipid metabolism, immune activation and synaptic function, which all have been linked to AD [25]. Nonetheless, the risk effects exerted by these individual genes are small, that is they confer only a ~0.10- to 0.15-fold increase in carriers versus non-carriers [16]. As such, the impact of genetic variation around these loci remains significantly lower than that of *APOE*, which for many years was the only major gene known to increase sAD risk, but on itself not sufficient enough to cause the disease [26]. Of note in this respect is the fact that many individuals may carry prominent genetic risk factors for sAD, as well as express profound Aβ and tau pathology, yet never develop the disease [27]. Indeed, even monozygotic twins can have indefinite sAD outcomes [28]. Thus, there is currently no single genetic model that explains the mode of disease transmission and it has been estimated that only half of the phenotypic variance in sAD is attributed to solely genetic factors [29]. Altogether, it is now hypothesized that both yet unidentified rare genetic variants with large effect sizes, interactions between different loci, as well as non-genetic factors, contribute to the “missing heritability” of sAD, which emphasizes its complexity.

Various conditions and environmental factors have been identified that either positively or negatively influence the development of sAD. While ageing is known

to be the major risk factor, pre-existing conditions such as obesity, diabetes, hypertension and a previous diagnosis of depression have been shown to increase the risk of developing the disease [30]. Additionally, exposure to heavy metals, smoking and alcoholism, are also considered as risk factors for sAD development [31, 32]. Physical exercise, cognitive reserve, high levels of education and various nutritional factors, on the other hand, have been depicted as protective factors [30, 31, 33]. Studies have furthermore demonstrated that affective experiences may also play a role in the pathogenesis of the disease. While prolonged and/or repetitive exposure to stress has been associated with a faster progression of sAD, mental well-being or having a general sense of purpose in life is associated with lower AD pathology [34, 35]. As already addressed earlier, it is known that in AD patients psychopathology occurs (long) before the cognitive symptoms appear, which suggests that adverse psychological experiences could potentially also cause or exacerbate AD-related neuropathology. In fact, affective symptomatology could also represent an early sign, *i.e.* a prodromal phase, of the disease. All in all, since environmental factors alone also fail to explain sAD development, it is currently becoming more clear that the etiology of the disease likely dependents on the synergistic effects of both genetic and environmental factors [36].

In this latter context, epigenetic processes, which mediate the interaction between the genome and the environment, could provide a mechanistic explanation in view of the etiology of sAD [37]. In addition to an individual's genetic susceptibility, environmental insults could disturb epigenetic mechanisms of gene expression with adverse consequences, concomitant with developing pathology as a result. These epigenetic changes, which can occur throughout life, could evolve into profound AD pathology or interact with the embedded effects of other contributing factors to both induce and influence disease development and progression. The model explained here advocates for a combination of genetics, environmental risk factors and epigenetic mechanisms, where these factors may operate synergistically during the pre-clinical phase of the disease, even decades before the appearance of the first clinical symptoms. In recent years, we have therefore seen a significant increase in studies focusing on the role of epigenetic mechanisms in the development and course of sAD.

Epigenetic mechanisms: DNA methylation and hydroxymethylation

The term epigenetics, initially formulated by Conrad Hal Waddington in 1942, implies "on top of" or "in addition to" genetics, referring to changes in gene expression that cannot be explained by changes in the DNA sequence [38]. The meaning of the term epigenetics, however, has itself undergone an evolution that resembles our dramatically deepened understanding of the mechanisms that

underlie gene regulation [39]. Our present definition has moved from being solely regarded as a regulatory process of gene silencing and activation during development, *e.g.* in view of cellular differentiation, to a crucial and highly dynamic plethora of processes involved in translating exposure to environmental stimuli to changes in gene expression patterns. As such, a unique characteristic that differentiates epigenetic from genetic variation is that epigenetic processes are responsive to the environment in a more dynamic fashion. Accordingly, their manifestation can have deleterious effects, *e.g.* in response to adverse environmental stimuli, a process which nowadays is thought to contribute to sAD [40]. These changes may be reflected at various stages throughout an individual's life, and even in subsequent generations, through a process termed transgenerational epigenetic inheritance [41, 42].

From a molecular point of view, epigenetic mechanisms consist of DNA modifications, histone modifications and noncoding RNAs, each offering a distinct – though often interdependent – layer of transcriptional control [43]. Collectively, these mechanisms determine the chromatin architecture, the concomitant accessibility of a genetic locus to the transcriptional machinery, while they are often also involved in the regulation of post-transcriptional processes [44-46]. Recent work argues that particularly covalent modifications to the DNA and histones are modifiable throughout the lifespan and that these can be altered by a person's experiences [47]. Alterations in the chemistry of these modifications can bring about chromatin remodeling processes, changing the accessibility of the DNA to the transcriptional machinery, and, thereby, gene expression patterns. Thus, environmental variation possesses means of changing these modifications, although some are more programmed and persistent than others, as seen during *e.g.* stem cell differentiation [48]. As such, these DNA modifications can be highly stable, as observed in genomic imprinting, or highly dynamic, such as epigenetic changes associated with memory formation [49, 50]. Importantly, whereas genetic alterations usually reflect permanent changes of the DNA sequence, epigenetic changes generally reflect reversible processes, still, with a pivotal role in regulating specific brain functions and related behaviors. Of note, the exact mechanisms by which the environment induces its effects on the epigenetic machinery remains largely unknown.

To date, the most extensively characterized epigenetic DNA modification is methylation, which involves the addition of a methyl group at the 5 position on the pyrimidine ring of a cytosine, *i.e.* 5-methylcytosine (5-mC) [44, 51]. DNA methylation occurs mainly at cytosine-phosphate-guanine (CpG) positions, although non-CpG cytosine, guanine and adenine methylation have also been reported [43]. CpG sites are primarily packed in so-called CpG islands, which are

stretches longer than 200 base pairs that have over 50% of CpG content [52]. These islands are often part of regulatory elements, including gene promoters. Historically, DNA methylation in promotor regions has been associated with transcriptional repression, although recent findings have also demonstrated that gene expression can be enhanced in a methylation-dependent manner [43, 53, 54]. Overall, the promotor methylation status directly influences the possible binding of transcription factors, which interact with histone-modifying enzymes that regulate chromatin accessibility, hence modulating transcriptional activity [55, 56]. In addition, DNA methylation within the transcribed portion of a gene is also observed and has been implicated in *e.g.* alternative splicing [57], although it remains to be further elucidated how gene expression is exactly regulated through gene body methylation. Aside from methylation, several other DNA modifications have also been described, including 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC). The derivation of these other modifications generally occurs through the process of active DNA demethylation, which implies the removal of the methyl group from cytosine bases [58]. This process is achieved by the oxidation of the methyl group by ten-eleven translocation (TET) enzymes, leading to the sequential production of each of these modifications, respectively.

Although initially considered solely as an intermediate modification during the process of active demethylation, DNA hydroxymethylation, has also been shown to exist as a stable epigenetic mark, similar to DNA methylation. Interestingly, it has been demonstrated that 5-hmC is highly enriched in the brain, where, in parallel to 5-mC, it is thought to fulfill an important role during neurodevelopment and memory formation, among others [59, 60]. Furthermore, studies have shown that 5-hmC is also important in pluripotent stem cell processes, such as proliferation and neural differentiation [61]. At the DNA level, 5-hmC can be found both at CpG sites and at non-CpG dinucleotides [62]. Genome-wide localization experiments have demonstrated that 5-hmC can be found in and around promoter regions, enhancers, transcription start sites and in intragenic regions [63, 64]. In addition, it has been suggested that changes in the 5-hmC architecture are involved in chromatin remodeling processes, thereby influencing gene expression patterns [65, 66]. Interestingly, DNA hydroxymethylation is both associated with transcriptional expression and repression [66, 67]. In conclusion, the studies presented above underline the importance of both 5-mC and 5-hmC in transcriptional regulation during healthy brain functioning. Given the importance of these modifications in the human brain, it is reasonable to assume that dysregulation of these marks could contribute to various brain disorders, including sAD.

DNA methylation and hydroxymethylation in sporadic Alzheimer's disease

Over the last decade, the number of publications on the role of epigenetic dysregulation in sAD has increased significantly. It is now well-accepted that environmental factors affect the epigenome and that epigenetic changes might provide a mechanistic link between adverse environmental exposure and the development and course of sAD. The rapidly growing, but still juvenile field of neuroepigenetics is therefore starting to provide novel insights into the underlying pathophysiological mechanisms of this disease. The increasing interest in epigenetic dysregulation is furthermore attributed to the clinical potential of epigenetic marks, as they could be used for the development of biomarker assays, potentially allowing early detection of sAD, as well as providing opportunities for more detailed disease classification. Moreover, the reversible nature of epigenetic modifications makes them suitable candidates for preventive treatment strategies and clinical interventions. For these reasons, the characterization of epigenomic profiles in brain tissue samples, as well as in blood, in both patients and people at risk of developing the disease is currently offering an appealing approach to study sAD.

When looking at the neuroepigenomic studies that have been conducted in sAD so far, one can observe that, over the last years, the field has moved away from the use of (antibody-based) technologies targeting global epigenetic marks [66, 68-73] towards more in-depth epigenome-wide arrays and sequencing-based platforms that provide mechanistic insight with a much broader coverage and higher resolution [43, 74]. Although a potential role for DNA (hydroxy)methylation in the pathogenesis of sAD became evident from these initial antibody-based studies, their findings have remained inconsistent and inconclusive – most likely attributable to the aspecificity and lack of reliability of the methodology used. Thus, the imperative need for more concise and conclusive data has shifted the field to the use of more novel, robust techniques, including the aforementioned platforms. The continuous advancements of the Illumina microarray, combined with its cost-effectiveness, have made this approach the most appealing method in sample cohort studies to date. Especially a combination of the microarray with independent validation using pyrosequencing is developing itself into a golden standard for DNA (hydroxy)methylation studies in sAD. The dawn of these approaches in combination with advanced bioinformatics analysis have led to the current era of epigenome-wide association studies (EWAS). Although these EWAS will be discussed in more detail in Chapter 2, it is worth mentioning that initial studies targeting different brain regions and blood samples derived from sAD patients have already identified AD-specific DNA (hydroxy)methylation changes in various genomic loci, emphasizing their importance in the disease [74, 75].

In spite of the fact that previous work has indicated a critical role for DNA (hydroxy)methylation changes in the development and course of sAD, many questions related to these marks remain yet unanswered. For instance, it is still unclear what exactly the functional consequences of these alterations are, how they specifically contribute to developing pathology, and whether these marks indeed represent valuable biomarkers. More research is necessary in order to obtain a better understanding of their role in sAD, to develop effective therapeutics and diagnostic alternatives, and thus to eventually be able to alleviate the clinical and socioeconomic burden of the disease.

Given the notion that aberrant DNA (hydroxy)methylation patterns could be involved in both the progression, as well as the etiology, of the disease, it will furthermore be necessary to assess whether these marks are dysregulated in regions of the brain affected early in the pathogenesis of AD. Interestingly, such studies have not been performed up until now, as most of the investigations outlined above targeted epigenetic dysregulation in cortical brain regions, or the hippocampus. Even though these areas of the brain are affected by the disease, as mentioned earlier, it is becoming more apparent that pathological changes in these regions particularly occur in later and more advanced stages of sAD. To this end, targeting epigenetic dysregulation in brainstem regions including the DRN and the LC might provide a more suitable strategy for identifying alterations involved in the early pathogenesis of sAD.

Nevertheless, in addition to simply profiling epigenetic marks in these brain regions, the field also requires further advances that could overcome remaining (methodological) challenges, thereby contributing to the production of more conclusive data. On the one hand, there are challenges directly attributed to the complex neurobiology of the disease [80]. First, DNA (hydroxy)methylation changes occur with aging, with different patterns observed depending on the timing, the tissue assessed and the genomic location analyzed [81, 82]. This emphasizes the importance of balancing age across experimental groups and confirms the need of discriminating epigenetic marks associated with the disease from those that occur during normal physiological aging. Second, disease comorbidities in sAD patients are common and could hinder the identification of sAD-specific epigenomic profiles [76]. It is therefore necessary to carefully select experimental groups based on the presence or absence of other pathology. Third, there is the problem of tissue and cell-type specificity of epigenetic modifications. sAD is characterized by a temporospatial pattern of changes, affecting different regions of the brain as the disease progresses. Brain regions therefore need to be selected carefully, as they could be affected differentially across the different disease stages. Of a similar nature, is the issue of cell-type specificity. sAD is

characterized by neuronal loss, altering the cell-type composition of diseased samples, which complicates the interpretation of generated data. In fact, the use of heterogeneous tissues could obscure cell-type specific modifications related to sAD, as changes in one cell type could negate or distort changes in another [77]. On the other hand, there are various methodological challenges that one should consider. Although some might be overcome based on adapting study designs, it will also require technological advances or modifications of existing protocols in order to allow more sophisticated and reliable analyses. Two examples of these include the aspecificity of the epigenetic marks assessed and causal interference. Techniques involving the use of microarrays and pyrosequencing rely on a DNA bisulfite treatment [78, 79]. However, it has been demonstrated that this treatment does not allow for the discrimination between 5-mC and 5-hmC [80]. In fact, a bisulfite methylation analysis without prior oxidation yields cumulative measures of both modifications [80]. Future studies will therefore need to adapt current protocols in order to produce more informative data. In addition, it is important to realize that EWAS in blood or brain tissue does not allow discrimination between cause and consequence of epigenetic changes in AD, as certain epigenetic marks could also result from AD-associated neuropathology and/or its neurobiological and behavioral consequences. Technological advances such as induced pluripotent stem cell (iPSC) models and epigenetic editing techniques, however, are anticipated to aid in overcoming this issue, allowing one to study the effect of epigenetic changes on cellular and molecular responses *in vitro* [81]. Thus, hopes are placed on such interdisciplinary approaches that will allow for a better understanding on the exact role of epigenetics in sAD in the future.

Advanced models of Alzheimer's disease: induced pluripotent stem cells

In 2006, a major scientific breakthrough was made with the discovery that mouse fibroblasts could be reprogrammed towards iPSCs by viral delivery of four transcription factors, i.e. *OCT4*, *SOX2*, *KLF4* and *cMYC* [82]. These cells demonstrated an expression profile and developmental potential similar to embryonic stem cells (ESCs), including their properties of self-renewal and their ability to differentiate into the three embryonic germ layers. One year later, it was demonstrated that also human fibroblasts could be reprogrammed, allowing one to produce personalized iPSCs from theoretically every individual [83, 84]. These studies were shortly followed by reports demonstrating the successful generation of iPSCs from various other human somatic cells and by alternative methods, such as blood cells and non-integrative, non-viral based techniques, respectively [85, 86]. After the discovery of the technology, a large proportion of research has moved directly into the establishment of protocols for the differentiation of iPSCs into neural populations. The possibility of creating disease-relevant neural cells

derived from patients' iPSCs received substantial interest for the generation of human disease models. In fact, many brain disorders that previously lacked a comprehensive cellular model, including sAD, could now potentially be modeled using patient-derived iPSCs. Consequently, the scientific community has put a lot of effort in developing robust and scalable methods for generating large numbers of differentiated neural cells from iPSCs.

Over the last years, protocols for differentiating iPSCs towards neural precursor cells (NPCs) and various neuronal subtypes, as well as astrocytes, microglia, and oligodendrocytes, became available [87, 88]. These protocols generally involve the chemical stimulation of iPSCs through patterning cues or the use of ectopic overexpression of lineage-specific transcription factors, which both play a central role during neural development in the early embryo [87]. In addition to the networks of patterning molecules and transcription factors that orchestrate neural differentiation, it is currently well-accepted that epigenetic mechanisms are also key players within these circuits [89]. Recent advances in epigenetic editing and RNA interference therefore offer novel, alternative strategies for neural differentiation *in vitro* [81, 87]. Aside from directed differentiation of iPSCs, which constrains the differentiation of these cells towards a specific neural subtype, protocols involving the direct reprogramming of somatic cells towards neural cells or NPCs also became available [87]. Even though these methods have demonstrated that functional neurons can be differentiated from iPSCs or fibroblast, the optimization and refinement of these approaches, as well as the establishment of novel methods to produce neural subtypes for which no protocols have been established yet, is currently an ongoing process [87]. Nevertheless, it is anticipated that further efforts in this field will play a significant role in the development of disease-relevant cultures for the establishment of cellular models of AD.

The opportunity of developing iPSC-based models based on the approach described above has ushered a new exciting era of scientific research into sAD. Cellular models consisting of neurons and glia derived from patients' iPSCs are hypothesized to harbor and manifest disease hallmarks *in vitro* [90]. As such, these models offer unprecedented opportunities to study these hallmarks in more detail, potentially contributing to the development of a better understanding on the underlying molecular mechanisms of the disease. Furthermore, the availability of patient-derived neural populations provides a personalized platform for high-throughput drugs and toxicity screenings that could aid in the development of novel therapeutics or diagnostic tools [91, 92]. Additionally, the opportunity of creating co-cultures or three-dimensional models incorporating multiple brain cell types offers a new approach towards modeling the complex

cellular interactions in integrated neural networks found in the human brain [90, 93]. These procedures are becoming increasingly popular nowadays, given the general advantages that iPSCs have in disease modeling compared with traditional animal and cellular models. Next to their human origin, these advantages include their easy accessibility, expandability, the ability to give rise to almost all major brain cells, relatively less ethical concerns when compared to work with human ESCs or animal models, and their potential for personalized medicine [94].

Besides these general advantages, iPSCs also have the potential to overcome challenges specifically related to sAD research. To date, AD studies have mainly relied on transgenic mouse models, immortalized cell lines and post-mortem brain tissues. Although various animal models carrying fAD mutations have displayed certain neuropathological and behavioral deficits, they are mainly modeling fAD and, furthermore, do not develop all phenotypic aspects characteristic of human AD, such as neuronal death with distinct neurofibrillary tangle formation [95]. In addition, the effort of generating mice carrying the human *APOE* gene has not led to a model that truly mimics sAD development [96]. Post-mortem brain tissues and immortalized cell lines on the other hand, have offered opportunities to study sAD in human cells. However, these tissues are limited for studying the earlier stages of the disease and do not allow one to study the cause-effect relationships between molecular hallmarks and developing pathology [97]. Although immortalized cell lines provide an unlimited resource for AD research, these models do not harbor the unique genomes of sAD patients. The characteristics of iPSCs, however, provide a unique platform to detect early-disease phenotypes, which may point towards underlying pathogenic mechanisms of sAD later in life [98]. In addition, since human iPSCs can be derived from individuals with sAD, they contain a patient-specific pathogenic background [99]. Furthermore, since these cells are cultivated in laboratory conditions, it allows for direct manipulations of various molecular mechanisms, as well as their microenvironment, allowing more mechanistic studies on cause-effect relationships.

In order to achieve this potential, neurons from patients' iPSCs should naturally manifest pathological mechanisms *in vitro*. Indeed, pioneering studies using iPSCs from fAD patients have found the presence of disease hallmarks in differentiated neural cells, including altered APP processing, elevated levels of neurotoxic Aβ peptides, increased levels of hyperphosphorylated tau protein and activation of GSK3β, a physiological kinase of tau [99, 100]. As such, these model systems have proven to be instrumental in modelling molecular alterations associated with genetic mutations in AD-causing genes, such as *APP* and

PSEN1. Although sAD is more prevalent, only a limited number of iPSC-based models have been developed for this form of the disease [99, 100]. In very few cases, iPSC-derived neurons from sAD patients were studied in parallel to lines obtained from fAD patients, with the main goal of seeking to find AD-associated cellular phenotypes for the validation of their potential for sAD modeling. Interestingly, these initial sAD models have shown a high degree of variability and inconsistencies in terms of disease hallmarks, which might be explained by the heterogeneous nature of sAD as observed in patients. On the other hand, given that these studies represent initial attempts of developing disease-relevant model systems, various methodological factors are likely contributing to these lacking hallmarks [99]. Further improving these models to facilitate more meaningful studies for sAD will therefore be a major goal in the upcoming years.

While it is well-established that epigenetic alterations contribute to sAD, detailed characterization of epigenomic profiles in iPSC-derived neural cells has not been performed yet. The complex and interconnected network of epigenetic changes, including DNA (hydroxy)methylation, has previously only been studied in animal models and in human postmortem brain samples. In future studies, it will therefore be vital to profile *e.g.* the methylome and hydroxymethylome and their downstream effects on the transcriptome in neural cells differentiated from patients' iPSCs. Furthermore, it will be crucial to compare these profiles with the same omics profiles in identical cells found in their brains in order to further validate the model for epigenomic and transcriptomic research. With the advent of iPSC technology, it can be expected that patient-derived neural models represent valid tools to explore epigenetic changes involved in sAD. Based on the availability of epigenetic editing [81] nowadays, iPSC-derived neuronal populations generated from sAD patients can be epigenetically probed and exposed to adverse environmental factors, as well as treated with drug libraries, to investigate their effects on molecular and cellular responses. Moreover, epigenetic-editing might even represent an interesting (proof-of-concept) therapeutic intervention on itself that could be explored further using these iPSC-based models. Along similar lines, probing the genomes of sAD patients and studying their interaction with AD-associated environmental factors could potentially contribute to a better understanding of complex gene-environment interactions, as well as the involvement of epigenetic mechanisms. To conclude, although the field of iPSC technology is still in its early development, it is anticipated that patient-derived neural populations differentiated from iPSCs could bridge the gap between animal models and clinical studies, which will ultimately contribute to a better understanding of the disease that could aid in establishing therapeutic interventions.

Overview of the thesis

The research that is compiled in this thesis can be divided in two parts. The first part, consisting of four chapters, is centered around the role of epigenetic dysregulation in the etiopathophysiology of sAD. In addition to providing insights into the most recent developments in neuroepigenomic studies of this disease, the first part of the thesis also touches upon remaining challenges, and provides a future outlook on possible developments in the field. The second part, which includes three more chapters, is focused on the application of iPSC-based disease models for the study of AD, including but not limited to mechanistic studies on epigenetic dysregulation using this platform. Aside from outlining the research that has been conducted using iPSC-based models for sAD to date, the second part of the thesis also provides insights into the acquisition of disease-relevant neural cultures based on directed differentiation of iPSCs, and furthermore includes an experimental approach for the establishment of such a model system.

The general introduction only briefly described epigenetic dysregulation, and specifically alterations in DNA methylation and hydroxymethylation, in relation to sAD. **Chapter 2** therefore offers a more in-depth perspective on EWAS in sAD. Among others, the degree of dysregulation in both DNA methylation and hydroxymethylation in various brain tissues and blood samples derived from patients are summarized. Furthermore, relevant caveats in relation to these studies that could affect correct interpretation of the obtained results are discussed and future outlooks on possible solutions, as well as developments, to overcome these challenges are provided.

Following on that notion, **Chapter 3** represents a methodological research article offering a possible solution to the issue of modification specificity in neuroepigenomic studies of sAD. In more detail, an approach for the targeted detection of unmodified cytosine, 5-mC and 5-hmC levels at single CpG sites by oxidative bisulfite pyrosequencing is provided. For this purpose, three spike-in pyrosequencing controls that can be added to a given DNA target sample were developed in order to assess the successful application of the described method. By applying these spike-in controls, as well as by analyzing both brain tissue and iPSC samples, the efficacy of the oxidative bisulfite pyrosequencing approach for the detection of the aforementioned cytosine states in the context of single CpG sites is demonstrated.

Chapter 4 introduces a possible solution to overcome the issue of cellular heterogeneity in neuroepigenomic studies of sAD. In this chapter, it is demonstrated, for the first time, that limiting dilution bisulfite pyrosequencing

(LDBSP) can be successfully applied on pools of individually isolated post-mortem neurons using laser capture microdissection (LCM). The general workflow of LDBSP for the detection of DNA bisulfite methylation in multiple target genes on single alleles derived from these neurons is provided. Furthermore, a novel approach for data correction when applying this technique is being described, aimed at reducing bias and allowing the user to more accurately estimate the DNA methylation status of target genes using this method.

In **Chapter 5**, the body of EWAS reviewed in previous chapters, is complemented with the most recent findings on epigenetic dysregulation in the brainstem in AD. More specifically, this chapter provides data on the role of DNA methylation and hydroxymethylation in the DRN and LC of sAD patients. Common and distinct epigenetic signatures in both brainstem nuclei were identified and validated by a subsequent bisulfite pyrosequencing analysis. The EWAS presented here is furthermore complemented by a highly innovative cell subtype-specific analysis in the DRN, which was based on the LCM-LDBSP approach described in the previous chapter.

Chapter 6 moves away from profiling epigenetic patterns in post-mortem tissues or cells and introduces the use of iPSC-based model systems in sAD research. This chapter reviews iPSC-based studies that have been conducted in sAD to date. In addition, important sources of variability related to such model systems, next to those that might be explained by the heterogeneous nature of sAD, are being discussed.

In the development of iPSC-based models for sAD, **Chapter 7** summarizes the progress that has been made in generating various neuronal subtypes from both stem cells and somatic cells. This review dissects the directed- and direct differentiation protocols for the generation of disease-relevant neural cultures with special emphasis on chemically defined systems, transcription factor-mediated reprogramming and epigenetic-based approaches. Furthermore, efforts that are being made to increase the efficiency of current protocols are discussed and the potential for the use of these cells in disease modelling, drug discovery and regenerative medicine are being addressed.

The final study presented in **Chapter 8** comprises an exploratory approach, in which the establishment and characterization of a human neuronal *in vitro* model system for AD using iPSCs is being described. Based on the step-wise application of neural patterning factors, iPSCs were differentiated towards NPCs and, subsequently, cortical neurons, which represents a cellular population known to be affected in the pathophysiology of AD. In order to further explore the

potential of this *in vitro* model for future AD studies, the expression of AD-associated proteins were assessed in the differentiated neuronal cells.

All in all, the research collected in this thesis covers a multitude of studies in which state-of-the-art techniques, brain tissue samples and model systems are being applied in order to study the role of epigenetic dysregulation in the etiopathogenesis of (s)AD.

References

1. Vatanabe IP, Manzine PR, Cominetti MR. Historic concepts of dementia and Alzheimer's disease: From ancient times to the present. Rev Neurol (Paris). 2019.
2. Maurer K, Volk S, Gerbaldo H. Auguste D and Alzheimer's disease. The Lancet. 1997;349(9064):1546-9.
3. Prince MJ. World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends: Alzheimer's Disease International; 2015.
4. Silva MVF, Loures CdMG, Alves LCV, de Souza LC, Borges KBG, das Graças Carvalho M. Alzheimer's disease: risk factors and potentially protective measures. Journal of biomedical science. 2019;26(1):33.
5. Cerejeira J, Lagarto L, Mukaetova-Ladinska E. Behavioral and psychological symptoms of dementia. Frontiers in neurology. 2012;3:73.
6. Huang H-C, Jiang Z-F. Accumulated amyloid-β peptide and hyperphosphorylated tau protein: relationship and links in Alzheimer's disease. Journal of Alzheimer's disease. 2009;16(1):15-27.
7. Scheltens P, Blennow K, Breteler MM, de Strooper B, Frisoni GB, Salloway S, et al. Alzheimer's disease. Lancet. 2016;388(10043):505-17.
8. Mohandas E, Rajmohan V, Raghunath B. Neurobiology of Alzheimer's disease. Indian journal of psychiatry. 2009;51(1):55.
9. Tönnies E, Trushina E. Oxidative stress, synaptic dysfunction, and Alzheimer's disease. Journal of Alzheimer's Disease. 2017;57(4):1105-21.
10. Van Leeuwen FW, de Kleijn DP, van den Hurk HH, Neubauer A, Sonnemans MA, Sluijs JA, et al. Frameshift mutants of β amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. Science. 1998;279(5348):242-7.
11. Kumar K, Kumar A, Keegan RM, Deshmukh R. Recent advances in the neurobiology and neuropharmacology of Alzheimer's disease. Biomedicine & pharmacotherapy. 2018;98:297-307.
12. Beason-Held LL, Goh JO, An Y, Kraut MA, O'Brien RJ, Ferrucci L, et al. Changes in brain function occur years before the onset of cognitive impairment. Journal of Neuroscience. 2013;33(46):18008-14.
13. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta neuropathologica. 1991;82(4):239-59.
14. Grinberg L, Rüb U, Ferretti R, Nitrini R, Farfel J, Polichiso L, et al. The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset? Neuropathology and applied neurobiology. 2009;35(4):406-16.
15. Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. Journal of Neuropathology & Experimental Neurology. 2011;70(11):960-9.
16. Tanzi RE. The genetics of Alzheimer disease. Cold Spring Harbor perspectives in medicine. 2012;2(10):a006296.

17. Stozicka Z, Zilka N, Novak M. Risk and protective factors for sporadic Alzheimer's disease. Acta virologica. 2007;51(4):205-22.
18. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nature genetics. 2011;43(5):429.
19. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nature genetics. 2011;43(5):436-41.
20. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. New England Journal of Medicine. 2013;368(2):117-27.
21. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009;41(10):1094-9.
22. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nature genetics. 2009;41(10):1088.
23. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. Jama. 2010;303(18):1832-40.
24. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013;45(12):1452.
25. Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. The Lancet Neurology. 2016;15(8):857-68.
26. Yamazaki Y, Zhao N, Caulfield TR, Liu C-C, Bu G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. Nature Reviews Neurology. 2019;15(9):501-18.
27. Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nature Reviews Neurology. 2013;9(2):106.
28. Creasey H, Jorm A, Longley W, Broe G, Henderson A. Monozygotic twins discordant for Alzheimer's disease. Neurology. 1989;39(11):1474-.
29. Ridge PG, Mukherjee S, Crane PK, Kauwe JS, Consortium AsDG. Alzheimer's disease: analyzing the missing heritability. PloS one. 2013;8(11).
30. Edwards III GA, Gamez N, Escobedo Jr G, Calderon O, Moreno-Gonzalez I. Modifiable risk factors for Alzheimer's disease. Frontiers in aging neuroscience. 2019;11.
31. Killin LO, Starr JM, Shiue IJ, Russ TC. Environmental risk factors for dementia: a systematic review. BMC geriatrics. 2016;16(1):1-28.
32. Bisht K, Sharma K, Tremblay M-È. Chronic stress as a risk factor for Alzheimer's disease: Roles of microglia-mediated synaptic remodeling, inflammation, and oxidative stress. Neurobiology of stress. 2018;9:9-21.
33. Sharp ES, Gatz M. The relationship between education and dementia an updated systematic review. Alzheimer disease and associated disorders. 2011;25(4):289.
34. Boyle PA, Buchman AS, Wilson RS, Yu L, Schneider JA, Bennett DA. Effect of purpose in life on the relation between Alzheimer disease pathologic changes on cognitive function in advanced age. Archives of general psychiatry. 2012;69(5):499-504.
35. Chen ST, Siddarth P, Saito NY, Rueda F, Haight T, Ercoli LM, et al. Psychological well-being and regional brain amyloid and tau in mild cognitive impairment. The American Journal of Geriatric Psychiatry. 2014;22(4):362-9.
36. Lesuis SL, Hoeijmakers L, Korosi A, de Rooij SR, Swaab DF, Kessels HW, et al. Vulnerability and resilience to Alzheimer's disease: early life conditions modulate neuropathology and determine cognitive reserve. Alzheimer's research & therapy. 2018;10(1):1-20.

37. Lahiri DK, Maloney B, Zawia NH. The LEARn model: an epigenetic explanation for idiopathic neurobiological diseases. Molecular psychiatry. 2009;14(11):992-1003.
38. Choudhuri S. From Waddington's epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research. Toxicology mechanisms and methods. 2011;21(4):252-74.
39. Felsenfeld G. A brief history of epigenetics. Cold Spring Harbor perspectives in biology. 2014;6(1):a018200.
40. Liu X, Jiao B, Shen L. The epigenetics of Alzheimer's disease: Factors and therapeutic implications. Frontiers in genetics. 2018;9:579.
41. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. Cell. 2014;157(1):95-109.
42. Guerrero-Bosagna C, Skinner MK. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. Molecular and cellular endocrinology. 2012;354(1-2):3-8.
43. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131:21-64.
44. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences. 2006;31(2):89-97.
45. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Nature reviews genetics. 2015;16(7):421-33.
46. Mattick JS. The central role of RNA in human development and cognition. FEBS letters. 2011;585(11):1600-16.
47. Roth TL, David Sweatt J. Annual research review: epigenetic mechanisms and environmental shaping of the brain during sensitive periods of development. Journal of Child Psychology and Psychiatry. 2011;52(4):398-408.
48. Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. Nature Reviews Genetics. 2017;18(11):643-58.
49. Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nature Reviews Genetics. 2001;2(1):21-32.
50. Woldemichael BT, Bohacek J, Gapp K, Mansuy IM. Epigenetics of memory and plasticity. Progress in molecular biology and translational science. 122: Elsevier; 2014. p. 305-40.
51. Kumar S, Chinnusamy V, Mohapatra T. Epigenetics of modified DNA bases: 5-methylcytosine and beyond. Frontiers in genetics. 2018;9:640.
52. Cross SH, Bird AP. CpG islands and genes. Current opinion in genetics & development. 1995;5(3):309-14.
53. Silva PNO, Gigek CO, Leal MF, Bertolucci PHF, de Labio RW, Payao SLM, et al. Promoter methylation analysis of SIRT3, SMARCA5, HTERT and CDH1 genes in aging and Alzheimer's disease. Journal of Alzheimer's Disease. 2008;13(2):173-6.
54. Guilleret I, Benhattar J. Unusual distribution of DNA methylation within the hTERT CpG island in tissues and cell lines. Biochemical and biophysical research communications. 2004;325(3):1037-43.
55. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356(6337):eaaj2239.
56. Portela A, Esteller M. Epigenetic modifications and human disease. Nature biotechnology. 2010;28(10):1057.
57. Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ, Amdam GV. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC genomics. 2012;13(1):480.
58. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-3.

59. Sun W, Zang L, Shu Q, Li X. From development to diseases: the role of 5-hmC in brain. Genomics. 2014;104(5):347-51.
60. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PloS one. 2010;5(12).
61. Cheng Y, Xie N, Jin P, Wang T. DNA methylation and hydroxymethylation in stem cells. Cell biochemistry and function. 2015;33(4):161-73.
62. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011;473(7347):398-402.
63. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473(7347):343-8.
64. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473(7347):394-7.
65. Valinluck V, Tsai H-H, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic acids research. 2004;32(14):4100-8.
66. Coppieters N, Dragunow M. Epigenetics in Alzheimer's disease: a focus on DNA modifications. Current pharmaceutical design. 2011;17(31):3398-412.
67. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & development. 2011;25(7):679-84.
68. Bihaqi S, Schumacher A, Maloney B, Lahiri D, Zawia N. Do epigenetic pathways initiate late onset Alzheimer disease: towards a new paradigm. 2012.
69. Mastroeni D, McKee A, Grover A, Rogers J, Coleman PD. Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. PloS one. 2009;4(8).
70. Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. Neurobiology of aging. 2010;31(12):2025-37.
71. Chouliaras L, Mastroeni D, Delvaux E, Grover A, Kenis G, Hof PR, et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiology of aging. 2013;34(9):2091-9.
72. Bradley-Whitman M, Lovell M. Epigenetic changes in the progression of Alzheimer's disease. Mechanisms of ageing and development. 2013;134(10):486-95.
73. Coppieters N, Dieriks BV, Lill C, Faull RL, Curtis MA, Dragunow M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiology of aging. 2014;35(6):1334-44.
74. Roubroeks JAY, Smith RG, van den Hove DLA, Lunnon K. Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. J Neurochem. 2017;143(2):158-70.
75. Smith R, Pishva E, Shireby G, Smith AR, Roubroeks JA, Hannon E, et al. Meta-analysis of epigenome-wide association studies in Alzheime's disease highlights 220 differentially methylated loci across cortex. BioRxiv. 2020.
76. Haaksma ML, Vilela LR, Marengoni A, Calderón-Larrañaga A, Leoutsakos J-MS, Rikkert MGO, et al. Comorbidity and progression of late onset Alzheimer's disease: a systematic review. PLoS One. 2017;12(5).
77. Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. J Chem Neuroanat. 2011;42(2):118-26.

78. Carless MA. Determination of DNA methylation levels using Illumina HumanMethylation450 BeadChips. Chromatin Protocols: Springer; 2015. p. 143-92.
79. Bassil CF, Huang Z, Murphy SK. Bisulfite pyrosequencing. Ovarian Cancer: Springer; 2013. p. 95-107.
80. Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841-51.
81. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nature methods. 2016;13(2):127.
82. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell. 2006;126(4):663-76.
83. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. cell. 2007;131(5):861-72.
84. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. science. 2007;318(5858):1917-20.
85. Raab S, Klingenstein M, Liebau S, Linta L. A comparative view on human somatic cell sources for iPSC generation. Stem cells international. 2014;2014.
86. Malik N, Rao MS. A review of the methods for human iPSC derivation. Pluripotent stem cells: Springer; 2013. p. 23-33.
87. Riemens RJM, van den Hove DLA, Esteller M, Delgado-Morales R. Directing neuronal cell fate in vitro: achievements and challenges. Progress in neurobiology. 2018;168:42-68.
88. Zheng W, Li Q, Zhao C, Da Y, Zhang H-L, Chen Z. Differentiation of glial cells from hiPSCs: potential applications in neurological diseases and cell replacement therapy. Frontiers in cellular neuroscience. 2018;12:239.
89. Boland MJ, Nazor KL, Loring JF. Epigenetic regulation of pluripotency and differentiation. Circulation research. 2014;115(2):311-24.
90. Penney J, Ralvenius WT, Tsai L-H. Modeling Alzheimer's disease with iPSC-derived brain cells. Molecular Psychiatry. 2019:1-20.
91. Haston KM, Finkbeiner S. Clinical Trials in a Dish: The Potential of Pluripotent Stem Cells to Develop Therapies for Neurodegenerative Diseases. Annu Rev Pharmacol Toxicol. 2015.
92. Ross CA, Akimov SS. Human-induced pluripotent stem cells: potential for neurodegenerative diseases. Hum Mol Genet. 2014;23(R1):R17-26.
93. Zhang D, Pekkanen-Mattila M, Shahsavani M, Falk A, Teixeira AI, Herland A. A 3D Alzheimer's disease culture model and the induction of P21-activated kinase mediated sensing in iPSC derived neurons. Biomaterials. 2014;35(5):1420-8.
94. Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nat Rev Drug Discov. 2016.
95. Drummond E, Wisniewski T. Alzheimer's disease: experimental models and reality. Acta neuropathologica. 2017;133(2):155-75.
96. Williams T, Borchelt DR, Chakrabarty P. Therapeutic approaches targeting Apolipoprotein E function in Alzheimer's disease. Molecular Neurodegeneration. 2020;15(1):8.
97. van den Hove DLA, Riemens RJM, Koulousakis P, Pishva E. Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges. Brain Pathology. 2020.
98. Yang J, Li S, He XB, Cheng C, Le W. Induced pluripotent stem cells in Alzheimer's disease: applications for disease modeling and cell-replacement therapy. Mol Neurodegener. 2016;11(1):39.
99. Riemens R, Kenis G, van den Beucken T. Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease. Neurobiology of learning and memory. 2020:107318.

100. Riemens RJM, Soares ES, Esteller M, Delgado-Morales R. Stem Cell Technology for (Epi) genetic Brain Disorders. Neuroepigenomics in Aging and Disease: Springer; 2017. p. 443-75.

Chapter 2

Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges

Renzo J.M. Riemens[1,3]*, Daniël L.A. van den Hove[1,2]*, Philippos Koulousakis[1]*, Ehsan Pishva[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Division of Molecular Psychiatry, Laboratory of Translational Neuroscience, Center of Mental Health, Department of Psychiatry, University of Würzburg, Würzburg, Germany.
[3]Institute of Human Genetics, Julius Maximilians University, Würzburg, Germany.
[4]College of Medicine and Health, University of Exeter Medical School, Exeter University, Exeter, UK.

*Authors contributed equally to this work.

Brain Pathology. 2020 Sep;30(5):978-983.
Doi: 10.1111/bpa.12880.

Abstract
Alzheimer's disease (AD) represents a devastating progressive neurodegenerative disease with a complex pathophysiology, affecting millions of people worldwide. Recent epigenome-wide association studies suggest a key role for epigenetic mechanisms in its development and course. Despite the fact that current evidence on the role of epigenetic dysregulation in aging and AD is convincing, the pioneering field of neuroepigenetics is still facing many challenges that need to be addressed to fundamentally increase our understanding about the underlying mechanisms of this neurodegenerative disorder. This perspective paper describes the current state of play for epigenetic research into AD and discusses how new methodological advances in the field of epigenetics and related data science disciplines could further spur the development of novel therapeutic agents and biomarker assays.

Keywords
Alzheimer's disease; DNA methylation; epigenetics; methylome-wide association studies.

Epigenome-scale studies in Alzheimer's disease

Alzheimer's disease (AD) is a chronic neurodegenerative disorder characterized by the deposition of amyloid-beta (Aβ) plaques, neurofibrillary tangles (NFTs) of hyperphosphorylated tau protein, which ultimately lead to neuronal cell death. Genetic research exploring the etiology of AD has helped elucidate some of the pathogenic mechanisms underlying the disorder. Mutations in the *APP*, *PSEN1* and *PSEN2* genes are key players in developing early onset familial AD (fAD), whereas *APOE* was discovered as the most important genetic risk factor for late onset sporadic AD (sAD) [1]. Genome-wide association studies (GWAS) have helped identify several AD loci that are common in the general population, but show small risk effects (the most recent genetic studies of AD are summarized elsewhere in this mini-symposium series [2]). However, as only one third of AD cases can be explained by common variants or missing heritability, it is highly likely that nongenetic factors play a substantial role in the development and course of the disorder. As such, the multitude of nongenetic risk factors for developing sAD suggests the involvement of epigenetic mechanisms [3].

Epigenetic mechanisms refer to processes that regulate gene expression without altering the DNA sequence. Changes in the epigenome are acquired throughout life and are subject to alterations based on the environment that the cell or organism is exposed to. There are various types of epigenetic modifications, which include DNA modifications, histone modifications, as well as noncoding RNAs (ncRNAs) [4]. This perspective particularly focuses on the DNA modification of 5-methylcytosine (5-mC) as it has been the best characterized in the context of AD, discussing recent scientific achievements and challenges in the context of AD. Over the last decade, major technological advances have allowed the first epigenome-wide association studies (EWAS) of DNA methylation (herein termed methylome-wide association studies (MWAS)) in AD brain tissue (for a timely review, see [5]; in addition, see [6]). The first MWAS on AD made use of Illumina Infinium Human Methylation 27K arrays to study more than 27.000 CpG sites in 14.475 genes within the frontal cortex [7] (see Table 1). The most significant AD-associated differentially methylated cytosine-phosphate-guanine (CpG) site was located within the *TMEM59* gene, and displayed DNA hypomethylation when comparing 12 AD patients with 12 age- and gendermatched controls. A similar 27K MWAS on 15 AD patients and 5 non-demented controls identified Braak-associated DNA methylation alterations in four loci, including two residing in *DUSP22*, displaying DNA hypermethylation, concomitant with decreased *DUSP22* RNA expression, with increasing Braak stage [8]. Subsequent approaches made use of the Illumina Infinium Human Methylation 450K array, which interrogates more than 485.000 CpG sites. Several larger sample-size 450K studies (N = 68–740 depending on study),

primarily focusing on cortical brain tissue, provided converging evidence for numerous genes displaying differential methylation in AD, that is, *ANK1*, *C10orf105*, *CDH23*, *DIP2A*, *LOC100507547*, *PPT2*, *PPT2-EGFL8*, *PRDM16*, *PRRT1*, *RHBDF2*, *RNF39*, *RPL13*, *SERPINF1* and *SERPINF2* [9–12].

Interestingly, of the genes identified, *ANK1*, *CHD23*, *DIP2A*, *RHBDF2*, *RPL13*, *SERPINF1* and *SERPINF2* were also found to display significant AD-associated gene expression changes, supporting a potential functional role for DNA methylation in AD [10]. Differential methylation of *ANK1* has been established in five studies [6, 9, 12–14]. Furthermore, two genes of the dual specificity phosphatase (DUSP) family of proteins were shown to be differentially methylated in two independent studies as well [8, 11]. An overview of key findings and study designs for all MWAS published to date can be found in Table 1. Notably, a recent cross-cortical meta-analysis performed on nine previously published MWAS data sets consisting of 1408 donors, identified 220 CpGs significantly associated with AD neuropathology. These probes were annotated to 121 unique genes, of which 96 had not been previously reported in AD MWAS [15].

Although these MWAS findings have been highly valuable for improving our understanding of the molecular etiology of AD, there are a number of limitations to these approaches. Such challenges would need to be addressed properly in future endeavors to allow for appropriate reliable interpretation of generated data in order to significantly increase our understanding about the disease processes underlying AD. In this mini-symposium article, we showcase DNA methylation EWAS in AD, addressing several achievements and challenges in this respect in more detail. While this particular perspective article focuses on DNA methylation studies, there is also a substantial body of literature on other levels of epigenetic regulation, including ncRNAs such as microRNAs (miRNAs). For a timely review, in the context of AD, see [16]. Notably, many of the considerations referred to below do apply to other epigenetic processes as well.

Genomic coverage

Current state-of-the-art EWAS platforms used for studying DNA methylation alterations in AD have mostly been microarray based. This approach has been largely successful owing to their sensitivity and relative ease of execution in terms of analysis and standardization of protocols, allowing for a reliable comparison between studies. Ultimately, the common platform used across the MWAS to date would also lend itself well to meta-analyses, the first of which has just been undertaken in the case of AD [15]. However, despite a significant increase in CpG coverage on the Illumina microarrays over recent years, current EWAS platforms

still interrogate only a small proportion of CpG sites and predominantly assess CpG-rich promoter regulatory regions, hence, missing a large proportion of the human DNA methylome. As an example, even though the latest Illumina Infinium Methylation EPIC BeadChip array, which allows for the interrogation of over 850.000 CpG sites across the genome, shows improved coverage of regulatory elements, such as enhancers, when compared to its predecessors, even now, only a limited proportion of distal and proximal regulatory elements are represented [17]. For a carefully annotated description of CpG coverage on Illumina arrays, see [18] and [19]. These EWAS microarray platforms further lack the possibility to thoroughly assess DNA methylation at non-CpG (cytosine and adenine) sites. The use of (more expensive) next-generation sequencing-based approaches allowing for methylation assessment across the entire genome shows great promise in this respect, although these come with other major challenges as well, including those related to power (requiring larger sample sizes in view of multiple testing; see below), sequencing depth and associated costs.

Statistical power

Existing MWAS findings should be interpreted with caution, as most published studies seem (severely) underpowered. Power calculations for EPIC-based DNA methylation studies demonstrate that existing studies with data on ~1000 samples are adequately powered to detect small differences at the majority of sites [20]. Furthermore, for future endeavors that may use next generation sequencing approaches to increase coverage, increased samples sizes will be required. Evidently, when taking into account the intrinsic methylation susceptibility of specific CpG sites (*i.e.* likelihood of an individual CpG to be methylated, and the nucleotide distance between neighboring CpG sites [21]), or when assessing differentially methylated regions (DMRs) consisting of adjacent differentially methylated positions (DMPs), smaller sample sizes are sufficient to reach genome-wide significance. In this respect, Bonferroni correction for the number of probes on the array is often presumed to be too conservative as DNA methylation values at neighboring probes are generally known to be correlated. As such, in addition to false-discovery rate (FDR) correction, recent endeavors attempt to estimate the number of independent tests performed in an EWAS and adjust the significance level accordingly [20,22]. Altogether, although it is important to consider sample size when interpreting EWAS results, it is worth noting that different CpG sites show different variance and they will, therefore, be differently powered at the same sample size.

Specificity of the epigenetic marks assessed

Accumulating evidence suggests an important role for other types of cytosine modifications such as 5-hydroxymethylcytosine (5-hmC) in the human brain [23].

However, the great majority of MWAS on AD published to date have made use of sodium bisulfite-treated DNA, which is unable to discriminate between 5-mC and its first oxidation product, 5-hmC. Recently, the first EWAS simultaneously assessing DNA methylation and hydroxymethylation in isolation in AD by making use of oxidative-bisulfite DNA treatment was published [13]. This study highlighted different levels of 5-mC, 5-hmC or unmodified cytosine (5-uC) in a number of different genes, including *WNT5B*, *ANK1*, *ARID5B*, *FBXL16*, *ALLC* and *JAG2*. More specifically, in view of the earlier MWAS findings on *ANK1*, this study demonstrated that *ANK1* DNA modification alterations in AD are as a result of DNA hypermethylation and hypohydroxymethylation, suggestive of a loss of active DNA demethylation of *ANK1* in AD. Recently, a similar approach was used in AD and control individuals using cortical tissue from the middle temporal gyrus and peripheral whole blood samples and led to the identification of a common DMR associated with AD, close to the transcription start site of *OXT* [24]. The study suggested *OXT* 5-mC and 5-hmC levels change in opposite directions within the middle temporal gyrus in AD. Interestingly, the detection of a differentially methylated region near *OXT* is in line with a recent report of a nearly identical AD-associated *OXT* DMR in the superior temporal gyrus, which is located directly above the middle temporal gyrus [11].

While earlier studies on DNA modification changes have primarily focused on 5-mC and 5-hmC, other marks such as 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), which both represent oxidized derivatives of 5-hmC, are heavily understudied. Little is known about the frequency or functionality of these modifications in the healthy or diseased human brain. Further studies are therefore required to develop a better understanding on their potential role in the development and course of AD. Alternative, more specific technologies such as DNA-immunoprecipitation (IP) sequencing approaches with DNA being captured by specific antibodies to distinct DNA modifications hold great promise for assessing these epigenetic marks. In addition, direct DNA sequencing using novel third generation sequencing platforms, such as the Oxford Nanopore Technologies (ONT) minion/ promethion or the Pacific Biosciences (Pacbio) Sequel, allows calling of both single nucleotide polymorphisms (SNPs) and a range of different DNA modifications by sequencing native DNA without prior amplification via PCR or labeling the sample chemically, which represents another promising approach [25]. Evidently, these more advanced sequencing approaches do still require distinct, more challenging data science investments for calling of different DNA modifications.

Table 1. Summary of findings from AD MWAS.

Author	Year	Sample size	Method	Tissue type	DNA treatment and modification identified	Top differentially methylated loci
Bakulski *et al.* [7]	2012	24	Illumina Methylation 27K array	Frontal cortex	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	TMEM59
Sanchez-Mut *et al.* [8]	2014	20	Illumina Methylation 27K array	Hippocampus	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	DUSP22
Lunnon *et. al* [6]	2014	122	Illumina Methylation 450K array	Superior temporal gyrus, prefrontal cortex, cerebellum, blood	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	ANK1
De Jaeger *et al.* [9]	2014	708	Illumina Methylation 450K array	Dorsolateral prefrontal cortex	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	ABCA7, BIN1, ANK1, CDH23, DIP2A, RHBDF2, RPL13, SERPINF1, SERPINF2
Watson *et al.* [11]	2016	68	Illumina Methylation 450K array	Superior temporal gyrus	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	MOV10L1, B3GALT4, DUSP6, TBX15, OXT
Gasparoni *et al.* [14]	2018	128	Illumina Methylation 450K array	Bulk brain tissue (Frontal + temporal cortex)	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	CLU, SYNJ2, NCOR2, RAI1, CXXC5, INPP5A, MCF2L, ANK1, MAP2, LRRC8B, STK32C, S100B
Smith *et al.* [34]	2018	147	Illumina Methylation 450K array	Prefrontal cortex, superior temporal gyrus, hippocampus	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	HOXA3
Semick *et al.* [12]	2019	73	Illumina Methylation 450K array	Hippocampus, entorhinal cortex, dorsolateral prefrontal cortex, cerebellum	Bisulfite treatment (total DNA modifications [5-mC+5-hmC])	ANK1, ANKRD30B, WDR81, SERPINF2, MYO1C
Smith *et al.* [13]	2019	96	Illumina Methylation 450K array	Entorhinal cortex	Oxidative-bisulfite treatment (5-mC, 5-hmC and 5-uC independently	WNT5B, FBXL16, ANK1, ARID5B, ALLC

Table 1. *(Continued)*

Author	Year	Sample size	Method	Tissue type	DNA treatment and modification identified	Top differentially methylated loci
Lardenoije *et al.* [24]	2019	80	Illumina Methylation 450K array	Middle temporal gyrus, blood	Oxidative-bisulfite treatment (5-mC, 5-hmC and 5-uC independently)	OXT, CHRNB1, RHBDF2, C3

Abbreviations: 5-hmC = hydroxymethylation; 5-mC = methylation; 5-uC = unmodified cytosine.

Cell-type specificity and composition

A major issue that challenges the field of neuroepigenetics is that of tissue and cell-type specificity of epigenetic modifications. AD is characterized by neuronal loss, and activation of glia cells, concomitant with alterations in the cell-type composition of brain samples, which challenges the correct interpretation of results when making use of heterogeneous bulk tissue samples. As such, cell-type specific modifications in one cell-type could, for example, mask changes in another [26]. Aside from cell-type specific changes in activity or changes in cellular proportions that can occur during the development and course of AD, differences in cell-type compositions between samples derived from different individuals that arise as a result of tissue sampling is also a consideration. A workable, though often expensive and tedious, solution to this issue could be specifically profiling cell-type specific samples, to be acquired via, for example, laser capture microdissection (LCM), fluorescence-activated cell sorting (FACS) or magnetic-activated cell sorting (MACS). This could even exploit single-cell sequencing technologies to investigate cell-specific epigenomes [27]. Alternatively, one can use advanced bioinformatic approaches to correct for cell-type composition in MWAS data generated in bulk (unsorted) tissue. As an example, one can estimate neuronglia proportions based on DNA methylation values in EWAS data and this approach has been adopted by many in the AD epigenetics field [28-30]. Akin to the issues associated with studying epigenetic changes in the brain, cell type specificity issues also apply to studies making use of whole blood samples and these can be addressed in similar ways, for example, by adjusting data for cell-type specific DNA methylation markers [20, 21].

Causal inference

At present, one of the major limitations of current endeavors investigating the role of epigenetic dysregulation in aging and AD, is the issue of causality. Alterations in DNA modifications could be either causal in the disease process or could themselves arise as a direct or indirect result of pathological or secondary behavioral and psychological changes associated with disease. Similarly, age,

comorbidity or the use of therapeutic agents may have a profound impact on the epigenome, both in the brain and blood. In addition, epigenetic signatures identified in advanced AD cases, particularly in brain regions affected relatively late in the disease process, may provide limited information on causality. An interesting notion in this regard is the study of epigenomic changes with respect to the spatiotemporal spread of pathology in AD. For example, utilizing postmortem AD brain tissue derived from donors at different disease stages and comparing the epigenome of brain regions implicated in the early stages of AD (*i.e.* brainstem) to those regions affected in more advanced phases (*i.e.* cortical regions) in the same individuals, could allow the identification of the relative, spatiotemporal contribution of potentially causal, region-specific epigenetic alterations in the development and course of AD. Similarly, when assessing the blood epigenome, longitudinal studies capturing early stages of cognitive decline, and preferably also exposure to environmental factors throughout life (which may evidently also impact on the blood epigenome), are needed [31]. Alternatively, rodent models of AD could also be used to explore causal epigenetic mechanism in AD, by studying alterations prior to the onset of pathology. However, one limitation in this regard is that these murine systems are models of familial AD, bearing autosomal dominant mutations and are not a true model of sporadic disease. Another approach to elucidate causal epigenetic mechanisms in AD would be through the use of epigenetic editing. The use of novel editing constructs based on the catalytically deactivated clustered regularly interspaced short palindromic repeats (CRISPR) associated protein 9 (*i.e.* CRISPR-dCas9) fused to chromatin modifying enzymes such as DNA methyltransferase 3a (DNMT3A) or TET1, also known as EpiEffectors [32], allows adding, or removing, of methyl groups to specific DNA loci, respectively [33]. The use of this system *in vitro* in, for example induced pluripotent stem cells (iPSCs) could ultimately allow the field to determine the functional implications of epigenetic signatures identified in MWAS and to establish whether these represent a cause, a consequence, or both, of the disease process. Further information on this concept can be found elsewhere in this mini-symposium series [35]. As epigenetic changes are often reversible, the identification of causative epigenetic signatures may provide promising targets for future therapeutic interventions.

Multi-omics approaches

Importantly, the ability to identify specific and early signatures of AD is hindered by the substantial clinical heterogeneity among patients, for instance there is considerable variability in disease duration and cognitive decline rate between patients, which likely reflects inter-individual variation in genetics, exposure to environmental factors and underlying neuropathology. In recent years, several “big data” approaches have allowed the identification of robust genomic,

epigenomic and transcriptomic changes in AD. However, these studies have largely been restricted to correlating a single layer of molecular information with respect to a single measure of disease severity, in a single tissue, and it is not yet possible to identify the exact mechanisms involved in the pathophysiology of AD or to predict a patient's disease trajectory with current peripheral biomarkers. The recent development of powerful computational frameworks now offers multilayer inter-regulatory approaches to understand the development and course of AD, while accounting for inter-individual differences in genotype and exposure to environmental risk factors. As such, multi-omics approaches – that include an integrative analysis of various layers of epigenetic regulation – and associated data science tools show great promise in the development of novel diagnostic tools and treatment strategies for AD, with further details on this approach provided elsewhere in this mini-symposium [36].

Concluding remarks

There is increasing interest in the role of epigenetic dysregulation in aging and AD, with the primary focus of most EWAS studies being on DNA methylation. Platforms such as the Illumina Infinium Methylation 450K and EPIC BeadChip arrays have enabled cost-efficient, high-throughput profiling of methylomic variation across large numbers of samples. EWAS results should be interpreted with caution though, particularly when dealing with low sample-size studies. Future endeavors should aim for specific, adequately powered approaches, with large sample sizes and well-characterized environmental, medication and ante- and/or postmortem data, while considering tissue and cellular heterogeneity. Moreover, candidate signatures should be functionally validated to determine whether they could represent novel causal mechanisms.

Acknowledgments

Funds have been provided by the Joint Program— Neurodegenerative Disease Research (JPND) for the EPI-AD consortium (http://www.neurodegeneration research.eu/wpcontent/uploads/2015/10/Factsheet_EPI-AD.pdf). The project is supported through the following funding organizations under the aegis of JPND; the Netherlands, The Netherlands Organization for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 643417.

References

1. Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: back to the future. Neuron. 2010;68(2):270-81.
2. Bertram L, Tanzi RE. Genomic mechanisms in Alzheimer's disease. Brain Pathology. 2020;30(5):966-77.
3. Liu X, Jiao B, Shen L. The epigenetics of Alzheimer's disease: Factors and therapeutic implications. Frontiers in genetics. 2018;9:579.
4. Chouliaras L, Rutten BP, Kenis G, Peerbooms O, Visser PJ, Verhey F, et al. Epigenetic regulation in the pathophysiology of Alzheimer's disease. Progress in neurobiology. 2010;90(4):498-510.
5. Lardenoije R, Pishva E, Lunnon K, van den Hove DL. Neuroepigenetics of Aging and Age-Related Neurodegenerative Disorders. Prog Mol Biol Transl Sci. 2018;158:49-82.
6. Lunnon K, Mill J. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2013;162B(8):789-99.
7. Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, et al. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. Journal of Alzheimer's Disease. 2012;29(3):571-88.
8. Sanchez-Mut JV, Aso E, Heyn H, Matsuda T, Bock C, Ferrer I, et al. Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. Hippocampus. 2014;24(4):363-8.
9. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17(9):1156-63.
10. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.
11. Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. Genome medicine. 2016;8(1):1-14.
12. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, et al. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. Acta neuropathologica. 2019;137(4):557-69.
13. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenetics. 2019;11(1):52.
14. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics & chromatin. 2018;11(1):41.
15. Smith R, Pishva E, Shireby G, Smith AR, Roubroeks JA, Hannon E, et al. Meta-analysis of epigenome-wide association studies in Alzheime's disease highlights 220 differentially methylated loci across cortex. BioRxiv. 2020.
16. Van den Hove DL, Kompotis K, Lardenoije R, Kenis G, Mill J, Steinbusch HW, et al. Epigenetically regulated microRNAs in Alzheimer's disease. Neurobiology of aging. 2014;35(4):731-45.
17. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome biology. 2016;17(1):1-17.

18. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702.
19. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389-99.
20. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC genomics. 2019;20(1):1-15.
21. Affinito O, Palumbo D, Fierro A, Cuomo M, De Riso G, Monticelli A, et al. Nucleotide distance influences co-methylation between nearby CpG sites. Genomics. 2020;112(1):144-50.
22. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. Genetic epidemiology. 2018;42(1):20-33.
23. Cheng Y, Bernstein A, Chen D, Jin P. 5-Hydroxymethylcytosine: a new player in brain disorders? Experimental neurology. 2015;268:3-9.
24. Lardenoije R, Roubroeks J, Pishva E, Leber M, Wagner H, Iatrou A, et al. T40. Alzheimer's Disease DNA (Hydroxy) Methylome in the Brain and Blood: Evidence for OXT Methylation as a Preclinical Marker. Biological psychiatry. 2019;85(10):S144.
25. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology. 2016;17(1):1-11.
26. Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. J Chem Neuroanat. 2011;42(2):118-26.
27. Bheda P, Schneider R. Epigenetics reloaded: the single-cell revolution. Trends in cell biology. 2014;24(11):712-23.
28. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics. 2013;8(3):290-302.
29. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13(1):86.
30. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):3156.
31. van den Hove DL, Kenis G, Rutten BP. Epigenetic dysregulation in Alzheimer's disease: cause or consequence? Epigenomics. 2014;6(1):9-11.
32. Kungulovski G, Jeltsch A. Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends in Genetics. 2016;32(2):101-13.
33. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA Methylation in the Mammalian Genome. Cell. 2016;167(1):233-47 e17.
34. Smith RG, Hannon E, De Jager PL, Chibnik L, Lott SJ, Condliffe D, et al. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimer's & Dementia. 2018;14(12):1580-8.
35. Schrauben M, Dempster E, Lunnon K. Applying gene-editing technology to elucidate the functional consequence of genetic and epigenetic variation in Alzheimer's disease. Brain Pathology. 2020;30(5):992-1004.
36. Ma Y, Klein HU, De Jager PL. Considerations for integrative multi-omic approaches to explore Alzheimer's disease mechanisms. Brain Pathology. 2020;30(5):984-91.

Chapter 3

Targeted detection of unmodified cytosine, 5-methylcytosine and 5-hydroxymethylcytosine levels at single CpG sites by oxidative bisulfite pyrosequencing

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Diego Mastroeni[3], Thomas Haaf[2], Daniël L.A. van den Hove[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]L.J. Roberts Alzheimer's Disease Center, Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

Abstract
Classical bisulfite pyrosequencing does not allow for the discrimination between DNA methylation and hydroxymethylation, *i.e.* between 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC), respectively, in the context of cytosine-phosphate-guanine (CpG) sites. As such, it provides cumulative measures on the levels of both modifications. Given that 5-hmC is known to be enriched in the human brain, neuroepigenetic studies that initially relied on this technique might therefore have over- or underestimated biological relevant effects related to each of these individual CpG modifications. In order to overcome this issue, an alternative approach based on a highly selective chemical oxidation step prior to bisulfite treatment has recently been developed, which allows for the discrimination between 5-mC and 5-hmC bases. However, to date, this approach has not been directly demonstrated on, let alone fully optimized for the pyrosequencing system. Moreover, a spike-in DNA standard specifically designed for this platform that would allow the user to accurately determine the oxidative bisulfite conversion efficiency is currently not commercially available. In the present study, we therefore developed three spike-in pyrosequencing controls that can be added to a given DNA sample prior to (oxidative) bisulfite treatment. As such, we here demonstrate a relative low-cost oxidative bisulfite pyrosequencing approach that we successfully applied for the detection of 5-mC and 5-hmC, as well as unmodified cytosine (5-uC) bases. We show that levels of all three modifications can be detected following this approach in *OXT* and *DNAJB13* in both post-mortem brain tissue and cultured induced pluripotent stem cells (iPSCs). Finally, we show that our 5-hmC spike-in control can be used as internal pyrosequencing control that does not interfere with the analysis of the accompanying sample, and vice versa.

Keywords
Pyrosequencing, oxidative bisulfite conversion, methylation, hydroxymethylation, epigenetics.

Introduction

Pyrosequencing is a method of DNA sequencing that allows for quantitative mapping of the methylation status of individual cytosine-phosphate-guanine (CpG) sites in a target region within the DNA [1]. Amongst others, this technique has been widely used to study CpG methylation levels in candidate loci in brain tissues derived from patient cohorts, including Alzheimer's disease patients [2-4]. As such, it is often used as an independent, alternative method to validate methylation differences identified by means of an epigenome-wide association study (EWAS). In addition, the pyrosequencing platform is also very popular for more explorative studies, assessing CpG methylation levels of biosamples in a specific predetermined region of a target gene. The technique itself is based on the 'sequencing-by-synthesis' principle, which is performed by detecting deoxynucleotide triphosphates (dNTPs) incorporated by a DNA polymerase on the complementary side of a single stranded DNA target molecule [1]. Similar to other bisulfite-based techniques, bisulfite pyrosequencing furthermore relies on treatment of genomic (target) DNA with sodium bisulfite [1]. This treatment is applied prior to polymerase chain reaction (PCR) amplification and subsequent sequencing, and results in a base-specific conversion during which unmodified cytosine (5-uC) bases deanimate to uracil. Consequently, converted cytosines are read as thymines during the sequencing stage (Figure 1). Methylated cytosines, *i.e.* 5-methylcytosine (5-mC) bases, on the other hand, are resistant to bisulfite treatment and are hence read as cytosines, thus allowing for the discrimination between unmodified and methylated cytosines in the context of CpG sites.

Even though classical bisulfite pyrosequencing has been widely applied for the detection of DNA methylation, it was recently discovered that treatment of 5-hydroxymethylcytosine (5-hmC) with bisulfite salts results in a stable methyl-sulfonate adduct that is also read as a cytosine when sequenced [5]. For this reason, 5-mC and 5-hmC, while functionally distinct, are indistinguishable from another after bisulfite conversion of genomic DNA. In fact, classical bisulfite techniques provide cumulative measurements on the levels of both modifications rather than those of 5-mC alone. This is problematic, as many neuroepigenetic studies to date have relied on this approach, while it is well recognized that 5-hmC is highly enriched in the human brain [6, 7]. Previously acquired neuroepigenetic data could therefore easily be confounded due to the presence of both modifications at a specific CpG site on different DNA molecules in a given sample. Consequently, these studies might have overestimated biologically relevant effects related to 5-mC, in reality attributable to 5-hmC, or, on the other hand, have overlooked or underestimated effects related to 5-mC and/or 5-hmC, *e.g.* when both measures were to cancel each other out. In order to overcome

this issue, an alternative approach to discriminate between both modifications has recently been introduced when performing EWAS, *i.e.* a procedure that makes use of an oxidation step prior to bisulfite treatment [5]. In this method, 5-mC and 5-hmC can be discriminated from another via a highly selective chemical oxidation of 5-hmC to 5-formylcytosine (5-fC) (Figure 1) [5]. In contrast to 5-hmC, which does not deaminate, bisulfite treatment of 5-fC results in deformylation and deamination, yielding uracil, which at the sequencing step reads as a thymine. Thus, the only base that is not deaminated and therefore read as a cytosine after oxidative bisulfite sequencing is 5-mC. This approach therefore gives a more accurate readout on 5-mC levels and simultaneously allows for the detection of 5-hmC by quantitative subtraction of the bisulfite and oxidative bisulfite data derived from a single DNA sample. In addition, levels of 5-uC can be also be ascertained after pyrosequencing by subtraction of the bisulfite signal from 100% [2].

Although oxidative bisulfite sequencing provides a solution to the issue of modification specificity in neuroepigenetic studies, the oxidative bisulfite treatment has mainly been developed and optimized for EWAS. This includes techniques such as reduced representation bisulfite sequencing (RRBS) and Illumina BeadArrays [8, 9]. Even though one might assume that the oxidative bisulfite method should be compatible with pyrosequencing as well, a spike-in DNA standard that would allow the user to accurately determine the conversion efficiency of 5-hmC for specifically this platform is currently not commercially available. Moreover, despite the fact that oxidative bisulfite sequencing has been extensively validated using the aforementioned genome-wide techniques, as well as by mass spectrometry, to our knowledge, it has never been directly validated on the pyrosequencing system itself. To this end, we have developed three aligned spike-in DNA pyrosequencing controls that can be added to a DNA target sample prior to (oxidative) bisulfite treatment. Additionally, as a showcase, we here demonstrate a relative low-cost oxidative bisulfite pyrosequencing approach that we successfully applied for the detection of both 5-uC, 5-mC and 5-hmC, in post-mortem brain tissue and cultured induced pluripotent stem cells (iPSCs). We demonstrate that levels of all three modifications can be detected following this approach in both brain tissue and iPSCs, by assessing DNA methylation and hydroxymethylation of *OXT* and *DNAJB13* as examples. Finally, we show that our spike-in controls can be used as internal pyrosequencing controls that do not interfere with the analysis of the accompanying sample, and vice versa.

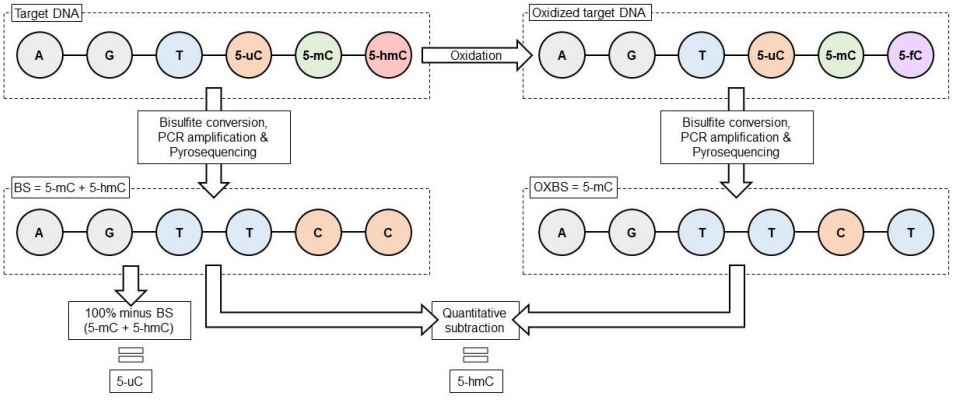


Figure 1. Overview of the oxidative bisulfite treatment for the analysis of both 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC). The figure demonstrates the general principle of the oxidative bisulfite treatment, allowing for the discrimination between 5-mC and 5-hmC on genomic target DNA. After a traditional bisulfite treatment and polymerase chain reaction (PCR) amplification, an unmodified cytosine (5-uC) reads as a thymine (T) during pyrosequencing, whereas both 5-mC and 5-hmC read as a cytosine (C). By applying a highly selective oxidation of 5-hmC to 5-formylcytosine (5-fC) prior to the bisulfite treatment and PCR amplification, both 5-uC and 5-hmC read as T during pyrosequencing, whereas only 5-mC reads as C, thereby providing a readout specifically for 5-mC. Levels of 5-hmC can then be obtained by applying a quantitative subtraction of both the readouts from a bisulfite- (BS) and oxidative-bisulfite (OXBS) converted target DNA sample, whereas levels of 5-uC can be determined by subtracting the bisulfite signal from 100% after pyrosequencing. In both of the approaches, adenosines (A), guanines (G) and thymines (T) remain unchanged.

Materials and methods

A step-wise overview of the oxidative bisulfite conversion protocol is available in the Supplementary material.

Sample collection and ethics statement

Post-mortem middle temporal gyrus (MTG) tissue from an Alzheimer's disease patient was received from the Banner Sun Health Research Institute (BSHRI) Brain and Body Donation Program (BBDP). A local institutional review board-approved informed consent form was signed by the participant, including specific consent to the use of donated tissue for future research. iPSCs were obtained by episomal reprogramming of mononuclear blood cells that were derived from a healthy Southern European male that was in his late twenties at the time of donation. The participant signed an institutional review board-approved informed consent form, including specific consent for the use of donated tissues for iPSC generation and scientific research. All laboratory procedures for iPSC reprogramming were performed at Maastricht University (Maastricht, the Netherlands). Clones of growing iPSC colonies were selected, expanded and characterized according to general quality control guidelines, including tests such as the evaluation of pluripotency marker expression, assessment of transgene

silencing, karyotyping by G-banding, embryoid body (EB) induction and directed differentiation into the three embryonic germ layers [10]. A single clone, *i.e.* CARIMi001-A, which surpassed the quality control criteria, was selected and used in the present study. For information on the used iPSC line, please also refer to https://hpscreg.eu/cell-line/CARIMi001-A. Further details on the subject demographics can be found in Supplementary Table 1.

Spike-in DNA pyrosequencing control production

A piece of synthetic DNA (syDNA) containing four unmodified CpG sites, as well as a compatible pyrosequencing assay, were manually designed using the PyroMark Assay Design 2.0 software (Qiagen, Hilden, Germany) (Supplementary Table 2). Both the syDNA, as well as the accompanying primers, were obtained from Metabion (Planegg, Germany). The syDNA underwent a reannealing procedure using a thermocycler in order to reassure duplex formation before further application. In brief, the syDNA was dissolved in an annealing buffer containing 60 mM KCL, 6 mM HEPES (pH 7.5) and 0.2 mM MgCl2, at a final concentration of 100 µM. The solution was incubated at 95°C for 1 minute and then the temperature was decreased by 1.5°C every minute thereafter until <25°C was reached. After the reannealing procedure, the solution was cooled down further and temporarily stored at 4°C up until further processing. The re-annealed syDNA duplex served as a template for PCR amplification for which either 5-uC, 5-mC or 5-hmC dNTP mixes (Zymo Research, Irvine, CA, United States) were used in order to produce the respective spike-in DNA pyrosequencing controls. Prior to amplification, the re-annealed syDNA was diluted to contain approximately 3000 copies per reaction, which is similar to conventional PCR reactions using 10 ng of genomic target DNA. Amplifications were performed with an initial denaturation step at 95°C for 5 minutes, followed by 35 cycles at 95°C, 60°C and 72°C for 30, 30, and at 15 seconds, respectively, with a final extension step at 72°C for 1 minute. Each individual reaction contained 2.5 µL PCR buffer (10X) with 20 mM MgCl2, 1 µL of each primer (10 µM stock), 0.2 µL (5 U/µL) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) and a final concentration of 0.2 mM dNTPs in a total volume of 25 µL. After PCR amplification, each of the respective spike-in DNA controls were purified using Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA, United States). The SSC buffer in the columns was exchanged with Milli-Q prior to purification. In brief, the lid of the column was removed, the packing buffer was drained by gravity and discarded. The column was then centrifuged at 1000 x g for 1 minute to remove the remaining buffer and washed three times with 500 µL of Milli-Q with intermediate centrifugation steps at 1000 x g for 1 minute. Finally, another 500 µL of Milli-Q was added and the column was centrifuged at 1.000 x g for 4 minutes. The complete 25 µL of PCR reaction was then added to the

center of the polyacrylamide gel matrix and the column was centrifuged at 1000 x g for 4 minutes. The purified spike-in DNA standards were then quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, United States). Each reaction was normalized to contain 3 ng/µL pyrosequencing standard and 1 µL was used per individual fraction to spike the DNA target samples.

DNA isolation and purification

DNA from ~50 mg MTG tissue and 5 x 10^6 iPSCs was isolated using a phenol/chloroform/isoamyl alcohol (PCI) extraction method. In brief, the samples were lysated in 500 µl lysis buffer containing 50 mM Tris (pH 8.0), 1 mM EDTA and 0.5% SDS. After adding 25 µl of proteinase K (Thermo Fisher Scientific, Waltham, MA, USA; Ambion AM2548; 20 mg/ml stock solution), the samples were incubated at 56°C in a shaking thermoblock at 1000 rpm for 4 hours until a clear solution appeared. Following incubation, the proteinase K was inactivated at 80°C for 10 minutes. After proteinase K inactivation, PCI (Sigma Aldrich, Saint Louis, MO, USA; 77617-100) was added in a ratio of 1:1, the samples were mixed for 5 minutes and then centrifuged at 14.000 rpm for 5 minutes. The upper phase was carefully removed from each sample and transferred to a new sterile 1.5ml Eppendorf tube. Another equal volume (1:1 ratio) of PCI was added, after which the samples were mixed for 5 minutes and centrifuged at 14.000 rpm for 5 minutes. The upper phase from each sample was once again transferred to new sterile 1.5 ml Eppendorf tube. The DNA was precipitated by adding 0.1 volume of 3 M NaAc (pH5.6) and 2.5 volumes of 100% cold (-20°C) ethanol, then incubated for 30 minutes at -80°C and centrifuged for 30 minutes at 14.000 rpm at 4°C. Subsequently, the solution was carefully removed, the DNA pellets were washed using 70% of cold ethanol and centrifuged for 30 minutes at 14.000 rpm at 4°C. Next, the ethanol was carefully removed and the DNA pellets were air dried at room temperature. Finally, the isolated DNA from each sample was eluted in a volume of 50µL Milli-Q and then stored at −20°C until further processing. At the day of analysis, the DNA was purified using Micro Bio-Spin™ P-6 Gel Columns. The SSC buffer in the columns was exchanged with Milli-Q prior to purifying the DNA as described earlier. After the buffer exchange, the DNA samples were added to the polyacrylamide gel matrixes, the columns were centrifuged at 1000 x g for 4 minutes and the samples were collected in fresh 1.5 mL Eppendorf tubes. The concentration of each sample was then quantified by using a Qubit™ dsDNA BR Assay Kit (Invitrogen). For all experiments, a total of 1 µg purified DNA from either the MTG tissue or the iPSCs was used for further analysis.

DNA denaturation, oxidation and pre-bisulfite purification

A 20X DNA denaturation solution was prepared by dissolving NaOH pellets in Milli-Q at a final concentration of 1 M. For the oxidant solution, a 10X stock was prepared by dissolving $KRuO_4$ (Sigma Aldrich) in a 0.5 M NaOH solution at a final concentration of 150 mM. The oxidant solution was stored in ready-to-use aliquots at -20°C until further use. At the day of the analysis, 1 µg of purified DNA was split into two single 1.5 mL Eppendorf tubes and each fraction was topped up to 17.1 µL using Milli-Q. The DNA was then denaturated by adding 0.9 µL 1 M NaOH to each of the fractions and incubated at 37°C for 30 minutes. During incubation, the oxidant solution was thawed on ice and diluted to a 1X working solution using Milli-Q. After incubation, 2 µL of 1X oxidant solution was added to only one of the two denaturated DNA fractions, while 2 µL of Milli-Q was added to the other. The fraction containing the oxidant solution was applied for oxidative bisulfite pyrosequencing, while the fraction without the oxidant was used for conventional bisulfite pyrosequencing. Both the brain and iPSC sample were incubated at 40°C for 10 minutes to allow DNA oxidation and were then purified using Micro Bio-Spin™ P-6 Gel Columns. The SSC buffer in the columns was exchanged with Milli-Q prior to purifying the denatured and oxidized DNA as described earlier.

DNA bisulfite conversion

Both the denatured and oxidized DNA fractions were individually bisulfite-converted using the EZ DNA Methylation-Direct Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions, with the exception of using an extended thermal cycler program. Bisulfite conversion was performed for two cycles at 98°C and 64°C for 8 minutes and 3.5 hours, respectively, with a final storage step at 4°C. After the bisulfite clean-up procedure, each sample was collected in a single 1.5 mL Eppendorf tube by flushing the spin column twice using 10 µL of elution buffer, resulting in a final concentration of 25 ng/µL for each fraction when assuming full recovery of the (oxidative) bisulfite-converted DNA.

Polymerase chain reaction

Primers targeting *OXT* and *DNAJB13* were designed using the PyroMark Assay Design 2.0 software (Qiagen) (Supplementary Table 3). A single amplicon for *OXT* and *DNAJB13* of 166 or 226 base pairs (bp), respectively, was amplified and sequenced using one sequencing primer covering eight or four CpG sites within an 85 or 46 bp region, respectively. PCR amplifications were performed with an initial denaturation step at 95°C for 5 minutes, followed by 45 cycles at 95°C, 58°C for *OXT* or 60°C for *DNAJB13* and 72°C for 30, 30 and 30 seconds, respectively, with a final extension step at 72°C for 1 minute. For each individual PCR reaction, 1 µl of the (oxidative) bisulfite-converted DNA was used as a

template. Each of the reactions contained 2.5 µL PCR buffer (10X) with 20 mM MgCl2, 0.5 µL 10 mM dNTP mix, 1 µL of each respective primer (5 µM stock) and 0.2 µL (5 U/µL) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) in a total volume of 25 µL. For PCR amplifications of the spike-in DNA standards after the oxidative bisulfite conversion, the same PCR conditions and thermocycler settings were used as described during the 'Spike-in DNA pyrosequencing control production'. The PCR products were visualized on a 2% agarose gel and 10 µL product was utilized per assay for (oxidative) bisulfite pyrosequencing.

Pyrosequencing

The Pyromark Q48 Autoprep system with the PyroMark Q48 Advanced CpG Reagents (Qiagen) and PyroMark Q48 Magnetic Beads (Qiagen) were used for oxidative bisulfite pyrosequencing according to the manufacturer's instructions. All genomic assays were tested for their sensitivity on various fractions, *i.e.* 0%, 20%, 40%, 60%, 80% and 100%, of methylated DNA standards that were constructed from the EpiTect PCR Control DNA Set (Qiagen). Modification levels at a single CpG resolution were analyzed with the Pyromark Q48 Autoprep software. Per CpG site, 5-uC values were calculated by subtracting the bisulfite signal from 100%, 5-mC values were considered equal to the oxidative bisulfite signal and 5-hmC values were estimated by subtracting the oxidative bisulfite signal from the bisulfite signal [2].

Results

Oxidation of DNA by $KRuO_4$ selectively converts 5-hmC to 5-fC

Previous studies have demonstrated that treatment of genomic DNA with $KRuO_4$ can be used in order to achieve specific oxidation of 5-hmC to 5-fC [11]. In order to confirm the selective and successful conversion of 5-hmC by the oxidant, three manually designed spike-in DNA standards, each containing four CpG sites with either solely 5-uC, 5-mC or 5-hmC bases, were subjected to both the classical bisulfite and novel oxidative bisulfite conversion protocol described above. Pyrosequencing of the bisulfite-converted standards revealed that 5-uC did convert to a uracil and was read as thymine, whereas both 5-mC and 5-hmC residues did not convert to a uracil and therefore were read as cytosine (Figure 2). Oxidative bisulfite pyrosequencing, on other hand, revealed that $KRuO_4$ selectively converted 5-hmC to 5-fC, which was demonstrated by the readout of a thymine at each CpG site. Similar to bisulfite pyrosequencing, the oxidative bisulfite treatment did convert 5-uC to uracil, which therefore was read as thymine, whereas 5-mC did not convert to uracil and was read as cytosine. Thus, these data confirm that the oxidative bisulfite pyrosequencing protocol described

above, based on the application of $KRuO_4$, selectively and successfully converts 5-hmC to 5-fC, while leaving 5-uC and 5-mC unaffected.

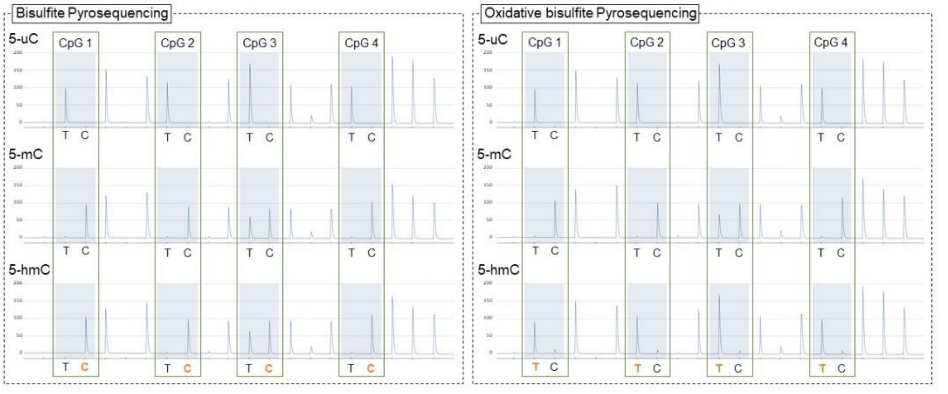


Figure 2. Pyrogram of both the bisulfite- and oxidative bisulfite-converted unmodified cytosine (5-uC), 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) spike-in DNA standards. The figure shows the pyrograms for all three DNA standards, *i.e.* 5-uC, 5-mC and 5-hmC, after both bisulfite- and oxidative bisulfite pyrosequencing. Displayed are all four the cytosine-phosphate-guanine (CpG) sites of each of the respective DNA standards. In the context of these CpG sites, bisulfite pyrosequencing revealed that 5-uC did convert to a uracil and was read as a thymine (T) whereas both 5-mC and 5-hmC residues did not convert to a uracil and therefore were read as a cytosine (C). Oxidative bisulfite pyrosequencing on other hand, revealed that the oxidant selectively converted 5-hmC to 5-formylcytosine (5-fC), which was demonstrated by the read-out of a T at each of the respective CpG sites. Similar to traditional bisulfite pyrosequencing, the oxidative bisulfite conversion did convert 5-uC to a uracil, which therefore was read as a T, whereas 5-mC did not convert to a uracil and therefore was read as a C.

The 5-hmC spike-in DNA standard allows to accurately assess 5-hmC conversion efficiency

An overall conversion efficiency of 5-hmC to uracil of >85% has been considered sufficient in order to accurately quantify 5-mC and 5-hmC bases [11]. Conversion efficiencies below this threshold are a sign of either incomplete or incorrect oxidation and/or bisulfite conversion of the DNA target sample. We therefore estimated the oxidative bisulfite conversion efficiency for 5-hmC based on three independent experiments using the aforementioned 5-hmC spike-in DNA standard in conjunction with the protocol described here. The 5-hmC conversion efficiency, which was considered equal to the measureable 5-hmC content of the fully hydroxymethylated standard, was estimated by quantitative subtraction of the CpG readouts after both bisulfite- and oxidative bisulfite pyrosequencing. On average, a value of 97.00 ± 1.90% and 8.73 ± 2.02% were obtained for each of the 4 CpG sites after bisulfite pyrosequencing and oxidative bisulfite pyrosequencing, respectively. Thus, we estimated, by using this fully hydroxymethylated DNA standard, that the measurable 5-hmC content, and therefore the average 5-hmC conversion efficiency, was 88.27 ± 2.56% per CpG

site (Figure 3). These data indicate that the manually designed spike-in 5-hmC DNA standard can be successfully applied in order to assess the conversion efficiency of 5-hmC during oxidative bisulfite pyrosequencing. Furthermore, based on the achieved overall conversion efficiency of 5-hmC to uracil, we demonstrate that the protocol described here can be considered suitable in order to accurately quantify and estimate 5-mC and 5-hmC bases, respectively.

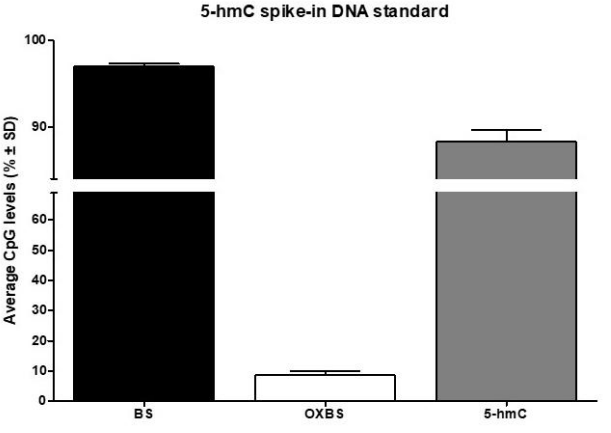


Figure 3. Average cytosine-phosphate-guanine (CpG) levels of the 5-hydroxymethylcytosine (5-hmC) spike-in DNA standard. The average CpG readout (% ± standard deviation (SD)), measured over 4 CpG sites on the 5-hmC spike-in DNA standard in three independent experiments, is displayed after both bisulfite (BS) and oxidative bisulfite (OXBS) pyrosequencing. On average, a readout of 97 ± 1.9% and 8.73 ± 2.02% were obtained for each of the 4 CpG sites after bisulfite- and oxidative bisulfite pyrosequencing, respectively. The measurable 5-hmC content, which was obtained by quantitative subtraction of both aforementioned readouts, was estimated to be 88.27 ± 2.56%. The 5-hmC conversion efficiency was considered proportional to the levels of quantifiable 5-hmC on the standard.

Proper conversion of 5-fC to uracil requires extended bisulfite times

Previous studies have demonstrated that 5-fC converts at a slower rate to a uracil when compared to 5-uC and, therefore, longer incubation times when it comes to bisulfite treatment are generally required than those recommended by most bisulfite kits [11]. Importantly, incomplete conversion of 5-fC to uracil affects the accurate quantification of 5-mC and 5-hmC, since non-converted 5-fC bases are read as a cytosine and not as a thymine after PCR amplification and subsequent pyrosequencing. For this reason, a lower conversion efficiency of 5-fC to uracil would cause an underestimation of the actual number of 5-fC bases, which consequently results in an overestimation of 5-mC residues and an underestimation of the 5-hmC content. In order to address this notion, we examined the effects of both the recommended bisulfite incubation period from the kit and a double incubation period, in parallel, on the levels of measureable 5-hmC when using the fully hydroxymethylated spike-in DNA standard. We observed that the shorter suggested bisulfite incubation time resulted in an

average conversion efficiency of 67.48 ± 3.40% per CpG (data not shown), which is substantially lower when compared to the average efficiency when using the double incubation time (88.27 ± 2.56%). These data therefore confirm that a proper conversion of 5-fC to uracil requires an extended bisulfite conversion time, *i.e.* at least a double incubation period, when compared to the time recommended by the bisulfite conversion kit used here.

Table 1. Modification levels of *OXT* in the middle temporal gyrus (MTG) and induced pluripotent stem cells (iPSCs).

OXT									
MTG									
CpG site	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean ± SD**
% 5-uC *(100%-BS)*	23.69	24.41	30.70	53.94	31.18	59.10	62.58	34.29	39.99 ± 15.92
% 5-mC *(OXBS)*	50.38	56.55	48.90	35.09	42.57	22.38	25.64	38.58	40.01 ± 12.01
% 5-hmC *(BS-OXBS)*	25.93	19.04	20.40	10.97	26.25	18.52	11.78	27.13	20.00 ± 6.30
% 5-mC+5-hmC *(BS)*	76.31	75.59	69.30	46.06	68.82	40.90	37.42	65.71	60.01 ± 15.92
iPSCs									
CpG site	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean ± SD**
% 5-uC *(100%-BS)*	33.53	27.24	29.06	39.21	37.01	51.86	63.61	50.82	41.54 ± 12.71
% 5-mC *(OXBS)*	57.48	70.04	66.29	58.36	57.38	44.10	34.64	41.20	53.69 ± 12.46
% 5-hmC *(BS-OXBS)*	8.99	2.72	4.65	2.43	5.61	4.04	1.75	7.98	4.77 ± 2.62
% 5-mC+5-hmC *(BS)*	66.47	72.76	70.94	60.79	62.99	48.14	36.39	49.18	58.46 ± 12.71

Displayed are the readouts of unmodified cytosine (5-uC), 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) for the eight cytosine-phosphate-guanine (CpG) sites, as well as the mean ± standard deviation (SD), in *OXT* in both the middle temporal gyrus (MTG) and the induced pluripotent stem cells (iPSCs) after oxidative bisulfite pyrosequencing. 5-uC values were calculated by subtracting the bisulfite (BS) signals from 100% (100%-BS), 5-mC values were considered equal to the oxidative bisulfite (OXBS) signal and 5-hmC values were calculated by the subtraction of the oxidative bisulfite signal from the bisulfite signal (BS-OXBS).

Oxidative bisulfite pyrosequencing allows for targeted detection of 5-uC, 5-mC and 5-hmC

Levels of 5-uC, 5-mC and 5-hmC can be found in many tissues and cell types, although with diverse levels of abundance, and with 5-hmC being specifically enriched in brain cells and stem cells [6, 7]. For this reason, and to demonstrate that the oxidative bisulfite pyrosequencing protocol described here can be successfully applied for the targeted detection of all three modifications in various biosamples, levels of each of these marks were assessed across eight individual

OXT and four individual *DNAJB13* CpG sites in human DNA isolated from both post-mortem MTG tissue and cultured iPSCs. For both samples, 5-uC values per CpG were calculated by subtracting the bisulfite signals from 100%, 5-mC values were derived from the oxidative bisulfite signal and 5-hmC values were calculated by subtracting the oxidative bisulfite signal from the bisulfite signal [2]. The levels of these modifications obtained for all the CpG sites in the sequenced region were then averaged per gene (for individual values per CpG site, please see Tables 1 and 2).

Table 2. Modification levels of *DNAJB13* in the middle temporal gyrus (MTG) and induced pluripotent stem cells (iPSCs).

DNAJB13					
MTG					
CpG site	**1**	**2**	**3**	**4**	**Mean ± SD**
% 5-uC *(100%-BS)*	60.78	70.32	47.97	38.45	54.38 ± 14.02
% 5-mC *(OXBS)*	29.72	23.24	44.21	53.45	37.66 ± 13.70
% 5-hmC *(BS-OXBS)*	9.50	6.44	7.82	8.10	7.97 ± 1.25
% 5-mC+5-hmC *(BS)*	39.22	29.68	52.03	61.55	45.62 ± 14.02
iPSCs					
CpG site	**1**	**2**	**3**	**4**	**Mean ± SD**
% 5-uC *(100%-BS)*	31.66	35.10	30.02	25.36	30.54 ± 4.05
% 5-mC *(OXBS)*	68.87	64.42	71.12	73.49	69.48 ± 3.86
% 5-hmC *(BS-OXBS)*	-0.53	0.48	-1.14	1.15	-0.01 ± 1.02
% 5-mC+5-hmC *(BS)*	68.34	64.90	69.98	74.64	69.47 ± 4.05

Displayed are the readouts of unmodified cytosine (5-uC), 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) for the eight cytosine-phosphate-guanine (CpG) sites, as well as the mean ± standard deviation (SD), in *DNAJB13* in both the middle temporal gyrus (MTG) and the induced pluripotent stem cells (iPSCs) after oxidative bisulfite pyrosequencing. 5-uC values were calculated by subtracting the bisulfite (BS) signals from 100% (100%-BS), 5-mC values were considered equal to the oxidative bisulfite (OXBS) signal and 5-hmC values were calculated by the subtraction of the oxidative bisulfite signal from the bisulfite signal (BS-OXBS).

For all 8 CpG sites analyzed within the *OXT* gene, the presence of all three cytosine states was confirmed in both post-mortem MTG tissue and cultured iPSCs (Table 1). In more detail, for 5-uC, 5-mC and 5-hmC, average levels of 39.99 ± 15.92%, 40.01 ± 12.01% and 20.00 ± 6.30%, respectively, were detected for the MTG, whereas for the iPSCs, average levels of 41.54 ± 12.71%, 53.69 ± 12.46% and 4.77 ± 2.62%, respectively, were found. For *DNAJB13,* the presence

of all three modifications was found for all four CpG sites analyzed in the MTG tissue. However, whereas in the iPSCs levels of 5-uC and 5-mC were detected, 5-hmC was either absent or only present at very low levels (Table 2). More specifically, in the MTG, average levels of 54.38 ± 14.02%, 37.66 ± 13.70% and 7.97 ± 1.25% were found for 5-uC, 5-mC and 5-hmC, respectively. In the iPSCs, average levels of 30.54 ± 4.05%, 69.48 ± 3.86 and -0.01% ± 1.02% were found for 5-uC, 5-mC and 5-hmC, respectively (see Table 2 for individual values per CpG site).

Table 3. 5-hydroxymethylcytosine (5-hmC) conversion efficiency in the middle temporal gyrus (MTG) and induced pluripotent stem cells (iPSCs).

5-hmC spike-in DNA standard					
MTG + 5-hmC standard					
CpG site	**1**	**2**	**3**	**4**	**Mean ± SD**
% Non-oxidized 5-hmC *(BS)*	96.48	96.81	100.00	95.63	97.23 ± 1.91
% Not converted 5-hmC *(OXBS)*	7.35	5.91	8.24	5.63	6.78 ± 1.23
% Measureable 5-hmC *(BS-OXBS)*	89.13	90.90	91.76	90.00	90.45 ± 1.13
iPSCs + 5-hmC standard					
CpG site	**1**	**2**	**3**	**4**	**Mean ± SD**
% Non-oxidized 5-hmC *(BS)*	96.21	96.68	100.00	96.09	97.25 ± 1.85
% Not converted 5-hmC *(OXBS)*	7.89	5.50	8.07	5.24	6.68 ± 1.51
% Measureable 5-hmC *(BS-OXBS)*	88.32	91.18	91.93	90.85	90.57 ± 1.57
5-hmC standard only					
CpG site	**1**	**2**	**3**	**4**	**Mean ± SD**
% Non-oxidized 5-hmC *(BS)*	96.62	96.85	100.00	96.85	97.58 ± 1.62
% Not converted 5-hmC *(OXBS)*	7.76	5.89	7.41	6.15	6.80 ± 0.92
% Measureable 5-hmC *(BS-OXBS)*	88.86	90.96	92.59	90.70	90.78 ± 1.53

Displayed are the readouts for the four cytosine-phosphate-guanine (CpG) sites, as well as the mean ± standard deviation (SD), in the 5-hmC spike-in DNA standard after oxidative bisulfite pyrosequencing in both the middle temporal gyrus (MTG) and the induced pluripotent stem cells (iPSCs). The DNA derived from the MTG and iPSCs that were spiked with the standard, as well as a reaction solely containing the 5-hmC standard and no genomic target DNA, were analyzed. The percentage of non-oxidized 5-hmC residues was considered equal to CpG values of the bisulfite (BS) converted fractions, whereas the percentage of not converted 5-hmC bases was considered equal to the CpG values of the oxidative bisulfite (OXBS) fractions. The percentage of measureable 5-hmC was obtained by quantitative subtraction of the OXBS- from the BS data (BS-OXBS).

The spike-in DNA standard does not interfere with the analysis of the accompanying sample

The data presented above was obtained from MTG- and iPSC-derived target DNA that was spiked with the 5-hmC DNA standard before conducting oxidative bisulfite pyrosequencing, which allowed us to assess the conversion efficiency of 5-hmC in each of the individual reactions. During the procedure, we also analyzed a reaction with solely the 5-hmC DNA standard and non-genomic target DNA in parallel, in order to demonstrate that a maximum conversion efficiency of 5-hmC can still be achieved after spiking the target DNA with the standard. After conducting the oxidative bisulfite pyrosequencing protocol on all three samples, *i.e.* MTG + 5-hmC standard, iPSC + 5-hmC standard and the 5-hmC standard only, we then calculated and compared the 5-hmC conversion efficiency for each of these reactions. Based on the average CpG readouts from the 5-hmC spike-in DNA standard, we obtained a conversion efficiency of 90.45 ± 1.13%, 90.57 ± 1.57% and 90.78 ± 1.53% for the MTG + 5-hmC standard, iPSC + 5-hmC standard, and the 5-hmC standard only reactions, respectively (Table 3). These data demonstrate that no striking differences could be identified between the different reactions and confirmed that a maximum 5-hmC conversion efficiency, similar to what we demonstrated earlier, could still be achieved after spiking the target DNA with the standard.

Table 4. Error-rates for *OXT* between paired measurement of spiked and non-spiked DNA target samples

OXT									
MTG vs. MTG + 5-hmC standard									
CpG	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean ± SD**
% Error BS	3.53	3.04	3.33	2.28	2.84	5.22	0.25	1.77	2.78 ± 1.44
% Error OXBS	3.46	2.05	3.15	4.26	0.00	2.36	0.08	3.95	2.41 ± 1.64
								Total	2.60 ± 1.50
iPSCs vs. iPSCs + 5-hmC standard									
CpG	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean ± SD**
% Error BS	4.10	3.39	1.57	0.33	3.21	1.44	2.32	0.21	2.07 ± 1.43
% Error OXBS	0.56	3.86	0.19	0.60	2.69	4.37	2.31	1.00	1.95 ± 1.60
								Total	2.01 ± 1.47

Shown are the absolute differences (% Error) for each individual cytosine-phosphate-guanine (CpG) site in *OXT*, as well as for the mean ± standard deviation (SD) over the entire sequenced region, between the paired oxidative bisulfite measurements in the spiked and non-spiked reactions per tissue. DNA derived the middle temporal gyrus (MTG) and the induced pluripotent stem cells (iPSCs) was analyzed with or without spiking the sample with the 5-hydroxymethylcytosine (5-hmC) standard, and CpG readouts for *OXT* were then compared between the paired bisulfite (BS) and oxidative bisulfite (OXBS) converted fractions per tissue.

Table 5. Error-rates for DNAJB13 between paired measurement of spiked and non-spiked DNA target samples

DNAJB13					
MTG vs. MTG + 5-hmC standard					
CpG	**1**	**2**	**3**	**4**	**Mean ± SD**
% Error BS	1.90	2.13	4.41	3.11	2.89 ± 1.14
% Error OXBS	1.57	1.99	5.78	1.30	2.66 ± 2.10
				Total	2.77 ± 1.57
iPSCs vs. iPSCs + 5-hmC standard					
CpG	**1**	**2**	**3**	**4**	**Mean ± SD**
% Error BS	3.41	4.94	5.29	3.48	4.28 ± 0.98
% Error OXBS	4.31	2.87	6.68	2.00	3.97 ± 2.05
				Total	4.12 ± 1.49

Shown are the absolute differences (% Error) for each individual cytosine-phosphate-guanine (CpG) site in DNAJB13, as well as for the mean ± standard deviation (SD) over the entire sequenced region, between the paired oxidative bisulfite measurements in the spiked and non-spiked reactions per tissue. DNA derived the middle temporal gyrus (MTG) and the induced pluripotent stem cells (iPSCs) was analyzed with or without spiking the sample with the 5-hydroxymethylcytosine (5-hmC) standard, and CpG readouts for DNAJB13 were then compared between the paired bisulfite (BS) and oxidative bisulfite (OXBS) converted fractions per tissue.

To investigate whether spiking of genomic target DNA with the manually designed DNA standard does not interfere with the analysis of the accompanying target sample, we also examined the bisulfite and oxidative bisulfite measurements for *OXT* and *DNAJB13* in MTG- and iPSC-derived DNA without spiking the reactions with the 5-hmC DNA standard. After conducting the oxidative bisulfite pyrosequencing protocol on all four samples, *i.e.* MTG + 5-hmC standard, iPSCs + 5-hmC standard, MTG only, and iPSCs only, we then assessed the error rate between the CpG values for the aforementioned genes in both the bisulfite- and oxidative bisulfite fractions. The error was expressed as the absolute difference in percentage between the bisulfite and oxidative bisulfite signals for each individual CpG site, as well as for the mean over the entire sequenced region, between the paired measurements in the spiked and non-spiked reactions per tissue. For *OXT*, an average CpG error rate of 2.60 ± 1.50% in both the bisulfite and oxidative bisulfite fractions of the MTG samples was found, whereas an error rate of 2.01 ± 1.47% was observed in the iPSCs (Table 4). For *DNAJB13*, on the other hand, an average CpG error rate of 2.77 ± 1.57% in both the bisulfite and oxidative bisulfite fractions of the MTG samples was found, whereas a slightly higher error rate of 4.12 ± 1.49% was observed in the iPSCs (Table 5). Measured over both genes and samples, we therefore estimated the overall error rate to be 2.69 ± 1.62%. Based on previous studies assessing the precision of repetitive measures using the same pyrosequencing assays, we do not consider these

error rates as strikingly different for any of the genes that were analyzed in both the post-mortem brain tissues and cultured iPSCs [12]. These data therefore demonstrate that our spike-in control could be used as an internal pyrosequencing control that did not interfere with the analysis of the accompanying sample.

Discussion

In the present study, we introduced an oxidative bisulfite pyrosequencing approach that can be used for the targeted detection of 5-mC and 5-hmC, as well as 5-uC, levels at individual CpG sites in a predetermined region of a target gene. The approach described here is based on chemical oxidation of genomic DNA by $KRuO_4$ that is applied prior to bisulfite treatment, PCR and pyrosequencing. Recent studies have demonstrated that this oxidant specifically converts 5-hmC to 5-fC bases [5, 11], which consequently allows for the discrimination between 5-mC and 5-hmC using downstream bisulfite-based techniques. More specifically, while after classical bisulfite treatment and PCR both 5-mC and 5-hmC are read as a cytosine, 5-fC bases are read as a thymine at the pyrosequencing step [5]. Thus, when genomic target DNA is treated with $KRuO_4$ prior to bisulfite treatment, the only base that reads as a cytosine is 5-mC. For this reason, oxidative bisulfite pyrosequencing provides a more accurate readout of the levels of 5-mC as compared to classical bisulfite pyrosequencing, while at the same time, it allows the user to estimate levels of 5-hmC after quantitative subtraction of the bisulfite and oxidative bisulfite data obtained from the same DNA sample. Additionally, levels of 5-uC at single CpG sites in a pool of DNA can also be obtained by subtraction of the bisulfite signal from 100%.

In order to demonstrate the efficacy of the approach described here, we developed three spike-in DNA standards, each containing four CpG sites with either 5-uC, 5-mC or 5-hmC bases. All three standards were then subjected to both the classical bisulfite and novel oxidative bisulfite pyrosequencing protocol described here. We confirmed that $KRuO_4$ successfully converts 5-hmC to 5-fC, while leaving 5-uC and 5-mC unaffected. This was demonstrated by a pyrogram readout of a thymine at each of the four CpG sites after oxidative bisulfite pyrosequencing of the 5-hmC standard, which was opposite to the CpG readouts of a cytosine after classical bisulfite pyrosequencing of the same standard. The readouts of all four CpG sites in the pyrograms of both the 5-uC and 5-mC standards, on the other hand, were identical after bisulfite- and oxidative bisulfite sequencing, *i.e.* thymines and cytosines for 5-uC and 5-mC, respectively. In our hands, we obtained an overall conversion efficiency of >88% per CpG site, which is similar to the data obtained by Booth *et al.* using Sanger sequencing [11]. Generally, a conversion efficiency >85% has been considered sufficient, while an

efficiency below this threshold is a sign of either incomplete or ineffective oxidation and/or bisulfite conversion [5]. Factors that could affect the oxidation and bisulfite conversion include, but are not limited to, incorrect preparation of the reagents, improperly denaturated DNA, precipitation of the oxidant and/or too short a time for bisulfite incubation.

In order to demonstrate that the aforementioned factors such as an insufficient bisulfite incubation time can negatively influence the 5-hmC conversion efficiency, we furthermore tested the effect of two different incubation times on the quantifiable 5-hmC content of the standard. The 5-hmC conversion efficiency is proportional to the levels of quantifiable 5-hmC on the standard, as non-converted residues will result in a false base calling for 5-hmC, hence resulting in lower estimated 5-hmC levels after subtraction of the bisulfite and oxidative bisulfite data of the standard. In relation to the bisulfite incubation time, it is know that 5-fC converts to uracil at a slower rate when compared to 5-uC [11]. Due to this difference, longer bisulfite incubation times than those recommended by most manufacturers are generally required to fully convert 5-fC to uracil after the oxidation of 5-hmC. Indeed, the normal incubation time suggested by the kit resulted in a substantially lower 5-hmC conversion efficiency, *i.e.* 67.48 ± 3.40%, when compared to the longer (double) incubation period used (>88%). In order to maximize conversion of 5-fC to uracil, we therefore recommend longer incubation periods. Even though we did not directly test the effect on the final converted DNA yield, previous studies have shown that little effects on DNA loss are observed with increasing incubation times [5, 11].

After validating and assessing the efficiency of the approach based on our manually designed DNA standards, we then set out to explore whether we could detect levels of 5-uC, 5-mC and 5-hmC in DNA derived from both post-mortem MTG tissue and iPSCs. Although levels for each of the three modifications have been found in all bodily tissues, 5-hmC is specifically enriched in the central nervous system and in stem cells, with the highest levels in certain neurons of the brain [6, 7]. As a proof of principle, we quantified the levels of these modifications in two previously established pyrosequencing assays for *OXT* and *DNAJB13*, covering 8 and 4 CpG sites in these genes, respectively. We were able to identify levels of all three modifications at all the CpG sites analyzed within the *OXT* gene in both post-mortem brain tissue and iPSCs. For *DNAJB13*, levels of all three modifications for the 4 CpG sites were detected in the brain tissue. However, in DNA derived from iPSCs, 5-hmC was either absent or present at very low levels.

Interestingly, for two CpG sites in *DNAJB13* in the iPSCs, as well as for the mean over the entire sequenced region, negative values were obtained when calculating the levels of 5-hmC. The procedure used for estimating 5-hmC levels, *i.e.* the quantitative subtraction of the bisulfite- and oxidative bisulfite pyrosequencing data, can indeed yield negative values [13], a phenomenon that is commonly observed in oxidative bisulfite data due to random assay noise or other technological variation linked to the paired measurements. Along similar lines, the sum of both 5-uC and 5-mC levels can therefore surpass 100%, as values for both these modifications are obtained via either the bisulfite- or oxidative bisulfite measurements. For any given pyrosequencing dataset, a statistical approach or the use of false discovery thresholds could be applied in order to correct for technical variation. Similar approaches are currently also being applied in EWAS, *e.g.* during the analysis of oxidative bisulfite datasets obtained using Illumina BeadArray technology [11, 13, 14]. Overall, based on our findings, we could confirm that oxidative bisulfite pyrosequencing allows for a rigorous detection of 5-uC, 5-mC and 5-hmC in various biosamples, including post-mortem brain tissue and iPSCs.

Finally, by spiking genomic DNA samples derived from iPSCs and post-mortem MTG tissue with our manually designed 5-hmC DNA standard, we were able to demonstrate that our standard can be used as an internal pyrosequencing control in order to accurately determine the conversion efficiency of 5-hmC for each of the individual reactions. Importantly, the spike-in control can be used as an internal pyrosequencing control that does not interfere with the analysis of the accompanying sample, and vice versa. As the approach described here is based on the use of $KRuO_4$ for oxidation of the primary alcohol of 5-hmC, it is of utmost importance to accurately assess the oxidation efficiency, as it can easily be affected by contaminants. This includes other primary alcohols or buffers that are often present in the target sample after DNA isolation, such as ethanol, Tris and phosphate [5]. It is therefore also essential that the DNA sample used for the purpose of oxidative bisulfite pyrosequencing is of high quality and purity, as well as dissolved in Milli-Q. Additionally, the color of the oxidation reaction is indicative of successful oxidation and should remain orange after the oxidation step. Deviating colors, such as green or black, suggest the presence of contaminants in the reaction that lead to decomposition of the oxidant. Please refer to the publication by Booth *et al.* [5] for an illustration on color changes as a result of side reactions due to contaminants. Importantly, DNA in contaminated reactions with partial decomposition of the oxidant might still be fully oxidized and therefore analyzable. We therefore highly recommend to spike each individual reaction with the internal pyrosequencing control and to assess the 5-hmC conversion efficiency in order to exclude that the DNA oxidation was sub-optimal.

To conclude, the strategy presented in this manuscript can be applied to quantify levels of 5-uC, 5-mC and 5-hmC at any DNA sequence of interest and only requires the establishment of a qualitative pyrosequencing assay, as well as isolated target DNA, prior to conducting the protocol. The manually designed DNA standards furthermore enable assessing whether the approach has been applied successfully, thereby adding validity to the findings. Future pyrosequencing studies based on similar strategies will allow researchers to overcome the issue of modification specificity as previously observed in neuroepigenetic studies. Certainly, to date, this issue has been most prominent in studies targeting the human brain, regardless of the platform, but should in the future not be disregarded in studies using bodily tissues with lower known levels of 5-hmC.

Acknowledgments

Funds have been provided by the Joint Program— Neurodegenerative Disease Research (JPND) for the EPI-AD consortium (http://www.epi-ad.eu/). The project is supported through the following funding organizations under the aegis of JPND; the Netherlands, The Netherlands Organization for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 643417.

References

1. Delaney C, Garg SK, Yung R. Analysis of DNA Methylation by Pyrosequencing. Methods Mol Biol. 2015;1343:249-64.
2. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenetics. 2019;11(1):52.
3. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.
4. Smith AR, Smith RG, Condliffe D, Hannon E, Schalkwyk L, Mill J, et al. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. Neurobiology of aging. 2016;47:35-40.
5. Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841-51.
6. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929-30.
7. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic acids research. 2010;38(19):e181-e.

8. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic acids research. 2005;33(18):5868-77.
9. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome biology. 2016;17(1):1-17.
10. Sullivan S, Stacey GN, Akazawa C, Aoyama N, Baptista R, Bedford P, et al. Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. Regenerative medicine. 2018;13(7):859-66.
11. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012;336(6083):934-7.
12. Crary-Dooley FK, Tam ME, Dunaway KW, Hertz-Picciotto I, Schmidt RJ, LaSalle JM. A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. Epigenetics. 2017;12(3):206-14.
13. Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S, et al. oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. Methods. 2015;72:9-15.
14. Slynko A, Benner A. Statistical methods for classification of 5-hmC levels based on the Illumina Inifinium HumanMethylation450 (450k) array data, under the paired bisulfite (BS) and oxidative bisulfite (oxBS) treatment. PloS one. 2019;14(6):e0218103.

Supplementary material

Oxidative bisulfite conversion protocol for pyrosequencing

Reagents

- Genomic target DNA (>1 µg)
- Milli-Q water
- Sodium hydroxide (NaOH; Sigma Aldrich, cat. no. S5881)
- Potassium perruthenate ($KRuO_4$, cat. no. 334537)
- FastStart™ Taq DNA Polymerase (Sigma Aldrich, cat. no. 12032929001)
- EZ DNA Methylation-Direct Kit (Zymo Research, cat. no. D5020)
- Micro Bio-Spin™ P-6 Gel Columns, SSC Buffer (Bio-Rad, cat. no. 7326200)
- Agarose
- Tris base, acetic acid and EDTA (TAE) buffer
- Ethanol (EtOH)
- DNA loading dye, *e.g.* Orange G DNA Loading Dye (ThermoFisher Scientific, cat. no. R0631)
- Nucleic acid gel stain, *e.g.* GelsGelStar™ Nucleic Acid Gel Stain (Lonza,cat. no. 50535)
- Ultra Low Range DNA Ladder (ThermoFischer Scientific, Invitrogen, cat. no. 10597012)
- 5-hmC Spike-in DNA standard and primers (Riemens *et al.* 2021)
- Qubit™ dsDNA BR Assay Kit (Invitrogen, cat. no Q32850)
- Qubit™ Assay Tubes (Invitrogen, cat. no. Q32856)
- PyroMark Q48 AutoPrep Starter Kit (Qiagen, cat. no. 974230)

Equipment

- Manual pipettes
- Pipette filtered tips
- Thermomixer
- Thermal cycler
- Microcentrifuge with PCR tube adaptors
- Eppendorf Tubes, 1,5 mL (Eppendorf, cat. no. 0030120086)
- Thin-walled PCR tubes, 0.2 ml
- Pyrosequencer Q48 Autoprep (Qiagen, cat. no. 9002470)
- Gel electrophoresis equipment
- Qubit fluorimeter (Invitrogen, cat. no. Q33238)

Reagent preparations

NaOH 1M solution

1. Dissolve 3.99 g NaOH pellets in 60 mL Milli-Q
2. Bring the volume up to 100 mL with Milli-Q in order to obtain a 1 M solution
3. Store the solution at room temperature up to several months

$KRuO_4$ solution (10X) in 0.5M NaOH

1. Dilute 3 mL of 1 M NaOH solution 1:1 with Milli-Q to obtain a 0.5 M NaOH solution
2. Dissolve 61.25 mg $KRuO_4$ in 1 mL 0.5 M NaOH
3. Bring the volume up to 2 mL with 0.5 M NaOH in order to obtain a 10X $KRuO_4$ stock solution
4. Aliquot the stock solution and store at -20°C up to several months

CT Conversion Reagent (EZ DNA Methylation-DirectTM Kit; D5020)

The CT Conversion Reagent provided with the EZ DNA Methylation-DirectTM Kit (D5020) is a solid mixture and must be prepared prior to first use. Please refer to the protocol provided with the kit for further details.

1. Add 790 µL of M-Solubilization Buffer to a tube of CT Conversion Reagent
2. Add 300 µL of M-Dilution Buffer to the tube of CT Conversion Reagent
3. Mix at room temperature with frequent vortexing or shaking for 10 minutes
4. Add 160 µL of M-Reaction Buffer to the tube of CT Conversion Reagent
5. Mix an additional 1 minute

M-Wash Buffer

The M-Wash Buffer provided with the EZ DNA Methylation-DirectTM Kit (D5020) must be prepared prior to first use. Please refer to the protocol provided with the kit for further details.

1. Add 24 mL of 100% ethanol to the 6 mL M-Wash Buffer concentrate (D5020)

Procedure

Replacing SSC buffer with Milli-Q in Micro Bio-Spin columns

The SSC buffer in the columns has to be replaced with Milli-Q prior to use. Please refer to the protocol provided with the columns for further details.

1. Remove the cap of a Micro Bio-Spin column
2. Allow the buffer to drain by gravity
3. Spin Micro Bio-Spin column 1 minute at 1000 x g

4. Discard eluate
5. Add 500 µL Milli-Q to the Bio-Spin column
6. Spin 1 minute at 1000 x g
7. Discard eluate
8. Add 500 µL Milli-Q to the Bio-Spin column
9. Spin 1 minute at 1000 x g
10. Discard eluate
11. Add 500 µL Milli-Q to the Bio-Spin column
12. Spin 1 minute at 1000 x g
13. Discard eluate
14. Add 500 µL Milli-Q to the Bio-Spin column
15. Spin 4 minutes at 1000 x g
16. Discard eluate

DNA denaturation and oxidation

1. Place the Micro Bio-Spin column from previous step in a 1.5 mL Eppendorf tube
2. Add the genomic target DNA to the matrix of the Micro Bio-Spin column
3. Spin 4 minutes at 1000 x g
4. Measure the DNA concentration using the Qubit fluorimeter
5. Split the sample into two tubes (500 ng DNA per tube; volume should be ≤16.1 µL per tube)
6. Add 1 µL of 5-hmC Spike-in DNA standard to both tubes
7. Top up each fraction to 17.1 uL using Milli-Q
8. Add 0.9 µL NaOH (1M) to each fraction
9. Vortex and then centrifuge both tubes briefly
10. Incubate for 30 minutes in a heating block at 37°C shaking at 1000 rpm
11. Thaw 10X oxidant solution on ice
12. Dilute oxidant to 1X using Milli-Q
13. Remove denatured DNA from the heating block
14. Add 2 µL of 1X oxidant to the oxidative bisulfite fraction, use 2 µL Milli-Q for the bisulfite fraction
15. Vortex and then centrifuge both tubes briefly
16. Incubate at 40°C for 10 minutes shaking at 1000 rpm

IMPORTANT: The color of the oxidized sample should remain orange after the oxidation step. Deviating colors, such as, green or black, suggest the presence of contaminants in the reaction that lead to decomposition of the oxidant. DNA in contaminated reactions with partial decomposition of the oxidant might still be fully oxidized and therefore analyzable. We therefore highly recommend to spike each individual reaction with the internal pyrosequencing control and to assess

the 5-hmC conversion efficiency in order to exclude that the DNA oxidation was sub-optimal.

Pre-bisulfite conversion clean-up

The SSC buffer in the columns has to be replaced with Milli-Q during the incubation periods of the previous step in the protocol in order to allow a direct clean-up after the DNA oxidation.

1. Replace the SSC buffer of two Micro Bio-Spin columns with Milli-Q as described earlier
2. Place each of the Micro Bio-Spin columns in a 1.5 mL Eppendorf tube
3. Add both samples from previous step (20 uL) to the matrix of a separate Micro Bio-Spin column
4. Spin 4 minutes at 1000 x g

Bisulfite conversion (EZ DNA Methylation-DirectTM Kit; D5020)

Although we expect that any bisulfite conversion kit can be used in combination with an appropriate bisulfite conversion time, for the present manuscript we optimized the protocol using the EZ DNA Methylation-Direct™ Kit D5020. Adequate oxidative bisulfite conversion using other kits can be assessed using the 5-hmC spike-in DNA standards.

1. Add each sample from the previous step to 130 µL of CT Conversion Reagent solution in a PCR tube
2. Mix the sample and then centrifuge briefly
3. Place the PCR tube(s) in a thermal cycler and perform the following steps:
    a. 98°C for 8 minutes
    b. 64°C for 3.5 hours
    c. 98°C for 8 minutes
    d. 64°C for 3.5 hours
    e. 4°C storage
4. Add 600 µL of M-Binding Buffer into a Zymo-Spin™ IC Column and place the column into a provided Collection Tube
5. Load the bisulfite converted sample into the Zymo-Spin™ IC Column containing the M-Binding Buffer
6. Close the cap and mix by inverting the column several times
7. Centrifuge at full speed (>10,000 x g) for 30 seconds
8. Discard the flow-through
9. Add 100 µl of M-Wash Buffer to the column
10. Centrifuge at full speed (>10,000 x g) for 30 seconds
11. Add 200 µl of M-Desulphonation Buffer to the column and let stand at room temperature (20-30°C) for 15-20 minutes

12. Centrifuge at full speed (>10,000 x g) for 30 seconds
13. Add 200 µl of M-Wash Buffer to the column
14. Centrifuge at full speed (>10,000 x g) for 30 seconds
15. Add 200 µl of M-Wash Buffer to the column
16. Centrifuge at full speed (>10,000 x g) for 30 seconds
17. Place the column into a 1.5 ml Eppendorf tube
18. Add 10 µl of M-Elution Buffer directly to the column matrix
19. Centrifuge for 30 seconds at full speed (>10,000 x g) to elute the DNA
20. Add 10 µl of M-Elution Buffer directly to the column matrix
21. Centrifuge for 30 seconds at full speed (>10,000 x g) to elute the DNA

NOTE: The (oxidative) bisulfite converted DNA fractions can be used directly for PCR amplifications or stored for later use at -20°C or -80°C. In order to assess the 5-hmC conversion efficiency, the provided PCR primers, as well as pyrosequencing primer for the 5-hmC Spike-in DNA standard, can be applied for PCR amplifications and subsequent pyrosequencing analysis as described below. For the present manuscript, the FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) was used for PCR amplifications. For pyrosequencing analysis of the obtained PCR product(s), please refer to the manufacturer's instructions provided with the pyrosequencer.

5-hmC conversion efficiency assessment

1. For each oxidative bisulfite converted sample that was spiked with the 5-hmC standard, prepare two PCR tubes containing the following reagents per tube (total volume of 25 µL):
    a. 1 µl of the bisulfite-converted DNA sample (PCR tube 1) and/or 1 µl of the oxidative bisulfite-converted DNA sample (PCR tube 2)
    b. 2.5 µL PCR buffer (10X) with 20 mM MgCl2
    c. 0.5 µL 10 mM dNTP mix
    d. 1 µL of the forward primer (5 µM stock)
    e. 1 µL of the (biotinylated-)reverse primer (5 µM stock)
    f. 0.2 µL (5 U/µL) FastStart™ Taq DNA Polymerase
2. Place the PCR tube(s) in a thermal cycler and perform the following steps:
    a. 95°C for 5 minutes
    b. 95°C for 30 seconds
60°C for 30 seconds
72°C for 15 seconds } x 35 cycles
    c. 72°C for 1 minute
    d. 4°C storage

3. Visualize the PCR products on a 2% agarose gel, using the Orange G DNA Loading Dye, the GelsGelStar™ Nucleic Acid Gel Stain and an Ultra Low Range DNA Ladder according to the manufacturer's instructions. A PCR product derived from the bisulfite and oxidative bisulfite converted fraction should be visible at 54 bp.
4. Use the required amount of PCR product as suggested by the manufacturer's instructions for subsequent pyrosequencing analysis, *i.e.* 10 µL PCR product for The Pyromark Q48 Autoprep system.
5. The 5-hmC conversion efficiency of the spiked target sample can be ascertained by quantitative subtraction of the CpG readouts for the 5-hmC standard after both bisulfite- and oxidative bisulfite pyrosequencing.

Supplementary Table 1. Subject demographics

	MTG
Pathology	Definite AD
Sex	Female
Age	83
Ethnicity/Race	Caucasian
PMI	4.75
APOE	4/4
Braak Stage	VI
Plaque total	14.5
Tangle total	15
	iPSCs
Cell line ID	CARIMi001A
Pathology	No
Sex	Male
Age	Late twenties
Ethnicity/Race	Southern European
Karyotype	Normal
Pluripotency markers	*OCT4*, *NANOG*, *SSEA3*, *SSEA4*, *TRA-1-60* and *TRA-1-81*
Germlayer differentiation	Yes

Middle temporal gyrus (MTG) tissue obtained from the Banner Sun Health Research Institute (BSHRI, Sun City, Arizona, US). Displayed from the subject are the pathology, sex, age, ethnicity/race, postmortem interval (PMI; hours), Apolipoprotein E (*APOE*) genotype, braak stage, total plaque and tangle load, *i.e.* the sum of average amyloid beta (Aβ) plaque and tangle densities, respectively, in the entorhinal cortex, hippocampus, parietal lobe, temporal lobe and frontal lobe cortex. Induced pluripotent stem cells (iPSCs) obtained from Cardiovascular Research Institute Maastricht (CARIM). Displayed from the subject are the cell line identifier (ID), pathology, sex, age, ethnicity/race, karyotype, the pluripotency markers assessed and whether successful differentiation to the three germ layers was achieved with the cell line.

Supplementary Table 2. Overview of the synthetic DNA (syDNA) sequence and associated assay primers

syDNA		
Sequence		
TTTGGGGGTGGAGTAGAGGGATTTGTTCGACGTCGACGGTTAGGAAGTTTAGGGAGGTGAGGGTTT		
PCR primers		
Forward primer (5'-3')	**Reverze primer (5'-3')**	**Product size (bp)**
GGTGGAGTAGAGGGATTTGT	(Bio-)TCACCTCCCTAAACTTCCTAAC	54
Pyrosequencing primers		
Sequencing primer (5'-3')	**Number of CpG sites**	**Sequence region (bp)**
GAGTAGAGGGATTTGTT	4	15 (28-42)

An overview of the syDNA sequence and both the polymerase chain reaction (PCR)- and pyrosequencing primers of the associated assay that were used in this study. Displayed is the syDNA sequence, the PCR primer sequences and the amplicon size in base pairs (bp). The cytosine-phosphate-guanine (CpG) sites and the covered sequencing region are listed for the sequencing primer. For DNA standard preparations, a non-biotinylated version of the reverse primer was used.

Supplementary Table 3. Primer overview *OXT* and *DNAJB13*

PCR primers				
Gene	**Forward primer (5'-3')**	**Reverze primer (5'-3')**	**Target region (GRCh37)**	**Product size (bp)**
OXT	ATGTTTAGGTATAAAAAGGTTAGGT	(Bio-)CTCTTACCTCCCAAAAACAATT	20:3071582-3071747	166
DNAJB13	GTATTTTGGGAGGATGAGTTATAATTG	(Bio-)CCCTCCCTCCTCCAAATAC	11:73957554-73957779	226

Pyrosequencing primers				
Gene	**Sequencing primer (5'-3')**	**Number of CpG sites**	**Sequencing region (GRCh37)**	**PyroMark Orientation**
OXT	TTAGGTAGGAGAGATAGTTATTAG	8	20:3071625-3071709	Upper strand (5'-3')
DNAJB13	TGTTTGAAAATTTTTGTTGAAGA	4	11:73957570-73957615	Lower strand (5'-3')

An overview of the polymerase chain reaction (PCR) and pyrosequencing primers for *OXT* and *DNAJB13* that were used in this study. Displayed are the primer sequences, as well as corresponding genomic coordinates (Ensembl GRCh37 assembly) and amplicon sizes in base pairs (bp) for each of the target genes. The total number of cytosine-phosphate-guanine (CpG) sites, the covered sequencing region and its orientation in the PyroMark assay design software, are listed per gene and sequencing primer.

Chapter 4

Targeted methylation profiling of single laser-capture microdissected post-mortem brain cells by adapted limiting dilution bisulfite pyrosequencing (LDBSP)

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Jennifer Nolz[3], Sonia C. Susano Chaves[1], Diane Duroux[4], Ehsan Pishva[1], Diego Mastroeni[3], Kristel van Steen[4,5], Thomas Haaf[2], Daniël L.A. van den Hove[1,6]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[4]WELBIO, GIGA-R Medical Genomics - BIO3, University of Liège, Liège, Belgium; Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[5]Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[6]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

Abstract
A reoccurring issue in neuroepigenomic studies, especially in the context of neurodegenerative disease, is the use of (heterogeneous) bulk tissue, which generates noise during epigenetic profiling. A workable solution to this issue is to quantify epigenetic patterns in individually isolated neuronal cells using laser capture microdissection (LCM). For this reason, we established a novel approach for targeted DNA methylation profiling of individual genes that relies on a combination of LCM and limiting dilution bisulfite pyrosequencing (LDBSP). Using this approach, we determined cytosine-phosphate-guanine (CpG) methylation rates in a multi-targeted manner on single alleles derived from 50 neurons that were isolated from unfixed post-mortem brain tissue. In the present manuscript, we describe the general pipeline and, as a showcase, demonstrate how targeted methylation analysis of various genes, in this case *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT*, can be performed simultaneously. By doing so, we describe an adapted data analysis pipeline for LDBSP, allowing one to include and correct CpG methylation rates derived from multi-allele reactions. In addition, we show that the efficiency of LDBSP on 50 LCM neurons is similar to the efficiency obtained in previously published studies using this technique on other isolated cell types. Overall, the method described here provides the user with a more accurate estimation of the DNA methylation status of each target gene in the analyzed cell pools, thereby adding further validity to this approach.

Keywords
Limiting dilution bisulfite pyrosequencing, laser-capture microdissection, DNA methylation, single cell, epigenetics.

Introduction

An increasing number of studies has implicated a central role for epigenetic mechanisms such as DNA methylation in the pathophysiology of neurodegenerative disorders, including Alzheimer's disease (AD) [1, 2]. The field of neuroepigenomics, however, still faces many challenges that impede attempts to disentangle the exact contribution of DNA methylation alterations in the development and course of disorders like AD. A central issue is the cellular heterogeneity of the studied bulk tissue samples, which represents a major source of noise in epigenetic profiling [3]. Next to drastically decreasing the signal-to-noise ratio, the use of bulk tissue samples does not allow one to conclude whether differences found between *e.g.* patient and control brains are related to disease status rather than changes in cell type composition [1]. In fact, neurodegenerative diseases like AD are characterized by neuronal loss, concomitant with alterations in the brain cellular architecture [2]. Thus, differences in cellular composition between tissue samples from patients and control cases can lead to misinterpretation of the acquired epigenetic data. Aside from disease-related changes in cellular proportions, differences in cell type composition between different samples that arise from tissue sampling represents another issue introducing bias. Moreover, even in the case of comparable cellular proportions, heterogeneous tissues could mask cell-type specific modifications related to the disease, as changes in one cell type could dilute or oppose changes in another, thereby obscuring important cell subtype-specific aberrations when analyzed together [4].

In order to overcome the issue of cellular heterogeneity in methylome-wide studies, it is possible correct for cell type composition using advanced bioinformatics approaches [5-7]. However, there is an ongoing dispute concerning the validity of this approach [3, 8, 9]. A more accurate alternative, and the only option in the case of targeted candidate gene-based approaches, is to profile epigenetic patterns in individually isolated cells that were selected *e.g.* based on a cell type-specific marker. This latter approach does not only avoid noise that is induced by differences in cell type composition of the studied bulk tissue samples, but it also provides a more detailed overview of epigenetic profiles in individual cell types. In recent years, limiting dilution bisulfite pyrosequencing (LDBSP) has emerged as a cost-effective approach, allowing targeted bisulfite methylation profiling at a single cytosine-phosphate-guanine (CpG) site resolution in a single or a few cells [10]. The principle of this technique relies on an excessive dilution of the bisulfite-treated target DNA obtained from a single cell so that maximum one allele (or none) is present in each of the downstream reactions. Subsequently, each individual DNA molecule is separately amplified by (semi-) nested polymerase chain reactions (PCRs) and

analyzed by means of pyrosequencing. Thus, when applied in conjunction with a highly selective isolation procedure, such as laser-capture microdissection (LCM) based on the immunoreactivity of a cell-specific marker, this technique offers an appealing approach for the detection of methylation patterns in individual cellular populations of the brain.

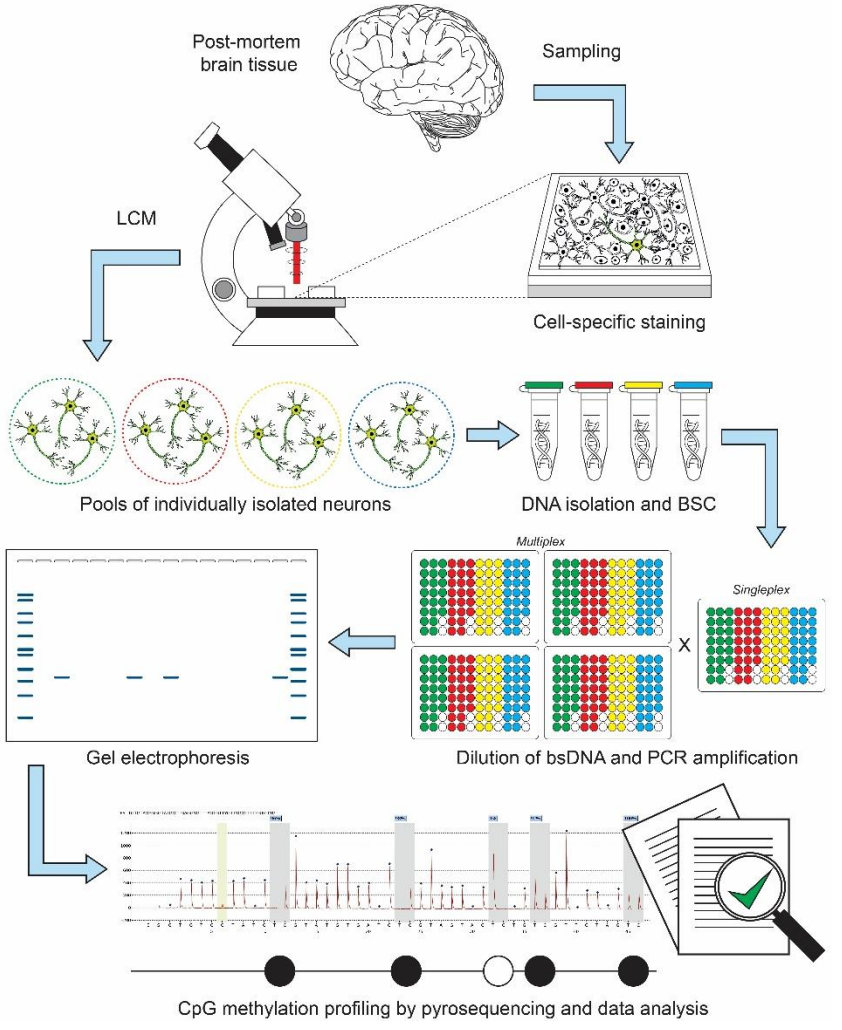


Figure 1. Overview of the limiting dilution bisulfite pyrosequencing (LDBPS) procedure for pools of individually isolated neurons using laser capture microdissection (LCM). Tissue from post-mortem brains is sampled and stained using a cell type-specific marker. Neurons are then isolated and divided in small cell pools. DNA derived from these neurons then undergoes a bisulfite conversion (BSC) treatment and is subsequently diluted to a single allele level. The bisulfite converted DNA (bsDNA) is then amplified twice by means of a (semi-)nested polymerase chain reaction (PCR) in multiplex-singleplex formation. Product-yielding reactions are then visualized on an agarose gel and the cytosine-phosphate-guanine (CpG) methylation status of these products is then profiled by means of pyrosequencing.

Although LDBSP has been shown to be compatible for individual 2-16 cell embryos, single germinal vesicle oocytes and haploid sperms [10, 11], to date, this technique has not been applied on *ex vivo* brain cells. In the present study, we therefore demonstrate, for the first time, that LDBSP can be successfully applied on pools of 50 neurons, isolated by LCM from unfixed post-mortem brain tissue (Figure 1). Here, we describe the general adapted LCM-LDBSP pipeline and, as a showcase, demonstrate how a targeted methylation analysis can be performed in multiple genes (*DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT*) simultaneously. Importantly, diluting a DNA target sample derived from 50 neurons to a single-allele-level will occasionally render reactions with more than one allele, a scenario that does not regularly occur when conducting LDBSP using one or just a few cells. We therefore describe an adapted data analysis pipeline specifically designed for assessing pools of LCM neurons that allow to include and correct the methylation data for multi-allele reactions. This novel approach, which compromises a CpG-site calling procedure, combined with an integrated in-depth analysis of the raw CpG methylation rates, aims at avoiding unintentionally induced bias due to the blunt exclusion of reactions that suggest to contain more than one allele. Overall, the method described here allows the user to more accurately determine the DNA methylation status of the target genes in the analyzed samples, thereby adding further validity to the experimental data acquired. In addition, we show that the efficiency of LDBSP on 50 neurons isolated with LCM from post-mortem brain tissues is similar to the efficiency achieved in previously published studies using this technique on other isolated cell types.

Materials and methods

A step-wise overview of the LDBSP protocol is available in the Supplementary material.

Ethics statement

Written informed consent for brain autopsies was obtained in compliance with institutional guidelines of the Banner Sun Health Research Institute (BSHRI, Sun City, Arizona, US). The Banner Sun Health Research Institute Review Board approved the entire study, including the recruitment, enrollment and autopsy procedures. Each individual and their respective relative(s) consented to brain autopsy for the purpose of scientific research as part of the BSHRI Brain and Body Donation Program (BBDP). The human brain tissue used in this study was derived from routine autopsies, fully qualifying for 4C exemption by the National Institute of Health (NIH) guidelines [12]. All samples were analyzed anonymously throughout the experimental procedures.

Sample collection

Frozen unfixed dorsal raphe nuclei (DRN) tissue from 24 individuals (12 female AD and 12 female age-matched non-demented control cases) were collected at the BSHRI. Brain samples were frozen and stored at -80°C after autopsy, with an average post-mortem interval (PMI) of 2.69±0.82 hours. A final diagnosis of AD or non-demented healthy control was made based on the NIH AD Center criteria [12]. Comorbidity with any other type of dementia, mild cognitive impairment, cerebrovascular disorders and presence of non-microscopic infarcts were applied as exclusion criteria. For demographic and other relevant information about the studied samples, please refer to Supplementary Table 1.

Immunohistochemistry

Frozen DRN tissue sections of 10 µm from were mounted onto polyethylene naphthalate (PEN) slides and fixed in ice-cold 50% acetone/50% ethanol solution for 5 minutes on ice. Sections were washed in ice cold phosphate buffered saline (PBS), blocked in 1% hydrogen peroxide for 2 minutes, followed by 3 quick submersions in ice-cold PBS. Sections were then placed in a dilution of primary antibody against serotonin (5-HT; Abcam, ab66047) in PBS for 10 minutes at room temperature. After the incubation, sections were washed three times in PBS and incubated with avidin-biotin complex in PBS for 10 minutes at room temperature. Next, sections were washed three times in 50 mM Tris buffer and immersed in 3.3'-diaminobenzidine (DAB) solution (9.3 ml 50 mM Tris; 200 µl DAB (5 mg/ml); 500 µl saturated nickel; and 4 µl of 1% hydrogen peroxidase) for 5 minutes, followed by two quick rinses in 50mM Tris to stop the reaction. All sections were stored at -80^{o}C until further processing.

Laser-capture microdissection

5-HT is a monoamine neurotransmitter that is specifically expressed by serotonergic neurons [13]. For this reason, LCM of serotonergic neurons from the DRN sections was performed based on 5-HT-immunoreactivity. In brief, sections were dipped in 100% ethanol, allowed to dry, and loaded onto a Leica AS-LMD LCM microscope (Leica, Wetzlar, Germany). Single serotonergic neurons were cut and then dropped into an inverted microcentrifuge cap containing 10 µl of Tris-EDTA (TE) buffer. Per individual subject, 150 serotonergic neurons were captured at 20X magnification and divided in small pools of 50 cells per microcentrifuge tube, *i.e.* three pools of 50 neurons per subject. All isolated cells were stored at -80°C until further processing.

DNA isolation and sodium bisulfite treatment

Genomic DNA from a pool of 50 neurons was isolated and bisulfite-converted using the EZ DNA Methylation-Direct Kit (Zymo Research, Irvine, CA, USA) with

the following adjustments. In brief, 1 µl of proteinase K (20 µg/µl) and 11 µl of M-Digestion buffer (2X) were added to a microcentrifuge tube containing the cells and incubated overnight at 50°C. Subsequently, the complete lysate was transferred to a PCR tube and 143 µl of bisulfite conversion reagent was used to wash out the digestion tube before adding it to the sample. Bisulfite conversion was performed in a thermal cycler running at 98°C for 8 minutes and then at 64°C for 3.5 hours. A volume of 200 µl binding buffer was added to the spin column before loading the bisulfite-converted sample. The PCR tube used for bisulfite conversion was washed out twice by first adding 200 µl of binding buffer to the tube and then by transferring this volume to the sample-containing column. After centrifugation (10,000 x g; 30 seconds), the column was washed with 100 µl washing buffer, incubated for 15 minutes with 200 µl desulfonation buffer and washed twice again with 200 µl washing buffer. The bisulfite-converted DNA was eluted in a single Eppendorf tube by running 20 µl of elution buffer through the column twice (Two times at 10,000 x g; 30 seconds). Eppendorf LoBind microcentrifuge tubes (Merck KGaA, Darmstadt, Germany) and TipOne Low Retention Tips (STARLAB, Hamburg, Germany) with low affinity for DNA were used throughout the whole procedure. Multiplex PCR amplifications were performed directly after elution of the bisulfite-converted DNA.

Multiplex polymerase chain reaction

All assays were based on a (semi-)nested PCR design and amplified in multiplex-singleplex formation. Primers were designed with the PyroMark Assay Design 2.0 software (Qiagen, Hilden, Germany; see Supplementary Table 2). Bisulfite-treated DNA derived from a pool of 50 neurons was diluted to a single allele level by adding a multiplex PCR mixture with a capacity of 22 individual reactions to the sample (determined empirically). Each individual multiplex PCR reaction made use of 2.5 µl PCR buffer (10X) with 20 mM $MgCl_2$, 0.5 µl 10 mM dNTP mix, 1 µl of each primer (10 µM stock) and 0.2 µl (5 U/µl) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) in a total volume of 25 µl. After adding the bisulfite DNA to the complete mixture, the sample was pipetted up-and-down in order to homogeneously disperse all bisulfite DNA molecules throughout the solution and fractions of 25 µl were divided over 22 wells of a microtiter plate. Multiplex PCRs were then performed with an initial denaturation step at 95°C for 5 minutes, followed by 43 cycles with denaturation at 95°C for 30 seconds, annealing at 56°C for 30 seconds and extension at 72°C for 1 minute, with a final extension step at 72°C for 7 minutes.

Singleplex polymerase chain reaction

For each individual singleplex PCR reaction, 1 µl of the multiplex product was used as a template. In addition, every single singleplex PCR reaction made use

of 2.5 µl PCR buffer (10X) with 20 mM MgCl2, 0.5 µl 10 mM dNTP mix, 1 µl of each primer (10 µM stock) and 0.2 µl (5 U/µl) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) in a total volume of 25 µl. Amplifications for each of the target genes were then performed with an initial denaturation step at 95°C for 5 minutes, followed by 45 cycles with denaturation at 95°C for 30 seconds, annealing at 58°C for 30 seconds and extension at 72°C for 1 minute, with a final extension step at 72°C for 7 minutes. Reactions that yielded a singleplex PCR product- were identified on an agarose gel and 3 µl of the product was utilized per assay for bisulfite pyrosequencing.

Bisulfite pyrosequencing

The PyroMark Q96 MD pyrosequencing system (Qiagen) with the PyroMark Gold Q96 CDT reagent kit (Qiagen, Hilden, Germany) were used according to the manufacturer's instructions. Methylation levels at a single CpG resolution were quantified with the Pyro Q-CpG 1.0.9 software (Qiagen). All assays were tested for their sensitivity using the EpiTect PCR Control DNA Set (Qiagen). For further details on the pyrosequencing assays and sequencing primers, please refer to Supplementary Table 2.

Statistical analysis

All statistical analyses were performed with the IBM SPSS Statistics software version 25. In order to determine whether the estimated CpG methylation rates were significantly different between the novel and the traditional LDBSP data analysis pipeline, a one-way repeated measures multivariate analysis of variance (MANOVA) was performed for each target gene. The multivariate analysis was performed in order to determine whether there was a combined difference in CpG methylation rates measured over all CpG sites, which was followed by a univariate test for each individual site. A p value of <0.05 was considered as statistically significant.

Results

LDBSP allows for methylation profiling of pools of 50 LCM-collected neurons

In order to assess the compatibility of LDBSP on neurons that were individually isolated from unfixed post-mortem brain tissue using LCM, batches consisting of 1, 5, 10 and 50 cells were initially processed following the working procedures described in the present manuscript. Following LDBSP, only pools of 50 cells yielded a favorable amount of sequenceable PCR products for each gene, whereas lower cell numbers generally yielded little to no amplicons (data not shown). Next, 150 neurons per individual from 24 donors were isolated, divided over three pools of 50 cells each and processed for further analysis. In order to

ensure adequate dilution to a single-allele-level, the dilution factor that is applied when conducting LDBSP should be larger than the maximum number of DNA molecules in the starting sample. For pragmatic reasons, at least half of the reactions should therefore not contain a PCR product after the procedure [11]. In our hands, on average, 4.46 PCR products were obtained per gene per pool of 50 neurons, hence a dilution factor of 22 was used. Of note, we did not observe any striking differences in the number of acquired products when increasing the number of dilutions (up to 96x; data not shown). Overall, for *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT*, respectively, 416, 233, 315, 189, 532 and 241 PCR products were obtained (Table 1). All of these were then successfully taken forward for methylation profiling using pyrosequencing.

Table 1. Overview of the limiting dilution bisulfite pyrosequencing (LDBSP) parameters

Singleplex PCR	**DNAJB13**	**PGLYRP1**	**RHBDF2**	**C3**	**LMX1B**	**OXT**	**Mean**
Number (+ percentage) of reactions with PCR product	416 (26.26)	233 (14.71)	315 (19.89)	189 (11.93)	532 (33.59)	241 (15.21)	321 (20.27)
Allele estimation traditional LDBSP criteria							
Number (+ percentage) of included reactions (1 allele)	346 (83.17)	205 (87.98)	298 (94.60)	186 (98.41)	474 (89.10)	221 (91.70)	288.33 (89.82)
Number (+ percentage) of excluded reactions (multi-allele/artifact)	70 (16.83)	28 (12.02)	17 (5.40)	3 (1.59)	58 (10.90)	20 (8.30)	32.67 (10.18)
Allele estimation novel LDBSP criteria							
Number (+ percentage) of reactions with 1 allele	356 (85.58)	208 (89.27)	300 (95.24)	186 (98.41)	475 (89.29)	232 (96.27)	292.83 (91.23)
Number (+ percentage) of reactions with 2 alleles	23 (5.53)	4 (1.72)	7 (2.22)	0 (0.00)	21 (3.95)	3 (1.24)	9.67 (3.01)
Number (+ percentage) of reactions with 3 alleles	37 (8.89)	21 (9.01)	8 (2.54)	3 (1.59)	36 (6.77)	6 (2.49)	18.50 (5.76)
Number (+ percentage) of multi-allele reactions	60 (14.42)	25 (10.73)	15 (4.76)	3 (1.59)	57 (10.71)	9 (3.73)	28.17 (8.77)
Number (+ percentage) of recovered alleles	513 (7.13)	279 (3.88)	338 (4.69)	195 (2.71)	625 (8.68)	256 (3.56)	367.67 (5.11)

A summary of all the LDBSP parameters for pools of 50 neurons isolated from unfixed post-mortem brain tissue using laser capture microdissection (LCM). Displayed are the total number of obtained singleplex polymerase chain reaction (PCR) products for each gene, the total amount of included and excluded products according to the traditional LDBSP method for the downstream data analysis, as well as the total amount of estimated alleles according to the novel LDBSP data analysis approach. Please refer to the written text for further specifications on both allele estimation methods and criteria.

LDBSP on pools of 50 neurons occasionally renders reactions with more than one target allele

As the principle of LDBSP is based on the seclusion of individual alleles, the vast majority of PCR products should theoretically represent an amplicon derived from one DNA molecule [10]. As a given CpG site on a single allele is simply methylated or not, pyrosequencing of these PCR products should result in obtaining binary CpG methylation readouts, *i.e.* the percentage of methylation should approach 0% for unmethylated and 100% for methylated CpG sites. Because the quantitative measurement of a CpG site by pyrosequencing does not only depend on its methylation status, but also on the sensitivity of the assay and other technological factors, earlier studies have considered CpG methylation values of <20% and >80% indicative of unmethylated and methylated CpG sites, respectively [10, 11, 14]. Moreover, products that contained at least one CpG methylation value between 20% and 80% were excluded from further downstream analyses, as these were considered to represent measures of multiple target alleles in a single reaction or from other technological artifacts. Importantly, as previous studies have mainly applied LDBSP on a single or a few cells, the occurrence of multi-allele reactions has often been negligible [11].

When applying LDBSP on pools containing a larger number of cells, *i.e.* 50 neurons in our case, one can observe that a substantial proportion of the obtained products do not survive the traditional LDBSP inclusion criteria. Evidently, processing DNA isolated from a pool of 50 neurons increases the chance of allele clumping when compared to conducting LDBSP on a single or just a few cells. As such, when two different alleles with an opposite methylation status for at least one CpG site end up in the same reaction, the methylation value obtained for specifically this site should approach ~50%. Along similar lines, when three alleles end up in the same reaction with a different methylation status for at least one CpG site, then the methylation value obtained for specifically this site should approach either ~33.33% or ~66.66%. From all the PCR products of *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT*, respectively, 16.83%, 12.02%, 5.40%, 1.59%, 10.90% and 8.30% would have been omitted from further downstream analyses, as at least one CpG site in these products demonstrated a value between 20% and 80% (Table 1). A thorough inspection of these deviating CpG methylation values revealed that the vast majority approached either 50%, or 33%/67%, indicative of the presence of two or three alleles. Thus, based on these typical patterns of CpG methylation values, we concluded that a considerable amount of the LDBSP data could be attributed to reactions that contained more than one target allele, hence suggesting an adapted protocol for estimating methylation rates using this approach is required. Evidently, one can also obtain methylation values of 0% or 100% in the case of two or three allele-

reactions, which demands for a reliable determination of the number of alleles in each reaction.

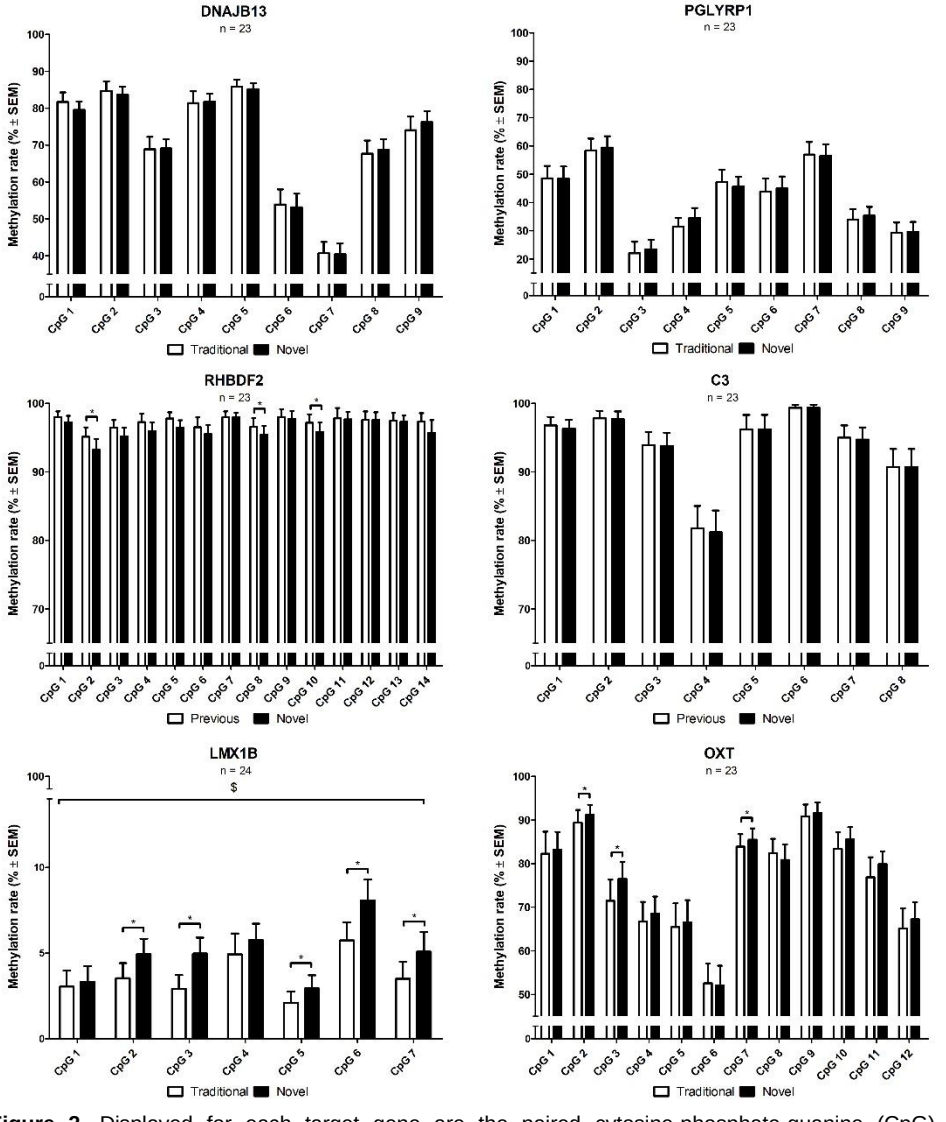


Figure 2. Displayed for each target gene are the paired cytosine-phosphate-guanine (CpG) methylation rates (mean percentage ± standard error of the mean (SEM)) estimated by both the traditional and novel downstream data analysis for limiting dilution bisulfite pyrosequencing (LDBSP) on pools of 50 neurons. A one-way repeated measures multivariate analysis of variance (MANOVA) was performed in order to determine whether there was a combined significant difference in the CpG methylation rates estimated by both methods. Subsequently a univariate analysis was performed in order to identify differences between the methylation rates for each individual CpG site. Significant findings for the multivariate and univariate tests are indicated with $ and *, respectively. A *p* value of < 0.05 was considered as statistically significant.

An integrated analysis of CpG readouts for LDBSP data with correction for multi-allele reactions

In order to correct for the number of estimated alleles present in each of the PCR reactions, we established a novel method for the downstream data analysis. For this purpose, thresholds of (1) ≤8.33% and ≥91.33%, (2) 50±8.33%, and (3) 33.33±8.33% and 66.66±8.33%, were set for the CpG methylation values, enabling assessing both 1-, 2- and 3-allele reactions, respectively. Accordingly, individual CpG sites that fell within the first, second and third threshold range were called and considered indicative for the potential presence of one, two and three alleles, respectively. A definitive (total) allele score for each individual product was then assigned following a multi-step filtering process that was based on the criteria described hereafter. All products solely displaying binary CpG methylation patterns (*i.e.* when every CpG site within an amplicon displays methylation levels within the first threshold range) were directly considered to be derived from single allele reactions, as the CpG methylation profiles for these products displayed a strong, typical binary pattern that is expected for a single DNA molecule. Similarly, and based on the assumption described in the previous section, all products displaying CpG methylation values that fell only in the first and second, but not third, or only in the first and third, but not second, threshold ranges were scored as two or three alleles, respectively.

Subsequently, products containing CpG methylation values that fell outside of the assigned threshold ranges, *i.e.* between 8.33%-25% and 75%-91.33%, and/or products with values that were indicative of both two and three alleles simultaneously, were assessed. The observation of such more ambiguous products displaying small deviations from the assigned threshold values are likely to be caused by technical variation induced during the PCR and pyrosequencing procedure and the more stringent thresholds used in the present study in this respect (8.33% versus 20% in previous studies; see above). All of these products were therefore thoroughly inspected by two investigators that were blinded to the experimental conditions, and a decision on the total number of alleles present in each individual reaction, *i.e.* one, two or three alleles, was made independently, while taking into account a combination of different factors. These included, but were not limited to, small technological variation that was previously observed during sensitivity testing of the assays, the directionality and methylation status of other CpG sites in the same product and the total number of dominant allele indicators, *i.e.* whether a product demonstrated more or less suggestive CpG sites for either two or three alleles. Furthermore, a likelihood estimation for each CpG site was made by taking into account the methylation status frequency on other gene-specific products obtained from the same individual, as well as from identical products obtained from other individuals. A cross comparison between

the independent score sheets was then performed (98.38% overlap per gene on average) and reactions with a deviating score between the first two investigators were assessed by a third (blinded) investigator. A final allele number was then assigned for these reactions based on the overlap between the score sheets of the third and first two investigators, *i.e.* when two out of the three investigators assigned the same score then this allele number was used for the respective reaction.

Table 2. One-way repeated measures multivariate analysis of variance (MANOVA)

Multivariate test	**RHBDF2**	**OXT**	**DNAJB13**	**PGLYRP1**	**C3**	**LMX1B**
Combined effect	0.390	0.474	0.078	0.392	0.385	0.012
Univariate tests						
CpG 1	0.163	0.544	0.194	0.962	0.328	0.435
CpG 2	0.027	0.048	0.354	0.493	0.328	0.010
CpG 3	0.066	0.005	0.905	0.527	0.328	0.009
CpG 4	0.100	0.407	0.799	0.090	0.312	0.314
CpG 5	0.062	0.581	0.560	0.519	1,000	0.040
CpG 6	0.205	0.692	0.662	0.517	1,000	0.001
CpG 7	0.936	0.033	0.859	0.863	0.328	0.006
CpG 8	0.026	0.349	0.567	0.544	1,000	-
CpG 9	0.328	0.381	0.117	0.857	-	-
CpG 10	0.033	0.087	-	-	-	-
CpG 11	0.810	0.228	-	-	-	-
CpG 12	0.963	0.394	-	-	-	-
CpG 13	0.779	-	-	-	-	-
CpG 14	0.054	-	-	-	-	-

Displayed are the p-values for both the multivariate test of the combined cytosine-phosphate-guanine (CpG) effects and the univariate tests for each individual CpG site per target gene. A p-value of <0.05 was considered as stastically significant.

Overall, we estimated that from the PCR products of *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, LMX1B and *OXT*, 85.58%, 89.27%, 95.24%, 98.41%, 89.29% and 96.27%, respectively, were derived from single alleles, whereas 5.53%, 1.72%, 2.22%, 0.00%, 3.95% and 1.24%, respectively, were derived from two alleles, and 8.89%, 9.01%, 2.54%, 1.59%, 6.77% and 2.49%, respectively, were derived from three alleles (Table 1). By taking into account these multi-allele reactions, we therefore estimated that in total 513, 279, 338, 195, 625 and 256 alleles were recovered for *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT*,

respectively, with an average recovery rate of 5.11%. It has previously been demonstrated that shorter assays generally have a higher recovery rate for LDBSP [11], something that could also be observed in our data set. In fact, the highest number of alleles were recovered for the shortest assay (*LMX1B*, 625 alleles; multiplex, 257 basepairs (bp); singleplex, 249 bp), while the lowest number was recovered for the longest assay (*C3*, 195 alleles; multiplex 441 bp; singleplex, 364 bp). Accordingly, a Pearson's correlation test revealed a strong negative correlation (r = -0.866, p = 0.026) between the number of recovered alleles and multiplex amplicon length, thereby confirming previous observation.

Based on the final allele estimations above, the CpG methylation rates, representing the percentage of methylated CpG sites from the total number of recovered alleles per gene, were calculated per individual. An overview of the average CpG methylation rates can be found in Figure 2 for each target gene. In addition to estimating these rates by the novel method described above, we also quantified them according to the traditional LDBSP approach that is based on excluding multi-allele reactions.

Gene specific changes in the CpG methylation data based on the novel integrated analysis

Next, for the CpG site methylation rates of each gene, a one-way repeated measures MANOVA was performed in order to determine whether there was a combined significant difference in the CpG methylation rates estimated by the traditional and novel LDBSP data analysis approach described above (Table 2). A significant effect (p = 0.012) was identified for *LMX1B*, demonstrating that, for this gene, the estimated CpG methylation rates are different depending on the applied calling procedure. A subsequent univariate analysis for each individual CpG site, indicated that the estimated methylation rates for 5 out of 7 sites were significantly different between the two approaches (CpG 2, p = 0.010; CpG 3, p = 0.009; CpG 5, p = 0.040; CpG 6, p = 0.001; CpG 7, p = 0.006). While no significant combined effect was identified for any of the other target genes, the univariate analyses did reveal that the estimated methylation rates for two CpG sites in both *RHBDF2* and *OXT* were different between the traditional and novel method. While for *RHBDF2* CpG site 8 (p = 0.026) and 10 (p = 0.033) were significantly different, CpG site 3 (p = 0.005) and 7 (p = 0.033) of *OXT* differed between the two approaches. All together, these findings demonstrate that the obtained methylation rates can be significantly affected by the method that is applied for downstream data analysis when conducting LDBSP on pools of 50 neurons. In other words, when excluding reactions that suggest to contain multiple alleles, instead of correcting the derived methylation rates based on the number of estimated alleles present in the reaction, the eventual LDBSP data can

differ significantly. For this reason, and in order to prevent potential bias in the experimental outcomes, we strongly suggest to apply this adapted approach for LDBSP when analyzing pools including larger numbers of cells. Overall, this novel pipeline approach provides a closer estimate of the true CpG methylation rates for a target gene.

Discussion

In the present study, we demonstrated, for the first time, that LDBSP can be successfully applied on LCM neurons derived from unfixed post-mortem brains. In brief, brain tissue sections were stained for a neuronal subtype-specific marker (*i.e.* 5-HT). Immuno-positive cells were then individually isolated using LCM and subsequently divided in small pools of 50 neurons. Next, the bisulfite-converted DNA isolated from a pool of neurons was diluted to a single allele-level and then amplified using (semi-)nested PCRs in multiplex-singleplex formation, targeting *DNAJB13*, *PGLYRP1*, *OXT*, *RHBDF2*, *C3* and *LMX1B* simultaneously. Finally, the methylation status of the target genes was then quantified using bisulfite pyrosequencing.

In contrast to most previous studies [10-15], LDBSP on pools of 50 neurons renders a considerable degree (8.77% on average per gene) of the downstream PCR reactions containing more than one target allele. The presence of these multiple alleles could be identified by the observed 'aberrant' CpG methylation values obtained from the respective reactions. Pyrosequencing of single alleles normally results in the derivation of a binary CpG methylation readout, that is the methylation values approach 0% for unmethylated and 100% for methylated sites [10]. Although in the vast majority of reactions a binary CpG methylation pattern could be detected (91.23%), for 3.01% and 5.76% of the sequenced products, respectively, we obtained values that approached 50% or 33/67% for at least one CpG site. As such, we discovered that these reactions contained either two or three alleles, in which at least one DNA molecule had an opposite CpG methylation status as compared to the other allele(s) present in the same reaction. Traditionally, thresholds of <20% and >80% have been applied in order to define the methylation status of CpG sites, and all products displaying more intermediate values (20-80%) suggesting the presence of more than one allele, were excluded from further downstream analysis [10-15]. However, we argued that, particularly when the degree of these reactions is more substantial, exclusion might actually influence the data negatively, either by inducing bias or by reducing or reinforcing effect sizes.

We therefore established a novel data analysis pipeline that allows one to include and correct the LDBSP data for these multi-allele reactions, hence providing a

more accurate estimation of the CpG methylation rates. For this purpose, novel CpG site thresholds of (1) ≤8.33% and ≥91.33%, (2) 50±8.33%, and (3) 33.33±8.33% and 66.66±8.33%, were applied in order to identify reactions that contained one, two or three alleles, respectively. Next, a CpG site calling procedure was performed. First products displaying CpG methylation values that fell within the assigned threshold ranges were categorized as either one, two or three alleles. Remaining products containing ambiguous CpG sites, *i.e.* i) products with multiple CpG sites showing presence of either two or three alleles, or ii) methylation values falling outside of the threshold values, were subsequently inspected by three researchers blinded to the experimental conditions. An allele score was then assigned independently taking into account factors that could explain variability in the CpG methylation data, such as known differences in the sensitivity of the sequencing assays that were assessed before conducting the experiment using DNA methylation standards. Additional measures such as the directionality and methylation status of other CpG sites in the same product or the total number of dominant allele indicators in a single product were also considered. Moreover, a likelihood estimation was performed for each CpG site taking into account the methylation status frequency for other gene-specific products obtained from the same individual, as well as from identical products obtained from other individuals. A cross comparison between the independent score sheets was then performed (98.38% overlap per gene on average), reactions with a deviating score between the first two investigators were assessed by a third (blinded) investigator and a final allele number was then assigned for these reactions based on the overlap between the score sheets of the third and first two investigators.

For each gene, we then compared the CpG methylation rates estimated both by the traditional (<20% and >80% cut-offs) and novel data analysis approach described above. Strikingly, significant differences in the combined CpG effects for *LMX1B*, as well as for individual CpG sites in *LMX1B*, *RHBDF2* and *OXT* were identified. Thus, these findings emphasized that the derived LDBSP methylation data can significantly differ depending on the method that is being applied for the downstream data analysis. Importantly, this will especially affect loci that for biological reasons display a varying methylation status over all the recovered allele, that is when part of them is fully methylated and part of them fully unmethylated. In such a scenario, the average methylation rates will likely fall in the intermediate range of 20-80% and, hence, the chances of obtaining a multi-allele reactions with values outside of the traditional threshold will therefore be higher. Therefore, the probability of detecting differences between both methods is also dependent on the average methylation scores. In either way, when LDBSP renders reactions with more than one target allele, independent of the cell types

that are analyzed, we strongly suggest to correct the CpG methylation rates based on the number of target alleles present in a single reaction. Overall, we strongly believe that estimating the CpG methylation rates based on the CpG methylation values provides a more accurate representation of the true methylation rates obtained for a specific target gene, hence adding more validity to the data.

One important consideration, is that dependent on the design of an individual experiment, potentially more than three alleles might be present in an individual reaction, which would mean the estimation of allele numbers in a single reaction based on CpG methylation values becomes too ambiguous, as one also needs to consider potential technical variation that might affect these read-outs. In our hands, the proportion of reactions that contain more than three alleles seemed to be very little or even negligible, and these reactions therefore hardly affect the estimated methylation data. Even when present, the method proposed in the present manuscript will most likely classify these reactions as three alleles and, hence, the methylation data obtained from these reactions has been partially corrected when compared to the traditional analysis pipeline. Evidently, when ending up with a high degree of (>3) multi-allele reactions, it is advised to dilute the DNA more.

Another important consideration is that LDBSP does not allow the user to detect reactions that appear to contain only one allele, which in reality contain 2 or more alleles with an identical methylation pattern. In such a scenario, the methylation profile of a PCR product would appear binary, hence suggesting the presence of only one target allele. Indeed, this represents an issue independent of the downstream data analysis that is used for the analysis of the CpG methylation rates, *i.e.* the traditional or novel approach proposed here. Overall, this becomes a bigger challenge when due to biological reasons the CpG methylation rates of a target gene approach either 0 or 100%, meaning that most of the recovered alleles will display either fully unmethylated or methylated CpG sites, respectively. In our study, this might therefore have affected genes such as *LMX1B*, *C3* and *RHBDF2*, although we often observed that at least one of the CpG sites in these gene-derived products displayed an opposite methylation status compared to the other sites derived from the same product. In other words, the methylation status of all CpG sites in a target gene were not always identical and often displayed a deviating methylation state for one site when compared to the others. For this reason, the occurrence of these opposite methylation patterns at single CpG sites still allowed us to detect the presence of multiple alleles when such molecules end up in a reaction with other alleles showing a more homogenous pattern of CpG methylation. It is therefore advisable to always target a substantial number

of CpG sites per target gene, *e.g.* 7-14 in the present manuscript, in order to increase the chance of obtaining, and hence detecting, at least one site with a 'non-binary' methylation status on one of the target alleles. As such, a readout of such a reaction would demonstrate an intermediate methylation value, *e.g.* ~50% and ~33% or ~66%, for specifically this CpG site. When targeting only a few or even a single CpG site, further diluting the DNA may be advisory.

Undoubtedly, when conducting LDBSP, it is paramount to ensure proper dilution of the isolated target DNA *a priori*, in order to minimize the occurrence of multi-allele reactions, even before one ought to correct the derived methylation values as proposed in the present manuscript. LDBSP is thought to follow a Poisson distribution, meaning that with increasing dilution the chances of obtaining multi-allele reactions should become smaller. As a benchmark, it is therefore advised to use more dilutions than the number of target alleles present in the reactions, whilst taking into account potential loss due to handling. Adequate dilution to a single allele-level can be assessed by the ratio between the number of product-yielding and non-product yielding reactions, of which the latter should occur more often than the former. Although in the present study the number of product-yielding reactions was substantially lower (20.27% on average) compared to the downstream reactions that did not contain a product (79.73% on average), we still obtained reactions with more than one target allele, suggesting that other technological factors, or biochemical or biophysical properties of the DNA, play a role as well. Nevertheless, in a scenario where the DNA has been diluted properly and where multi-allele reactions are still encountered it will likely remain more appropriate to correct methylation values for these reactions as compared to excluding them from the data analysis. Of note, simply further diluting the DNA will be on the expense of throughput and comes along with additional costs.

Finally, in order to demonstrate the overall efficiency of the approach, we made a final estimation on the total number of alleles that were recovered after correcting for the number of alleles present in a single reaction. Assuming a 100% DNA recovery starting from three pools of 50 neurons, we estimated that the allele recovery rate of LDBSP on pools of 50 neurons isolated with LCM is 5.11% per gene of the total number of alleles. In comparison, previous studies have estimated recovery rates ranging from 0.3% up to 25% using intact cell pools consisting of 10 oocytes, two-cell embryos and 16-cell embryos [11, 13, 14]. The recovery rate achieved in the present study on the pools of 50 neurons is therefore similar to other studies using these other cell types, despite the fact that a higher amount of starting material is needed when conducting the technique on neurons isolated with LCM. Nevertheless, one should not overlook the differences in sample collection procedures that were applied in these studies. In

fact, the success of extracting DNA from cells isolated with LCM is dependent on several critical factors, *e.g.* cutting procedures, thickness of the sections, staining procedures, exposure to heat produced by the laser, and the DNA isolation process, which all could affect the eventual quantity of DNA that is recovered from the cells. Furthermore, the degradation and low complexity of bisulfite-converted DNA opposes another challenge for the methylation analysis of small amounts of DNA from only few cells. As such, a previous study demonstrated that shorter assays generally have a higher recovery rate for LDBSP [11]. Interestingly, also in our study the highest number of alleles were recovered for the shortest assay (*LMX1B*, 625 alleles; multiplex, 257 basepairs (bp); singleplex, 249 bp), while the lowest number was recovered for the longest assay (*C3*, 195 alleles; multiplex 441 bp; singleplex, 364 bp). Accordingly, a strong negative correlation between the number of recovered alleles and multiplex amplicon length was identified, thereby confirming previous observation.

In conclusion, the approach described here, relying on a combination of LDBSP with LCM, offers a novel and alternative strategy to single cell bisulfite sequencing techniques that can be applied for the study of DNA methylation marks in the human brain. Moreover, the approach offers a workable solution for the challenge of tissue and cell-type specificity as encountered in the field of neuroepigenomics. In fact, LDBSP on pools of neuronal populations allows one to determine DNA methylation profiles in a multi-targeted and cell subtype-specific manner, hence avoiding potential noise in epigenetic data that is induced by analyzing heterogeneous tissue samples. Aside from allowing the identification of methylation marks in individual neuronal cells, we expect that similar strategies using other isolation techniques and other cell subtypes in combination with LDBSP will be increasingly valuable for future neuroepigenomic studies.

Acknowledgments
Funds have been provided by the Joint Program— Neurodegenerative Disease Research (JPND) for the EPI-AD consortium (http://www.epi-ad.eu/). The project is supported through the following funding organizations under the aegis of JPND; the Netherlands, The Netherlands Organization for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 643417.

References

1. van den Hove DL, Riemens RJ, Koulousakis P, Pishva E. Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges. Brain Pathology. 2020.
2. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131:21-64.
3. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics & chromatin. 2018;11(1):41.
4. Roubroeks JAY, Smith RG, van den Hove DLA, Lunnon K. Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. J Neurochem. 2017;143(2):158-70.
5. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13(1):86.
6. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics. 2014;30(10):1431-9.
7. Rahmani E, Zaitlen N, Baran Y, Eng C, Hu D, Galanter J, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. Nature methods. 2016;13(5):443.
8. Rahmani E, Zaitlen N, Baran Y, Eng C, Hu D, Galanter J, et al. Correcting for cell-type heterogeneity in DNA methylation: a comprehensive evaluation. Nature methods. 2017;14(3):218-9.
9. Zheng SC, Beck S, Jaffe AE, Koestler DC, Hansen KD, Houseman AE, et al. Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. Nature methods. 2017;14(3):216-7.
10. Hajj NE, Kuhtz J, Haaf T. Limiting Dilution Bisulfite Pyrosequencing(R): A Method for Methylation Analysis of Individual DNA Molecules in a Single or a Few Cells. Methods Mol Biol. 2015;1315:221-39.
11. El Hajj N, Trapphoff T, Linke M, May A, Hansmann T, Kuhtz J, et al. Limiting dilution bisulfite (pyro)sequencing reveals parent-specific methylation patterns in single early mouse embryos and bovine oocytes. Epigenetics. 2011;6(10):1176-88.
12. Beach TG, Sue LI, Walker DG, Roher AE, Lue L, Vedders L, et al. The Sun Health Research Institute Brain Donation Program: description and experience, 1987-2007. Cell Tissue Bank. 2008;9(3):229-45.
13. Charnay Y, Léger L. Brain serotonergic circuitries. Dialogues in clinical neuroscience. 2010;12(4):471.
14. Mattern F, Heinzmann J, Herrmann D, Lucas-Hahn A, Haaf T, Niemann H. Gene-specific profiling of DNA methylation and mRNA expression in bovine oocytes derived from follicles of different size categories. Reprod Fertil Dev. 2017;29(10):2040-51.
15. Trapphoff T, El Hajj N, Zechner U, Haaf T, Eichenlaub-Ritter U. DNA integrity, growth pattern, spindle formation, chromosomal constitution and imprinting patterns of mouse oocytes from vitrified pre-antral follicles. Human reproduction. 2010;25(12):3025-42.
16. Heinzmann J, Hansmann T, Herrmann D, Wrenzycki C, Zechner U, Haaf T, et al. Epigenetic profile of developmentally important genes in bovine oocytes. Molecular reproduction and development. 2011;78(3):188-201.
17. Diederich M, Hansmann T, Heinzmann J, Barg-Kues B, Herrmann D, Aldag P, et al. DNA methylation and mRNA expression profiles in bovine oocytes derived from prepubertal and adult donors. Reproduction. 2012;144(3):319.

Supplementary material

Limiting dilution bisulfite pyrosequencing protocol for 50 laser capture microdissected cells

Reagents

- Laser capture microdissected cells (pools of 50 cells) in 8 µl Tris-Ethylenediaminetetraacetic acid (TE) buffer or phosphate buffer saline (PBS)
- FastStart™ Taq DNA Polymerase (Sigma Aldrich, cat. no. 12032929001)
- EZ DNA Methylation-Direct Kit (Zymo Research, cat. no. D5020)
- Agarose
- DNA loading dye, *e.g.* Orange G DNA Loading Dye (ThermoFisher Scientific, cat. no. R0631)
- Nucleic acid gel stain, *e.g.* GelsGelStar™ Nucleic Acid Gel Stain (Lonza, cat. no. 50535)
- DNA 100 bp ladder, *e.g.* 100 bp DNA Ladder (ThermoFisher Scientific, cat. no. 15628050)
- PyroMark Gold Q96 CDT reagent kit (Qiagen, cat. no. 972824)

Equipment

- Manual pipettes
- Multichannel pipette
- Pipette filtered tips
- Eppendorf LoBind microcentrifuge tubes (Merck KGaA, Darmstadt, Germany, cat. no. 022431021)
- TipOne Low Retention Tips (STARLAB, Hamburg, Germany, cat. no. S1180-8810)
- Eppendorf® ThermoMixer® F1.5 (Sigma Aldrich, cat. no. EP5384000012)
- Microcentrifuge with PCR tube adaptors
- Thermal cycler
- Thin-walled PCR tubes, 0.2 ml
- PyroMark Q96 MD pyrosequencing system (Qiagen)
- Gel electrophoresis equipment

Reagents preperations

CT Conversion Reagent (EZ DNA Methylation-DirectTM Kit; D5020)

The CT Conversion Reagent provided with the EZ DNA Methylation-DirectTM Kit (D5020) is a solid mixture and must be prepared prior to first use. Please refer to the protocol provided with the kit for further details.

1. Add 790 µL of M-Solubilization Buffer to a tube of CT Conversion Reagent
2. Add 300 µL of M-Dilution Buffer to the tube of CT Conversion Reagent
3. Mix at room temperature with frequent vortexing or shaking for 10 minutes
4. Add 160 µL of M-Reaction Buffer to the tube of CT Conversion Reagent
5. Mix an additional 1 minute

M-Wash Buffer

The M-Wash Buffer provided with the EZ DNA Methylation-DirectTM Kit (D5020) must be prepared prior to first use. Please refer to the protocol provided with the kit for further details.

1. Add 24 mL of 100% ethanol to the 6 mL M-Wash Buffer concentrate (D5020)

Procedure

DNA isolation (EZ DNA Methylation-DirectTM Kit; D5020)

Laser capture microdissected cells in pools of 50 should be collected in 8 µl TE buffer or PBS, as this is the optimal working volume for the EZ DNA Methylation-Direct Kit. In our case, cells were collected in 10 µl, hence volumes of the CT Conversion Reagent were adjusted accordingly in order to maintain the suggested reagent concentrations during the bisulfite conversion. For the original reagent volumes using laser capture microdissected cells in a total volume of 8 µl, please refer to the protocol provided with the kit. Further details on LDBSP sample handling can be find below.

1. Briefly centrifuge the tube with laser capture microdissected cells
2. Add 11 µl M-Digestion Buffer (2X) to the sample
3. Add 1 µl of Proteinase K to the sample
4. Briefly vortex and centrifuge the tube
5. Incubate the sample in a heating block at 50°C for overnight

Bisulfite conversion of DNA (EZ DNA Methylation-DirectTM Kit; D5020)

When pipetting the sample, always use TipOne Low Retention Tips in order to minimize loss of DNA due to handling. Prepared CT Conversion Reagent from the EZ DNA Methylation-Direct Kit should be divided in 143 µl ready-to-use aliquots in separate Eppendorf tubes, as this will allow one to use the same pipet tip per sample throughout the entire procedure. Similarly, the M-Binding Buffer was prepared in 600 µL ready-to-use aliquots per sample. Please find further details on handling below.

1. After incubation, transfer the digested sample to a PCR tube
2. By using the same pipet tip, take 143 µl CT Conversion Reagent

3. Use the CT Conversion Reagent to wash out the digestion tube
4. Add the CT Conversion Reagent to the sample containing PCR tube
5. Briefly vortex and centrifuge the tube
6. Place the PCR tube in a thermal cycler and perform the following steps:
    a. 98°C for 8 minutes
    b. 64°C for 3.5 hours
    c. 4°C storage
7. Add 200 µL of M-Binding Buffer into a Zymo-Spin™ IC Column and place the column into a provided Collection Tube
8. Pipet the bisulfite converted sample into the Zymo-Spin™ IC Column
9. By using the same pipet tip, take 200 µL of M-Binding Buffer
10. Use the M-Binding Buffer to wash out the PCR tube used for bisulfite conversion
11. Add the 200 µL of M-Binding Buffer to the sample containing Zymo-Spin™ IC Column
12. Repeat step 9-11 once more
13. Close the cap and mix by inverting the column several times
14. Centrifuge at full speed (>10,000 x g) for 30 seconds
15. Discard the flow-through
16. Add 100 µl of M-Wash Buffer to the column
17. Centrifuge at full speed (>10,000 x g) for 30 seconds
18. Add 200 µl of M-Desulphonation Buffer to the column and let stand at room temperature (20-30°C) for 15-20 minutes
19. Centrifuge at full speed (>10,000 x g) for 30 seconds
20. Add 200 µl of M-Wash Buffer to the column
21. Centrifuge at full speed (>10,000 x g) for 30 seconds
22. Add 200 µl of M-Wash Buffer to the column
23. Centrifuge at full speed (>10,000 x g) for 30 seconds
24. Place the column into a 1.5 ml LoBind Eppendorf tube
25. Add 10 µl of M-Elution Buffer directly to the column matrix
26. Centrifuge for 30 seconds at full speed (>10,000 x g) to elute the DNA
27. Add 10 µl of M-Elution Buffer directly to the column matrix
28. Centrifuge for 30 seconds at full speed (>10,000 x g) to elute the DNA

NOTE: We strongly advice to process the samples directly for PCR amplification without intermediate storage by freezing in order to avoid potential loss of bisulfite-converted DNA due to free-thaw-cycles.

Multiplex PCR

The volumes for the Multiplex PCR can vary depending on the number of dilutions and target genes used for the LDBSP analysis. For illustrative purposes, we here

provide an example based on 6 target genes and 22 dilutions, using a total volume of 25 µl for each individual PCR reaction. Overall, the content of each individual PCR reaction in relation to the buffer, dNTPs and Taq DNA polymerase should be identical, whereas volumes of H_2O vary depending on the number of target genes. Furthermore, always prepare several extra reactions for negative and positive controls, as well as for counteracting pipetting errors. The calculation demonstrated here is based on 22 dilutions (or reactions) for the samples, 2 control reactions and 1 extra reaction for pipetting errors, giving a total of 25 reactions.

6. Prepare a single PCR mix containing the following reagents:
    a. 2.5 µL PCR buffer (10X) with 20 mM MgCl2 per reaction
        ➔ *25 x 2.5 µL= 62.5 µL*
    b. 0.5 µL 10 mM dNTP mix per reaction
        ➔ *25 x 0.5 µL= 12.5 µL*
    c. 1 µL of the forward primer (10 µM stock) per gene per reaction
        ➔ *25 x 6 genes x 1 µL = 150 µL*
    d. 1 µL of the reverse primer (10 µM stock) per gene per reaction
        ➔ *25 x 6 genes x 1 µL = 150 µL*
    e. 0.2 µL (5 U/µL) FastStart™ Taq DNA Polymerase per reaction
        ➔ *0.2 * 25 = 5 µL*
        ➔ *Components 'a'-'e'; total volume = 380 µl*
7. Mix by shaking and then centrifuge the PCR mix
8. Divide the PCR mix in two separate Eppendorf tubes according to the necessary volumes
    a. Tube 1, 22 reactions: 334.4 µl PCR mix
        ➔ *380 / 25 x 22 = 334.4 µl*
    b. Tube 2, 2 reactions: 30.4 µl PCR mix
        ➔ *380 / 25 x 2 = 30.4 µl*
9. Add a required volume of H_2O to both the PCR mixes (See details below)
    a. Tube 1, 22 reactions: 195.6 µl H_2O
        ➔ *22 reactions x 25 µl total volume = 550 µl*
        ➔ *Minus 334.4 µl PCR mix = 215.6 µl*
        ➔ *Minus 20 µl bisulfite converted DNA = 195.6 µl*
    b. Tube 2, 2 reactions: 17.6 µl H_2O
        ➔ *2 reactions x 25 µl total volume = 50 µl*
        ➔ *Minus 30.4 µl PCR mix = 19.6 µl*
        ➔ *Minus 2 x 1 µl of template (Control DNA or H_2O) = 17.6 µl*
10. Mix by shaking and then centrifuge the PCR mixes
11. Add the complete PCR mix from 'Tube 1' to the 20 µl of bisulfite converted DNA
12. Mix thoroughly by pipetting up-and-down

13. By using the same pipet tip, dispense 25 µl of the mixture over 22 wells of a micro-titer plate (mix thoroughly inbetween dispensation steps)
14. Dispense 24 microliter of the PCR mix from 'Tube 2' into 2 wells of a micro-titer plate
15. Add the necessary control templates to the latter 2 wells (*e.g.* positive and negative)
16. Seal the micro-titer plate
17. Place the micro-titer plate in a thermal cycler and perform the following steps:
    a. 95°C for 5 minutes
    b. 95°C for 30 seconds
56°C for 30 seconds
72°C for 1 minute } x 43 cycles
    c. 72°C for 7 minutes
    d. 4°C storage

NOTE: The optimum annealing temperature and required number of cycles should be tested a priori, both independently, as well as in multiplex formation. For the multiplex PCR amplifications, all primers need to be compatible for the same cycler conditions. The program displayed here was applied in the present study.

Singleplex PCR

Prepare a total volume of singleplex PCR mix that is sufficient for the number of sample containing reactions (i.e. 22 in this case), negative controls and some extra to counteract pipetting errors.

1. Each PCR reaction should contain the following reagents per tube (Total volume of 25 µL):
    a. 1 µl of Multiplex PCR product
    b. 2.5 µL PCR buffer (10X) with 20 mM MgCl2
    c. 0.5 µL 10 mM dNTP mix
    d. 1 µL of the forward primer (10 µM stock)
    e. 1 µL of the reverse primer (10 µM stock)
    f. 0.2 µL (5 U/µL) FastStart™ Taq DNA Polymerase
    g. 18.8 µl of H_2O
2. Place the micro-titer plate in a thermal cycler and perform the following steps:
    a. 95°C for 5 minutes
    b. 95°C for 30 seconds
58°C for 30 seconds
72°C for 1 minute } x 45 cycles

c. 72°C for 7 minutes
d. 4°C storage

3. Visualize the PCR products on a 2% agarose gel, using the Orange G DNA Loading Dye, the GelsGelStar™ Nucleic Acid Gel Stain and a 100 bp ladder according to the manufacturer's instructions.
4. Identify which of the reactions yield a gene-specific PCR product
5. Use the required amount of PCR product as suggested by the manufacturer's instructions for subsequent pyrosequencing analysis.
6. Please refer to the manufacturer's instructions for details on bisulfite pyrosequencing

NOTE: The optimum annealing temperature and required number of cycles should be tested a priori for each target gene, using both bisulfite converted DNA, as well as multiplex product. The program displayed here was applied in the present study.

Supplementary Table 1. Cohort demographics

	AD patients	**Non-demented controls**
N	12	12
Sex	Female	Female
Age of death; Mean (SD)	84.5 (5.04)	77.92 (13.00)
PMI; Mean (SD)	2.89 (0.96)	2.48 (0.62)
Plaque total; Mean (SD)	13.69 (1.71)	2.16 (3.54)
Tangle total; Mean (SD)	14.0 (1.36)	3.33 (2.38)
Braak stage; Range	V-VI	I – IV

Dorsal raphe nucleus (DRN) tissue obtained from the Banner Sun Health Research Institute (BSHRI, Sun City, Arizona, US) as part of the Brain and Body Donation Program (BBDP). Displayed is the number of samples in each group, the sex of the subjects and the mean plus standard deviation (SD) of their age of death, postmortem interval (PMI, in hours), total plaque load and total tangle load, as well as their range of Braak stage.

Supplementary Table 2. Primer overview

PCR primers

Gene	**Forward primer (5'-3')**	**Reverze primer (5'-3')**	**Target region (GRCh37)**	**Product size (bp)**
DNAJB13 (Multiplex)	GGTATTTTGGGAGGATGAGTTA	CCCTTCAAAAACAAACCAACTAAT	11:73668561-73668825	265
DNAJB13 (Singleplex)	GTATTTTGGGAGGATGAGTTATAATTG	(Bio-)CTTCAAAAACAAACCAACTAATAC	11:73668563-73668824	262
PGLYRP1 (Multiplex)	GTTGGGGAGAGTTTAGGTAAG	ATTCCTATTAAATTATTCAATATTCCACT	19:46526238-46526634	397

Supplementary Table 2. *(Continued)*

PCR primers

Gene	Forward primer (5'-3')	Reverze primer (5'-3')	Target region (GRCh37)	Product size (bp)
PGLYRP1 (Singleplex)	TGGGGAGAGTTTAGGTAAGTA	(Bio-) TCCTATTAAATTATTCAATATTCCACTAA	19:46526240-46526632	393
LMX1B (Multiplex)	TTTTTTAAGGGGGTGGAGTAGAG	CCTCACCTCCCCAAACTTCCTAACATT	9:129375928-129376184	257
LMX1B (Singleplex)	GGGGGTGGAGTAGAGGGA	(Bio-) CCTCACCTCCCCAAACTTCCTAACATT	9:129375936-129376184	249
OXT (Multiplex)	GTTTAGGTTTTGTTAATGAAGAGGAA	ACCAAAACTAAAAACTCACCTTAC	20:3052117-3052440	324
OXT (Singleplex)	GTTTAGGTTTTGTTAATGAAGAGGAA	(Bio-) TCTTACCTCCCAAAAAACAATTC	20:3052117-3052392	276
C3 (Multiplex)	GATTGGGTTTTATTTGAGTGTAAGAT	CCAACAATAAATTAAAAACTCCAATCTTC	19:6713187-6713627	441
C3 (Singleplex)	GATTGGGTTTTATTTGAGTGTAAGAT	(Bio-) AACCTAAAACCCTCCTTATCTAT	19:6713187-6713550	364
RHBDF2 (Multiplex)	AGGGTTTTGGGGATTTAGTG	CCAAAAACCTCACAAACAAATC	17:74475178-74475497	320
RHBDF2 (Singleplex)	AGGGTTTTGGGGATTTAGTG	(Bio-) CATCACCCCTTCCACACACTCAA	17:74475178-74475435	258

Pyrosequencing primers

Gene	Sequencing primer (5'-3')	Number of CpGs	PyroMark Orientation	Target region (GRCh37)	Primer
DNAJB13	GGATGAGTTATAATTGGG	5	Lower strand (5'-3')	11:73668745-73668795	1
DNAJB13	TTTTGTTATTTGTTTGAAAA	4	Lower strand (5'-3')	11:73668618-73668673	2
PGLYRP1	TTTTTTTTTGGTTGGGTTAG	1	Upper strand (5'-3')	19:46526451-46526454	1
PGLYRP1	GAGGGATGTATTGTGG	3	Upper strand (5'-3')	19:46526468-46526509	2
PGLYRP1	GTTTAGTAGGGAGGG	5	Upper strand (5'-3')	19:46526524-46526555	3
LMX1B	GTTTAGAAGAAGATTAAAATTTTTG	7	Upper strand (5'-3')	9:129375982-129376016	1
OXT	GGTTTTGTTAATGAAGAGGAA	5	Upper strand (5'-3')	20:3052143-3052173	1
OXT	ATATTATTAATTTTTTAAATAGAG	2	Upper strand (5'-3')	20:3052215-3052224	2
OXT	AATGTTTAGGTATAAAAAGGT	5	Upper strand (5'-3')	20:3052248-3052297	3
C3	TTTTAATTTGAGAAGGGAGA	2	Upper strand (5'-3')	19:6713340-6713376	1
C3	TGGGGTAGGGATTAAA	4	Upper strand (5'-3')	19:6713398-6713426	2
C3	ATTTAGGTAGGGAAA	2	Upper strand (5'-3')	19:6713450-6713475	3
RHBDF2	TTTTTGGTTGGGGAG	5	Upper strand(5'-3')	17:74475230-74475271	1

Supplementary Table 2. *(Continued)*

Pyrosequencing primers

Gene	Sequencing primer (5'-3')	Number of CpGs	PyroMark Orientation	Target region (GRCh37)	Primer
RHBDF2	GTATGTTGTAGTGGTG	5	Upper strand(5'-3')	17:74475287-74475332	2
RHBDF2	GATTTTAAATTATTGGGTTG	4	Upper strand(5'-3')	17:74475355-74475403	3

Displayed is an overview of the polymerase chain reaction (PCR) and pyrosequencing primers that were used in this study. Provided for each target gene are the multiplex and singleplex primer sequences (Bio, biotinylated), as well as corresponding genomic coordinates (Ensembl GRCh37 assembly) and amplicon sizes in base pairs (bp). For the pyrosequencing primers, the total number of CpG sites, the targeted region for sequencing and the orientation of the assay in the PyroMark assay design software 2.0 are listed per gene.

Chapter 5

Brain-region- and cell type-specific epigenetic profiling strongly implicates a role for dysregulation of *TNXB* and other loci in the brainstem in Alzheimer's disease

Renzo J.M. Riemens[1,2*], Ehsan Pishva[1,3*], Artemis Iatrou[4], Janou Roubroeks[3], Jennifer Nolz[5], Roy Lardenoije[1,6], Muhammad Ali[1,7], Antonio Del Sol[7], Raul Delgado-Morales[1,8], Manel Esteller[9], Gunter Kenis[1], Bart P.F. Rutten[1], Klaus-Peter Lesch[10], Jonathan Mill[3], Diego Mastroeni[1,5,11], Alfredo Ramirez[12], Thomas Haaf[2], Katie Lunnon[3†], Daniël L.A. van den Hove[1,10†]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]University of Exeter Medical School, University of Exeter, Exeter, United Kingdom.
[4]Rush Alzheimer's Neurodisease Center, Rush University Medical Center, Chicago, Illinois, United States of America.
[5]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[6]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany.
[7]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg.
[8]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute - IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[9]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain; Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Madrid, Spain; Institucio Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain.
[10]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[11]L.J. Roberts Center for Alzheimer's Research Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[12]Division for Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Medical Faculty, University of Cologne, Cologne, Germany; Department for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, Bonn, Germany.

* Authors contributed equally to this work
† Authors contributed equally to this work

Abstract

Increasing evidence suggests that dysregulation within brainstem nuclei such as the dorsal raphe nuclei (DRN) and the locus coeruleus (LC) represents one of the first steps in the pathogenesis of Alzheimer's disease (AD). Furthermore, recent epigenome-wide association studies (EWAS) have implicated a central role for epigenetic alterations in both the development and course of the disease. However, the extent of disease-specific deviant epigenetic signatures in the brainstem have not been investigated yet. The present study represents the first large-scale epigenetic analysis in the brainstem of AD, targeting both the DRN and the LC in parallel. In bulk tissues derived from both brainstem nuclei, differentially modified cytosine-phosphate-guanine (CpG) sites and regions were quantified at the level of three cytosine states, *i.e.* 5-methylcytosine (5-hmC), 5-hydroxymethylcytosine (5-mC) and unmodified cytosine (5-uC), utilizing the Illumina Infinium MethylationEPIC BeadChip array and making use of both customary bisulfite and innovative oxidative bisulfite treatment. Following methodological validation in a subset of patients from the same cohort using bisulfite pyrosequencing, we then targeted the bisulfite methylation signatures of our discovery findings in an independent patient cohort, where we profiled single laser capture microdissected serotonergic and non-serotonergic cells isolated from the DRN by means of modified limiting dilution bisulfite pyrosequencing (LDBSP). Overall, within both brainstem regions assessed, our EWAS revealed a strong overlapping dysregulation in the Tenascin XB (*TNXB)* gene, next to the identification of both previously identified and novel epigenetic loci that we hypothesize to play a pivotal role in the early development of AD. Furthermore, we demonstrated, for the first time, that the methylation profile of *TNXB* in the DRN is dependent on both the disease phenotype and the cell type analyzed, which warrants the need for future cell type-specific neuroepigenetic studies in AD.

Keywords

Alzheimer's disease, brainstem, dorsal raphe nucleus, locus coeruleus, DNA methylation (5-mC), DNA hydroxymethylation (5-hmC), Infinium MethylationEPIC BeadChip Kit, epigenome-wide association study (EWAS), single cell.

Introduction

Over the last years, prime sites of early interrogation of Alzheimer's disease (AD) have been extended beyond brain regions such as the hippocampus and the entorhinal cortex, since increasing evidence suggests that also the brainstem is among the first regions affected in the disease [1-3]. In fact, recent studies have demonstrated that neurofibrillary tangles in the brains of AD patients can already be observed in brainstem nuclei such as the dorsal raphe nuclei (DRN) and the locus coeruleus (LC), even before the manifestations of the first clinical symptoms [1-4]. In view of this notion, it is currently hypothesized that from these brainstem nuclei pathology could spread to subcortical areas, including the hippocampus, and, subsequently, to areas of the neocortex, marking the clinical stages of this neurodegenerative disease. Other evidence for a central role of the brainstem comes from observations using magnetic resonance imaging, indicating that this brain region is subject to both significant volume reductions and structural deformations in AD patients [5, 6]. Moreover, the occurrence of various non-cognitive behavioral and neuropsychological symptoms in AD, particularly in prodromal stages, including depression, disturbances in mood, emotions, appetite, respiratory and circadian rhythm, suggests the involvement of the DRN and the LC in the disorder [2].

Research aimed at exploring the etiopathophysiology of AD has furthermore implicated a crucial role for epigenetic mechanisms, in addition to both independent and interdependent genetic, environmental and life-style factors that influence the development and course of the disease [7, 8]. For this reason, the characterization of epigenetic profiles in brain tissue samples derived from AD patients and people at risk of developing AD is currently offering a highly attractive approach to study this disease. As such, several recent epigenome-wide association studies (EWAS) have identified robust differences in DNA methylation and hydroxymethylation in the AD cortex across independent patient cohorts [9-20]. Aside from studies targeting the brain, other EWAS using peripheral blood samples derived from patients have also identified dysregulated loci that harbor biomarker potential towards predicting the development and/or course of AD [15, 21]. Despite the increasing number of EWAS that offer novel insights into its pathophysiology, examining the brainstem for potential disease-specific deviant epigenetic signatures, indicative of the more incipient stages of AD, has, however, not been performed yet.

In the present study, we therefore conducted the first large-scale EWAS in the brainstem of AD to date. Our analytical approach consisted of two stages and focused on identifying alterations in cytosine-phosphate-guanine (CpG)

modifications associated with Braak stage at the level of unmodified cytosine (5-uC), 5-methylcytosine (5-mC; reflecting DNA methylation) and 5-hydroxymethylcytosine (5-hmC; reflecting DNA hydroxymethylation). First, in DNA derived from both DRN and LC bulk tissues, differentially modified positions and regions for each of the aforementioned modifications were quantified by using Illumina Infinium MethylationEPIC BeadChip arrays in combination with customary bisulfite and novel oxidative bisulfite treatment. These discovery findings were subsequently confirmed in a subset of patients from the same cohort using bisulfite pyrosequencing. Second, we followed up our EWAS by a validation study in an independent patient cohort targeting bisulfite methylation signatures in serotonergic and non-serotonergic cells isolated from the DRN by laser capture microdissection (LCM). Overall, within both brainstem regions assessed, our EWAS revealed a strong overlapping dysregulation in a locus of Tenascin XB (*TNXB*), next to highlighting both previously identified and novel epigenetic loci that we hypothesize to play a pivotal role in the early development of AD. Furthermore, we demonstrated, for the first time, that AD-associated dysregulation of bisulfite methylation patterns in the DRN, including that of *TNXB*, is dependent on both the disease phenotype and the cell type analyzed, which emphasizes the need for future cell type-specific neuroepigenetic studies in AD.

Materials and methods

Subjects and samples

EWAS discovery and pyrosequencing validation cohorts

The post-mortem DRN and LC bulk tissue samples that were used in the discovery EWAS and pyrosequencing validation analysis were obtained from the MRC London Brain bank for Neurodegenerative Disease (London, UK). Samples were provided with informed consent according to the Declaration of Helsinki (1991) and ethical approval for the study was provided by the NHS South East London REC 3. In total, 94 DRN samples and 81 LC samples from both AD patients and neurologically healthy controls were obtained and used for the EWAS. These two cohorts were derived from a total number of 109 individual subjects, meaning that 66 (60.55%) of them had matched tissue samples for both brainstem regions, thereby contributing to both cohorts. Upon screening for AD neurofibrillary pathology at autopsy, subjects were distributed to the relevant Braak stage ranging from Braak stage 0 to Braak stage VI. The tangle burden is described in a range from no AD pathology (Braak 0), to primary pathology in early-affected regions such as the brainstem (Braak I-II), to mid-stage subcortical pathology (Braak III-IV) and to severe widespread AD pathology (Braak V-VI). All samples were dissected by trained specialists, snap frozen and stored at -80°C up until further processing. For each sample, genomic DNA (gDNA) was extracted from ~50 mg of tissue using a standard phenol-chloroform extraction

method. For validation purposes, 62 and 52 identical isolated gDNA samples from a subset of these DRN and LC samples, respectively, were used for bisulfite pyrosequencing analysis of *TNXB*. Demographic and relevant information about the samples and cohorts are provided in Table 1.

Cell-specific DRN validation cohort

The post-mortem DRN tissue samples that were used in the cell-specific replication analysis were obtained from the Banner Sun Health Research Institute (Sun City, Arizona, USA). Age-matched DRN tissue samples from 12 female subjects without- (Control cases) and 12 female subjects with a clinical diagnosis of AD (AD cases) were collected. Each subject and her respective relative(s) consented to brain autopsy for the purpose of scientific research as part of the Banner Sun Health Research Institute Brain and Body Donation Program (BBDP) [22]. Written informed consent for autopsy was obtained in compliance with institutional guidelines. The institutional review board approved the entire study, including the recruitment, enrollment and autopsy procedures. A final diagnosis of AD at autopsy was made by following the National Institutes of Health (NIH) AD Center criteria [22]. Comorbidity with any other type of dementia, mild cognitive impairment, cerebrovascular disorders and presence of non-microscopic infarcts were applied as exclusion criteria. After autopsy, brain samples were snap frozen and stored at -80°C up until further processing. For demographic and relevant information about the studied samples, please refer to Table 1.

EWAS discovery analysis

Bisulfite and oxidative-bisulfite treatment

The TrueMethylTM 24 Kit version 2.0 by CEGXTM (Cambridge Epigenetix Limited, Cambridge, UK) was used for bisulfite and oxidative bisulfite conversion of gDNA isolated from the DRN and LC bulk tissues. All laboratory procedures were performed at GenomeScan (GenomeScan B.V., Leiden, the Netherlands) and according to the manufacturer's instructions. Prior to the conversion, high molecular weight gDNA was quantified using a PicoGreen assay (Invitrogen, Carlsbad, California, USA) and the gDNA quality was assessed by gel-electrophoresis, confirming that all samples were of sufficient quantity and quality. From each subject and brainstem region, 1 µg of gDNA was used, which, after purification and denaturation, was split up into two samples that underwent either the DNA oxidation or a mock DNA oxidation, for the oxidative bisulfite and bisulfite treated samples, respectively. Subsequently, all samples were bisulfite converted and the yields of the (oxidative) bisulfite DNA were assessed by a Qubit ssDNA assay (Invitrogen). An additional restriction quality control was performed for a

qualitative assessment of 5-hmC oxidation and bisulfite conversion using the Fragment Analyzer.

Illumina Infinium MethylationEPIC BeadArray

From each bisulfite- and oxidative bisulfite-treated DNA sample, 8 µL was amplified and hybridized on the Infinium MehtylationEPIC BeadChip (EPIC array; Illumina, Inc., San Diego, CA, U.S.A.). All samples were randomized with respect to sex and Braak stage in order to avoid batch effects and were processed in batches of 8 per BeadChip. Bisulfite- and oxidative bisulfite-treated samples from the same individuals were run on the same chip to avoid batch effects. Illumina iScan was used for imaging of the array. Sample preparation, hybridization, and washing steps were performed according to the manufacturer's instructions.

Pyrosequencing validation analysis

Bisulfite treatment

The EZ-96 DNA Methylation-Gold Kit (D5008; Irvine, CA, USA) was applied for bisulfite conversion of an identical subset of gDNA samples used in the EWAS (*i.e.* pyrosequencing validation cohort), which were isolated from both the DRN and LC bulk tissues. Per sample, 200 ng of gDNA was processed according to the manufacturer's instructions and all samples were bisulfite converted simultaneously per brainstem region in order to avoid batch effects. Finally, bisulfite-converted samples were eluted in 20 µl of elution buffer and 1 µl was used for polymerase chain reaction (PCR) amplifications, and subsequent bisulfite pyrosequencing.

Bisulfite pyrosequencing

Pyrosequencing was used to quantify the bisulfite methylation (5-mC+5-hmC) levels across eight individual *TNXB* CpG sites, including cg10365886, cg14188106 and cg07524919, spanning from 32063869 to 32063940 within chromosome 6 (Ensembl GRCh37 assembly). A single amplicon of 209 bp (6:32063788-32063996) was PCR amplified using primers designed with the PyroMark Assay Design software 2.0 (Qiagen, Hilden, Germany) (Supplementary Table 13). PCRs were performed with an initial denaturation step at 95°C for 5 minutes, followed by 45 cycles at 95°C for 30 seconds, 62°C for 30 seconds and 72°C for 30 seconds, with a final extension step at 72°C for 1 minute. All PCR reactions, contained 2.5 µl PCR buffer (10X) with 20 mM MgCl2, 0.5 µl 10 mM dNTP mix, 1 µl of each primer (5 µM stock) and 0.2 µl (5 U/µl) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) in a total volume of 25 µl. Pyrosequencing of the amplicon was performed using a single sequencing primer covering the 72-bp region. DNA bisulfite methylation was quantified in the validation cohort using the PyroMark Q48 Autoprep system

(Qiagen) and the Pyro Q48 Autoprep 2.4.2 software following the manufacturer's instructions. The sensitivity of the assay was initially assessed using methylated and unmethylated DNA standards from the EpiTect PCR Control DNA Set (Qiagen).

Cell-specific DRN validation analysis

Laser capture microdissection

Frozen DRN tissue sections of 10 µm from were mounted onto polyethylene naphthalate (PEN) slides and fixed in ice-cold 50% acetone/50% ethanol solution for 5 minutes on ice. Sections were washed in ice cold phosphate buffered saline (PBS), blocked in 1% hydrogen peroxide for 2 minutes, followed by 3 quick submersions in ice-cold PBS. Sections were then placed in a dilution of primary antibody against serotonin (5-HT; Abcam, ab66047) in PBS for 10 minutes at room temperature. After the incubation, sections were washed three times in PBS and incubated with avidin-biotin complex in PBS for 10 minutes at room temperature. Next, sections were washed three times in 50 mM Tris buffer and immersed in 3.3'-diaminobenzidine (DAB) solution (9.3 ml 50 mM Tris; 200 µl DAB (5 mg/ml); 500 µl saturated nickel; and 4 µl of 1% hydrogen peroxidase) for 5 minutes, followed by two quick rinses in 50mM Tris to stop the reaction. All sections were stored at -80ºC until further processing. 5-HT is a monoamine neurotransmitter that is specifically expressed by serotonergic neurons [23]. For this reason, LCM of serotonergic neurons from the DRN sections was performed based on 5-HT-immunoreactivity. In brief, sections were dipped in 100% ethanol, allowed to dry, and loaded onto a Leica AS-LMD LCM microscope (Leica, Wetzlar, Germany). Single serotonergic neurons were cut and then dropped into an inverted microcentrifuge cap containing 10 µl of Tris-EDTA (TE) buffer. Per individual subject, 150 serotonergic neurons were captured at 20X magnification and divided in small pools of 50 cells per microcentrifuge tube, *i.e.* three pools of 50 neurons per subject. For the non-serotonergic cells, 50 regions of similar single cell-size were dissected in areas free of 5-HT-immunoreactivity and collected as described above in 10 µl of TE buffer. In each of the experimental groups, two sections rendered unusable for additional isolation of non-serotonergic cells and were therefore excluded from further analysis (Non-serotonergic cells: AD, n = 10; Control, n = 10). All cells isolated with LCM were stored at -80°C until further processing.

DNA isolation and bisulfite treatment

Genomic DNA from a pool of serotonergic neurons or non-serotonergic cells was isolated and bisulfite-converted using the EZ DNA Methylation-Direct Kit (Zymo Research, Irvine, CA, USA) with the following adjustments. In brief, 1 µl of proteinase K (20 µg/µl) and 11 µl of M-Digestion buffer (2X) were added to a

microcentrifuge tube containing the cells and incubated overnight at 50°C. Subsequently, the complete lysate was transferred to a PCR tube and 143 µl of bisulfite conversion reagent was used to wash out the digestion tube before adding it to the sample. Bisulfite conversion was performed in a thermal cycler running at 98°C for 8 minutes and then at 64°C for 3.5 hours. A volume of 200 µl binding buffer was added to the spin column before loading the bisulfite-converted sample. The PCR tube used for bisulfite conversion was washed out twice by first adding 200 µl of binding buffer to the tube and then by transferring this volume to the sample-containing column. After centrifugation (10,000 x g; 30 sec), the column was washed with 100 µl washing buffer, incubated for 15 min with 200 µl desulfonation buffer and washed twice again with 200 µl washing buffer. The bisulfite-converted DNA was eluted in a single Eppendorf tube by running 20 µl of elution buffer through the column twice (2 times at 10,000 x g; 30 sec). Eppendorf LoBind microcentrifuge tubes (Merck KGaA, Darmstadt, Germany) and TipOne Low Retention Tips (STARLAB, Hamburg, Germany) with low affinity for DNA were used throughout the whole procedure. PCR amplifications were performed directly after elution of the bisulfite-converted DNA.

Polymerase chain reaction

The targeted bisulfite pyrosequencing assays for *TNXB* was based on a (semi-)nested PCR. Primers were designed with the PyroMark Assay Design 2.0 software (Qiagen), based on the Ensemble GRCh37 assembly (See Supplementary Table 13). Only for the LDBSP analysis, bisulfite-treated DNA from a pool of serotonergic cells was diluted to a single allele level by adding a PCR mixture with a capacity of 22 individual reactions to the sample (determined empirically). Each individual PCR reaction made use of 2.5 µl PCR buffer (10X) with 20 mM MgCl2, 0.5 µl 10 mM dNTP mix, 1 µl of each primer (10 µM stock) and 0.2 µl (5 U/µl) FastStart™ Taq DNA Polymerase (Roche Diagnostics GmbH, Mannheim, Germany) in a total volume of 25 µl. After adding the bisulfite DNA to the complete mixture, the sample was pipetted up-and-down in order to homogeneously disperse all bisulfite-converted DNA molecules throughout the solution and fractions of 25 µl were divided over 22 wells of a microtiter plate. A single amplicon of 440 bp (6:32063558-32063997) using the outer primers for *TNXB* was amplified based on an initial denaturation step at 95°C for 5 minutes, followed by 43 cycles at 95°C for 30 seconds, 56°C for 30 seconds and 72°C for 1 minute, with a final extension step at 72°C for 7 minutes. For each individual singleplex PCR reaction, 1 µl of the multiplex product was used as a template. A single amplicon of 432 bp (6:32063566-32063997) was amplified for *TNXB* using the inner primers using the same PCR compound concentrations and under similar thermocycler conditions as described above, although with 45 cycles and

an annealing temperature of 58°C. Bisulfite-treated DNA derived from pools of non-serotonergic cells was also amplified using (semi-)nested PCRs (*i.e.* extended pre-amplification); although with omitting the step of over-diluting the bisulfite converted DNA pool to single alleles. All PCR settings, as well as subsequent bisulfite pyrosequencing procedures, were performed under the exact same conditions for both of the analyzed cell types.

Bisulfite pyrosequencing

Pyrosequencing was used in order to determine the bisulfite methylation (5-mC+5-hmC) status of the Braak-associated region in *TNXB* in both the serotonergic neurons and non-serotonergic cells of the DRN. For *TNXB*, the DNA bisulfite methylation status was quantified over a region spanning from 32063774 to 32063913 within chromosome 6 (Ensemble GRCh37 assembly), including six Illumina probes: cg10365886, cg07524919, cg14188106, cg26266427, cg01337207 and cg02989255. The PyroMark Q96 MD pyrosequencing system (Qiagen) with the PyroMark Gold Q96 CDT reagent kit (Qiagen) were used according to the manufacturer's instructions. Methylation levels at a single CpG site resolution were quantified with the Pyro Q-CpG 1.0.9 software (Qiagen). The sensitivity of both assays was tested as described earlier using (un)methylated DNA standards from the EpiTect PCR Control DNA Set (Qiagen). The assay for *TNXB* was based on three pyrosequencing primers in order to maximize CpG site coverage over the targeted region (See Supplementary Table 13).

Data processing and statistical analyses

EWAS quality control and data processing

All computations and statistical analyses were performed using R version 3.3.2 [24] and Bioconductor version 3.5 [25]. Raw signal intensities generated from both bisulfite (BS) and oxidative bisulfite (oxBS) treated samples for each brain region were used to construct *MethylumiSet* object using 'readEPIC' function in *wateRmelon* package [26] and *RGChannelSet* objects using 'read.metharray.exp' function in *minfi package* [27]. We confirmed that samples that were treated for bisulfite and oxidative bisulfite conversion from the same identifiers were sourced from the same DNA samples using 59 single nucleotide polymorphism (SNP) probes on the array. Probes with common (minor allele frequency (MAF) > 5%) SNPs in the CG or single base extension position or cross-reactive probes were flagged and discarded. A principal component analysis (PCA)-based approach was used to examine a potential mismatch between reported and predicted sex. Using the 'pfilter' function within the *wateRmelon* package samples with a detection $p > 0.05$ in more than 5% of probes, probes with more than three beadcount in 5% of the samples and probes having 1% of samples with a detection p value > 0.05 were identified and

removed. Next the BS and oxBS datasets were split, and by using 'preprocessRaw' function the red and green channels for an Illumina methylation array were converted into methylation signals followed by a *Noob* background correction method with dye-bias normalization [28]. In order to estimate the proportion of DNA modifications, the maximum likelihood (ML) method described by Qu *et al.* [29] was used. The MLML function within the *MLML2R* package [30] was applied, which uses combined methylated and unmethylated signals from BS and oxBS arrays as an input, and returns the estimated proportion of 5-uC, 5-mC, and 5-hmC for each CpG site. Only probes with a mean beta value > 0.1 on both the bisulfite- and oxidative-bisulfite EPIC arrays were included in the MLML method. Finally, for 5-hmC, we only used sites in our analysis when 5-hmC was present in more than half of the sample population.

Differentially modified position analyses

In the DRN and LC EWAS, differentially methylated, hydroxymethylated, and unmodified cytosine positions (*i.e.* DUPS, DMPs and DHPs) associated with Braak stage were identified using linear regression models for each individual CpG probes. All analyses were adjusted for age, sex, and surrogate variables. Surrogate variable analysis (SVA) was performed in order to determine and estimate variation stemming from unknown sources including the cell type compositions. The identified surrogate variables (SVs) were correlated with variables known to affect methylation status, *i.e.* age, sex, post mortem interval, and the first five surrogate variables were included as covariates in the model along with age and sex. The *p* values abstained from EWAS were assessed using inflation index lambda (λ) value and the respective QQ plot. The *p* values were adjusted for multiple testing using the Benjamin-Hochberg, false discovery rate (FDR) procedure. Probes with nominal *p* values < 1.0E-3 were ordered based on a combined *p* values and regression estimate ranking. Illumina EPIC array probes were annotated using the Illumina UCSC gene annotation (Ensemble GRCh37 assembly).

Differentially modified region analyses

In order to identify differentially methylated, hydroxymethylated and unmodified regions (DMR, DHR, DUR) in our EWAS data, *comb-p* tool was used [31] with a distance of 500bp and a seeded *p* value of 1.0E-3.

Gene ontology enrichment analysis

For the top-ranked DUPs, DMPs and DHPs in DRN and LC, underlying biological processes and pathways were examined using a Gene Ontology (GO) enrichment analysis. Analysis were performed using the *missMethyl* package

[32], which takes into account the different number of probes per gene on the array.

Pyrosequencing validation data analysis

For the pyrosequencing validation analysis of *TNXB*, CpG bisulfite methylation percentages that passed quality control were exported from the Q48 Autoprep 2.4.2 software to the R statistical environment. Subsequently, a linear regression analysis per brainstem region was performed with the bisulfite methylation signals per CpG site as outcome, Braak Stage as predictor, and with age and sex added as covariates in the model. Cases with missing CpG bisulfite methylation values after pyrosequencing (DRN, n = 2; LC, n = 2) were excluded pairwise. In addition, a Pearson's correlation analysis was performed between the bisulfite methylation values obtained by bisulfite pyrosequencing and the EPIC array for each of the three Illumina probes in *TNXB*, *i.e.* cg10365886, cg14188106 and cg07524919. For cg07524919 in the LC, one case was excluded due to missing bisulfite methylation values for this CpG site after pyrosequencing.

Limiting dilution bisulfite pyrosequencing

For LDBSP on the serotonergic neurons derived from the DRN, sequenceable amplicon-yielding reactions for *TNXB* after the second PCR step were assessed for their representation of single-, two or three-alleles based on the CpG methylation readouts, as described previously [33]. In brief, thresholds of (1) ≤8.33% and ≥91.33%, (2) 50±8.33%, and (3) 33.33±8.33% and 66.66±8.33%, were set for the CpG methylation values. Accordingly, individual CpG sites that fell within the first, second and third threshold range were called and considered indicative for the potential presence of one, two and three alleles, respectively. A definitive (total) allele score for each individual product was then assigned following a multi-step filtering process that was based on the following criteria. All products solely displaying binary CpG methylation patterns, *i.e.* when every CpG site within an amplicon displays methylation levels within the first threshold range, were directly considered to be derived from single allele reactions, as the CpG methylation profiles for these products displayed a strong, typical binary pattern that is expected for a single DNA molecule [33, 34]. All products displaying CpG methylation values that fell only in the first and second, but not third, or only in the first and third, but not second, threshold ranges were scored as two or three alleles, respectively. Subsequently, products containing CpG methylation values that fell outside of the assigned threshold ranges, *i.e.* between 8.33%-25% and 75%-91.33%, and/or products with values that were indicative of both two and three alleles simultaneously, were assessed. All of these products were therefore thoroughly inspected by two investigators that were blinded to the experimental conditions, and a decision on the total number of alleles present in each individual

reaction, *i.e.* one, two or three alleles, was made independently, while taking into account a combination of different factors. These included, but were not limited to, small technological variation that was previously observed during sensitivity testing of the assays, the directionality and methylation status of other CpG sites in the same product and the total number of dominant allele indicators, *i.e.* whether a product demonstrated more or less suggestive CpG sites for either two or three alleles. Furthermore, a likelihood estimation for each CpG site was made by taking into account the methylation status frequency on other gene-specific products obtained from the same individual, as well as from identical products obtained from other individuals. A cross comparison between the independent score sheets was then performed (98.31% overlap in scoring) and reactions with a deviating score between the first two investigators were assessed by a third (blinded) investigator. A final allele number was then assigned for these reactions based on the overlap between the score sheets of the third and first two investigators, *i.e.* when two out of the three investigators assigned the same score then this allele number was used for the respective reaction. Overall, it was estimated that from the 237 (C = 104, AD = 133) PCR products, 91.56% (C = 93.27%; AD = 90.23%) were derived from a single allele, whereas 5.49% (C = 4.81%; AD = 6.02%) were derived from two alleles, and 2.95% (C = 1.92%; AD = 3.76%) from three alleles (Supplementary Table 14). By taking into account these multi-allele reactions, it was therefore estimated that in total 264 alleles (C = 113; AD = 151) were recovered with a recovery rate of 3.67% (C = 3.14%; AD = 4.19%). For each subject, the methylation rate for each CpG site was determined by expressing the number of methylated CpG sites as a percentage of the total number of CpG sites over the estimated alleles, whilst correcting for the number of alleles present in each of the reactions, *i.e.* one, two or three alleles.

General linear model

For the cell-specific DRN validation analysis, average bisulfite methylation levels across the targeted region of *TNXB* were calculated in both the serotonergic neurons and non-serotonergic cells. In the serotonergic neurons, the bisulfite methylation rates obtained for the targeted CpG sites in both genes were first averaged over the region per subject and subsequently over the experimental groups. For the non-serotonergic cells, only cases for which all CpG sites surpassed quality control by the Pyro Q-CpG 1.0.9 software were included and averaged per subject, as well as the experimental groups. One AD case and one control case were excluded from the analysis for *TNXB* (Non-serotonergic cells, *TNXB*; AD, n = 9; C, n = 9), as these did not met the quality control criteria. A general linear model with experimental condition (AD- and control cases) as between-subject factor and cell type (Serotonergic neurons and non-serotonergic

cells) as within-subject factor was conducted in order to test for interaction and main effects. A p value < 0.05 was considered as statistically significant.

Table 1. Demographics of studied cohorts

	EWAS (Discovery)		**Pyrosequencing (Validation)**		**LDBSP and Pyrosequencing (Replication)**	
	DRN	**LC**	**DRN**	**LC**	**5-HT+**	**5-HT-**
N	94	81	62	52	12 AD; 12 C	10 AD; 10 C
Sex (F/M)	53/40*	46/35	33/29	30/22	24/0	20/0
Age of death; Mean (± SD)	83.85 (9.04)	83.96 (9.53)	84.90 (9.16)	84.50 (10.13)	81.21 (10.21)	81.35 (11.31)
PMI; Mean (± SD)	32.66 (20.04)	34.69 (22.94)	43.87 (89.68)	34.69 (22.94)	2.69 (0.82)	2.77 (0.87)
Braak Stage; Range (Median)	0-VI (V)	0-VI (V)	0-VI (V)	0-VI (V)	I-VI (IV)	I-VI (IV)

Displayed are the demographics of the patient cohorts that were used in the present study. Dorsal raphe nuclei (DRN) and locus coerulues (LC) bulk tissues used in the discovery epigenome-wide association study (EWAS) and pyrosequencing validation study were obtained from the Medical Research Council (MRC) London Brainbank for Neurodegenerative Disease (London, UK). Serotonergic (5-HT+) neurons and non-serotonergic (5-HT-) cells isolated from the DRN by means of laser capture microdissection (LCM) used in the replication study were obtained from the Banner Sun Health Research Institute (BSHRI; Sun City, Arizona, USA) Brain and Body Donation Program (BBDP). For all cohorts, the number of samples in each group, the distributions of sex (F, female; M, male), age of death (mean ± standard deviation (SD)), postmortem interval (PMI; in hours) and Braak stage are shown. * = From one individual sex was not specified.

Results

Differentially modified positions associated with Braak stage in the DRN and LC

In the first stage of our analysis, epigenetic profiles at the level of 5-uC, 5-mC and 5-hmC were quantified in the DRN and LC in over 850.000 single CpG sites. For this purpose, we used largely matched bulk tissue samples of the DRN and the LC from 94 and 81 donors, respectively, which were collected from the Medical Research Council (MRC) London Brain bank for Neurodegenerative Disease (see Table 1 for the demographics of our studied cohorts). Our discovery cohorts in both brainstem regions represented the entire spectrum of AD pathology defined by Braak stage, which is a standardized measure of neurofibrillary tangle burden determined at autopsy [3]. In these subjects, we focused at identifying Braak stage-associated epigenetic alterations. As such, we aimed at capturing alterations that might be associated with both the early development and the course of AD. Thus, the acquired array data derived from both brainstem regions was pre-processed, normalized and subjected to quality control, after which linear regression analyses were conducted using either the 5-uC, 5-mC or 5-hmC signals as outcome and Braak stage as predictor. All analyses were performed

whilst controlling for sex and age, with surrogate variables added to the model as appropriate to correct for unmeasured confounding factors (see Materials and methods).

For the DRN, the top-ranked differentially unmodified positions (DUPs), differentially methylated positions (DMPs) and differentially hydroxymethylated positions (DHPs) are shown in Supplementary Tables 1-3, respectively. None of the CpG sites identified for each of the three cytosine states reached the false discovery rate (FDR) threshold for experiment-wide significance that has been established for the Illumina Infinium MethylationEPIC BeadChip array, *i.e.* $p < 9.0E-08$ [35]. The displayed nominally significant probes with a cut-off of $p < 0.001$ were ranked based on a combined *p* value and regression estimate ranking, resulting in 1029 DUPs, 647 DMPs and 831 DHPs in the DRN. For the LC, the top-ranked Braak-associated DUPs (n=611), DMPs (n=777) and DHPs (n=741) are provided in Supplementary Tables 4-6, respectively. Similar to the DRN, none of the individual CpG sites identified in the LC at the level of 5-uC, 5-mC and 5-hmC surpassed the FDR threshold for experiment-wide significance.

Most of the top-ranked CpG sites identified in the DRN and the LC represented unique gene-associated probes for each of the three cytosine states. However, we observed numerous nominally significant DUPs in both brainstem regions that were annotated to *TNXB*. The six probes in the DRN, *i.e.* cg19455923 (regression estimate (RE) = -110.1127, $p = 1.74E-04$), cg17832639 (RE = 34.49144887, $p = 3.25E-04$), cg14196170 (RE = 14.71958219, $p = 2.93E-05$), cg27387193 (RE = 11.28174141, $p = 2.76E-04$), cg05473289 (RE = 17.93969959, $p = 4.94E-04$) and cg16385684 (RE = -23.52117157, $p = 8.32E-04$), were dispersed throughout the gene. The six DUPs in the LC however, *i.e.* cg27387193 (RE = 10.54119093, $p = 2.51E-04$), cg10923662 (RE = 27.81154951, $p = 5.66E-04$), cg10890302 (RE = 27.80155724, $p = 8.13E-04$), cg00525277 (RE = 11.73343449, $p = 5.99E-04$), cg04753078 (RE = 15.07520956, $p = 8.34E-04$) and cg19267551 (RE = 11.55809141, $p = 9.16E-04$), were clustered within a 639 base pair (bp) window (6:32063619-32064258) in the gene body of *TNXB*. Interestingly, cg27387193, which overlapped in both brainstem regions, also demonstrated a similar RE, while cg14196170, which was identified in the DRN only, just lays 24 bp upstream (6:32063595) to the six probes identified in the LC. Moreover, the RE for each of the aforementioned DUPs that were identified within this region of *TNXB* (6:32063595-32064258) was positively associated with progressing Braak stage, suggesting that an increase in 5-uC in these loci is related to more advanced AD neuropathology. In the LC, we furthermore found cg04753078 (RE = -10.15576914; $p = 7.05E-04$) and cg14196170 (RE = -10.00801167; $p = 6.38E-05$) listed among the top-ranked DMPs, demonstrating a negative association

with Braak stage. When looking at the DHPs in the LC, also cg14196170 (RE = 12.14700315; *p* = 1.72E-04) showed up as top hit, displaying a positive association of 5-hmC with increasing AD neuropathology. Altogether, these findings suggest that active DNA demethylation of *TNXB* is typical with increasing Braak stage in both brainstem regions.

Table 2. Differentially modified regions in the dorsal raphe nuclei (DRN)

DURs (5-uC)										
Rank	**USCS annot.**	**Chr.**	**Start**	**End**	**Probes**	***p* value**	**Šidák *p* value**	**Gene feature**	**CpG island feature**	**Assoc.**
1	PGLYRP1	19	46526321	46526652	9	3.19E-13	7.24E-10	TSS; 5'UTR; CDS	Island	↑
2	TNXB	6	32063607	32064033	11	2.63E-12	4.64E-09	CDS	Island	↑
3	CAST	5	95997186	95997356	4	3.35E-09	1.48E-05	Intergenic	Shore	↑
4	GNAT1	3	50230792	50230886	3	1.14E-08	9.09E-05	Intron; CDS	Island	↑
5	RBMXL2	11	7110074	7110149	5	1.07E-08	1.07E-04	Intergenic	Island	↑
DMR (5-mC)										
Rank	**USCS annot.**	**Chr.**	**Start**	**End**	**Probes**	***p* value**	**Šidák *p* value**	**Gene feature**	**CpG island feature**	**Assoc.**
1	MALAT1	11	65266482	65266562	4	7.03E-09	6.61E-05	NC exon	Shore	↑

Displayed are the differentially unmodifed regions (DURs) and the differentially methylated region (DMR) associated with Braak stage in the dorsal raphe nuclei (DRN). For each region, the Illumina gene annotation (UCSC annotation), the chromosomal position and coordinates (Ensembl GRCh37 assembly), the number of probes, the *p* value, the multiple testing corrected Šidák *p* value, the gene feature (TSS, transcription start site; 5'UTR, 5'untranslated region; CDS, coding sequence; NC, non-coding), the cytosine-phosphate-guanine (CpG) island feature and the association with increasing Braak stage are shown. All regions are ranked based on their Šidák *p* value with Šidák $p < 0.05$ as a cut-off.

Differentially modified regions associated with Braak stage in the DRN and LC

Subsequently, a regional analysis, looking at the spatial correlation of adjacent modified positions, was performed, using *comb-p* [31], in order to identify regions containing three or more neighboring DUPs, DMPs or DHPs that displayed a Šidák-corrected *p* value < 0.05 within a 500-bp sliding window. In the DRN, we identified five significant differentially unmethylated regions (DURs) and one differentially methylated region (DMR), which are listed in Table 2. For the LC, we identified three DURs and two DMRs, which are listed in Table 3. We then examined whether any of these hits were shared between the two brainstem regions. Strikingly, we observed a strong overlap in our findings within the coding sequence of *TNXB.* While the second most significant DUR in the DRN (Šidák-corrected *p* = 4.64E-09) consisted of a region spanning 426 bp (6:32063607-32064033) and 11 probes, in the LC, a larger but overlapping region in *TNXB* (6:32063726-32064259) spanning 533 bp and covering 18 probes showed up as

most significant DUR (Šidák-corrected p = 1.39E-17). Furthermore, in the LC, a nearly identical region (6:32063607-32064259) that consisted of 18 probes within a 652-bp window was identified as most significant DMR (Šidák-corrected p = 7.47E-20). All other significant DURs, DMRs and DHRs that were discovered in our regional analysis were specific to the individual brainstem regions.

Table 3. Differentially modified regions in the locus coeruleus (LC)

DURs (5-uC)										
Rank	**USCS annot.**	**Chr.**	**Start**	**End**	**Probes**	***p* value**	**Šidák *p* value**	**Gene feature**	**CpG island feature**	**Assoc.**
1	TNXB	6	32063726	32064259	18	1.39E-17	2.02E-14	CDS	Island	↑
2	ZFYVE28	4	2366555	2366673	4	1.65E-09	1.08E-05	Intron	Island	↑
3	ANKRD2	10	99338056	99338189	3	3.03E-08	1.76E-04	Intron; CDS	Island	↑
DMRs (5-mC)										
Rank	**USCS annot.**	**Chr.**	**Start**	**End**	**Probes**	***p* value**	**Šidák *p* value**	**Gene feature**	**CpG island feature**	**Assoc.**
1	TNXB	6	32063607	32064259	18	7.47E-20	8.86E-17	CDS	Island	↓
2	LOC105377777; DLGAP2; DLGAP2	8	833012	833090	3	4.45E-06	4.32E-02	NC exon; NC intron; Intron	Shore	

Displayed are the differentially unmodified regions (DURs) and differentially methylated regions (DURs) associated with Braak stage in the locus coeruleus (LC). For each region, the Illumina gene annotation (UCSC annotation), the chromosomal position and coordinates (Ensembl GRCh37 assembly), the number of probes, the p value, the multiple testing corrected Šidák p value, the gene feature (CDS, coding sequence; NC, non-coding), the cytosine-phosphate-guanine (CpG) island feature and the association with increasing Braak stage are shown. All regions are ranked based on their Šidák p value with Šidák $p < 0.05$ as a cut-off.

We then explored whether any of these differentially modified regions in our array data have been nominated in previous EWAS or whether these represented novel AD-associated hits. For the DRN, similar regions as identified for the DURs in *TNXB* and *PGLYRP1* have been previously annotated in EWAS targeting various regions of the AD cortex, including the entorhinal cortex, the superior temporal gyrus and the prefrontal cortex [11, 19]. The remaining DURs in *GNAT1*, *CAST* and *RBMXL2*, as well as the DMR in *MALAT1*, represented novel Braak stage-associated regional hits. For the LC, similar regions as the DUR and DMR in *TNXB,* as well as the DUR in *ZFYVE28*, have been previously associated with AD in the aforementioned cortical tissues [11], while the DUR in *ANKRD2* and the DMR in *LOC105377777*/*DLGAP2* have not been previously identified in other AD EWAS. Overall, based on the substantial overlap in both the DRN and the LC, as well as previous associations in the AD cortex, our findings strongly suggest that epigenetic dysregulation in *TNXB* may play an important role in both the early development and course of AD. Furthermore, our brainstem EWAS highlighted several other previously identified loci, as well as novel epigenetic

signatures, that might play a pivotal role in both the early pathogenesis and progression of AD.

Structural and functional genomic annotation enrichment analysis

Next, we were interested to identify biological, cellular and molecular pathways in the DRN and LC that were enriched among Braak stage-associated probes. For this purpose, we performed a Gene Ontology (GO) enrichment analysis in each brainstem region using the top-ranked DUPs, DMPs and DHPs. The top 20 enriched GO terms for each cytosine state in the DRN and the LC are displayed in Supplementary Figures 1A and 1B, respectively. Additionally, a more elaborated overview of the top 50 ranked GO terms, as well as corresponding identifiers, ontologies, the number of (differentially modified) genes, *p* values and FDR-adjusted *p* values for 5-uC, 5-mC and 5-hmC in both the DRN and LC can be found in Supplementary Tables 7-12.

For the DRN, we identified two pathways related to the DHPs that surpassed the FDR threshold for multiple testing correction (FDR-adjusted $p < 0.05$). The most significant GO term in our 5-hmC enrichment analyses was found in 'homophilic cell adhesion via plasma membrane adhesion molecules' (FDR adjusted p = 1.16E-03), demonstrating that from the 167 annotated genes, 27 were differentially modified. The second significantly enriched GO term that was associated with our 5-hmC data from the DRN-EWAS was found for 'calcium ion binding' (FDR-adjusted p = 1.09E-02), revealing that from the 690 genes, 56 were altered. All other GO terms identified in the DRN at the level of each cytosine state did not surpass the FDR threshold for significance.

In the LC, we identified four pathways that were significantly enriched among our DMPs after multiple testing correction. The most significant enriched GO term was identified in 'homophilic cell adhesion via plasma membrane adhesion molecules' (FDR-adjusted p = 7.25E-11), in which 35 of the 161 genes were differentially modified. The second most significant GO term was found for 'cell-cell adhesion via plasma-membrane adhesion molecules' (FDR-adjusted p = 1.19E-06) with 35 out of 161 genes that were altered. The third most significantly enriched pathway in our 5-mC LC-EWAS data was found for 'calcium ion binding' (FDR-adjusted p = 1.17E-03), revealing that the modification levels of 53 genes from a total of 167 were affected. Finally, the fourth significantly enriched GO term concerned 'cell-cell adhesion' (FDR adjusted p = 2.75E-03), in which 57 of 782 genes were differentially modified. The remaining GO terms identified in the LC for the DUPs, DMPs and DHPs did not survive the FDR threshold for multiple testing correction.

Table 4. Regression analysis during the validation of the *TNXB* Braak-associated region in the dorsal raphe nuclei (DRN).

DRN						
CpG	**Probe ID**	**RE**	**SE**	***t* value**	***p* value**	
1	cg10365886	-0.10599	0.02909	-3,644	5.75E-04	***
2	-	-0.10048	0.02656	-3,783	3.69E-04	***
3	cg14188106	-0.12461	0.03575	-3,486	9.43E-04	***
4	cg07524919	-0.12121	0.03006	-4,032	1.63E-04	***
5	-	-0.12586	0.03256	-3,866	2.82E-04	***
6	-	-0.09968	0.03033	-3,286	1.73E-03	*
7	-	-0.12871	0.03687	-3,491	9.36E-04	***
8	-	-0.12202	0.03827	-3,188	2.34E-03	*

The DNA bisulfite methylation patterns quantified by pyrosequencing over eight cytosine-phosphate-guanine (CpG) sites in *TNXB,* including cg10365886, cg14188106 and cg07524919, were all significantly associated with Braak stage in the DRN. Displayed for each CpG site is their respective number in the sequence, the Illumina probe identifier (probe ID – if applicable), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the *t*-statistics and accompanying *p* values. * = $p < 0.05$; *** = $p < 0.001$

Table 5. Regression analysis during the validation of the *TNXB* Braak-associated region in the locus coeruleus (LC).

LC						
CpG	**Probe ID**	**RE**	**SE**	***t* value**	***p* value**	
1	cg10365886	-0.0415	0.03159	-1,314	1.95E-01	
2	-	-0.0386	0.03387	-1,139	2.60E-01	
3	cg14188106	-0.04263	0.04395	-0.97	3.37E-01	
4	cg07524919	-0.05173	0.03744	-1,382	1.74E-01	
5	-	-0.04597	0.04051	-1,135	2.62E-01	
6	-	-0.05208	0.03993	-1,304	1.99E-01	
7	-	-0.08598	0.04086	-2,104	4.08E-02	*
8	-	-0.02311	0.04794	-0.482	6.32E-01	

The DNA bisulfite methylation pattern quantified by pyrosequencing for each of the eight cytosine-phosphate-guanine (CpG) sites in *TNXB,* including cg10365886, cg14188106 and cg07524919, was significantly associated with Braak stage for CpG site 7 in the LC. Displayed for each CpG site is their respective number in the sequence, the Illumina probe identifier (probe ID – if applicable), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the *t*-statistics and accompanying *p* values. * = $p < 0.05$

Pyrosequencing validation of *TNXB* in the DRN and LC

In order to validate our discovery findings from the brainstem EWAS, we compared our EWAS-derived data sets to bisulfite pyrosequencing data generated from a subset of the DNA samples obtained from the DRN and LC bulk tissues (Table 1). In these samples, the DNA bisulfite methylation status of *TNXB*

was quantified across a region of 72 bp (6:32063869-32063940), spanning eight CpG sites. The targeted region overlapped with the DUR in the DRN, and with both the DUR and DMR identified in the LC, including three Illumina probes that were part of these identified Braak-associated regions, *i.e.* cg10365886, cg14188106 and cg07524919. Importantly, at this stage of the analysis, we aimed at a technical validation of the observed Braak-associated alterations by assessing the level of bisulfite methylation, *i.e.* the cumulative measure of 5-mC and 5-hmC, which is inversely proportional to the measurements of 5-uC [29]. All the acquired pyrosequencing data were adjusted for sex and age, similar to the EWAS described above. In addition to the performed linear regression analyses in both individual brainstem regions, we furthermore conducted Pearson's correlation analyses between the bisulfite methylation values estimated by the two technologies; *i.e.* by the Illumina Infinium MethylationEPIC BeadChip array and by bisulfite pyrosequencing. For this validation analyses, a p value < 0.05 was considered statistically significant.

The results from the linear regression analyses on the *TNXB* pyrosequencing data derived from the DRN and the LC are provided in Tables 4 and 5, respectively. We identified a strong significant association with Braak stage for all eight CpG sites in the DRN (Lowest p = 1.63E-04) and for one CpG site in the LC (CpG site number 7, p = 4.08E-02). All analyzed *TNXB* CpG sites in both brainstem regions demonstrated hypomethylation with increasing AD neuropathology, which was indicated by the negative REs. The results from the correlation analyses in both the DRN and the LC are shown in Figure 1. The pattern of DNA bisulfite methylation in *TNXB* quantified during the EWAS of the DRN and the LC was identical to the pyrosequencing data in each of the brainstem regions, with a highly significant and strong correlation between the values estimated by the two independent technologies for all three analyzed probes (lowest p = 2.2E-16 with highest r = 0.84). Thus, despite the use of a smaller sample size for our validation approach (for reasons of limited tissue availability), we observed a robust validation of our EWAS discovery findings in this Braak-associated region of *TNXB*. In addition to the analyzed Illumina probes, we furthermore demonstrated that adjacent CpG sites in this gene support the notion of DNA hypomethylation with increasing Braak stage, both in the DRN, as well as in the LC. Altogether, these data provide compelling evidence for an association between active DNA demethylation in this *TNXB* locus and an increase of AD-related neuropathological features in both brainstem regions assessed.

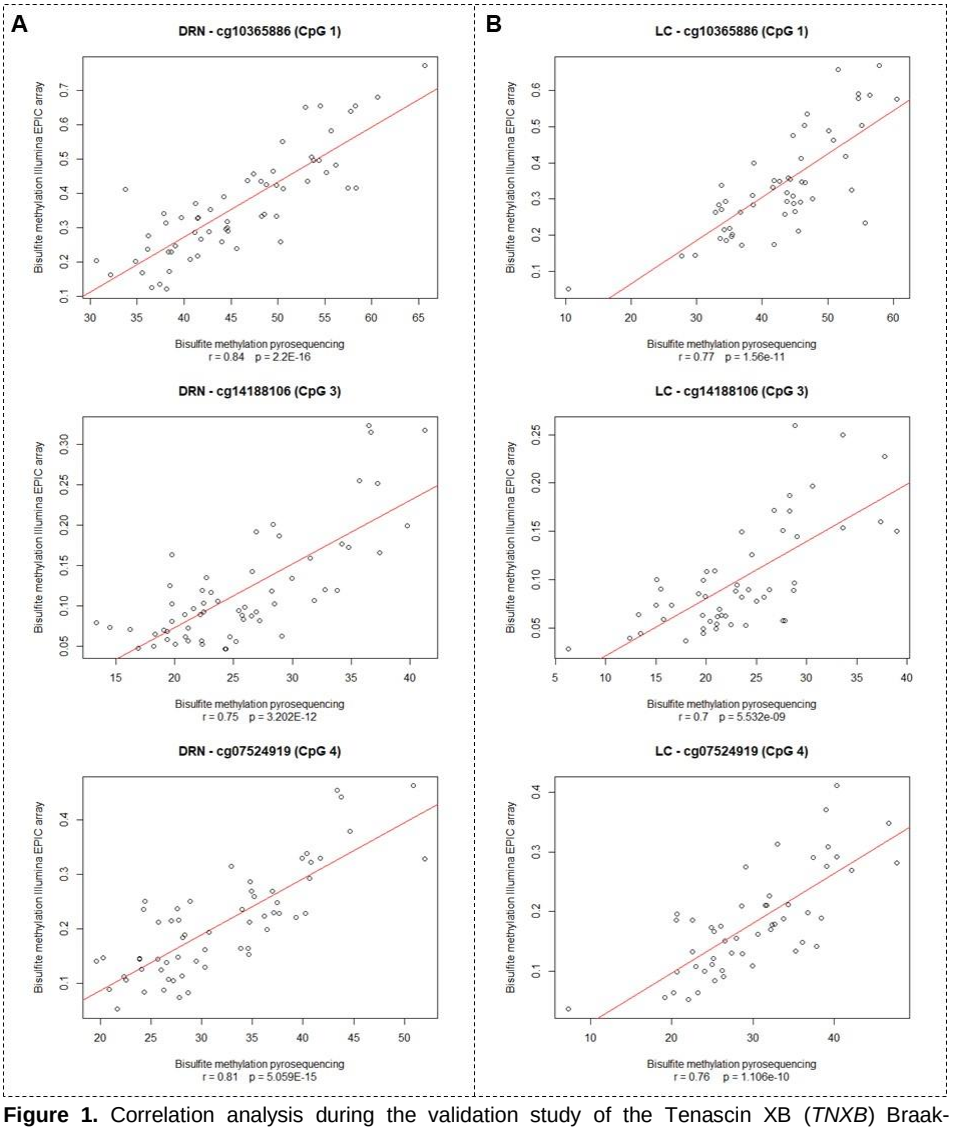


Figure 1. Correlation analysis during the validation study of the Tenascin XB (*TNXB*) Braak-associated region. DNA bisulfite methylation patterns quantified by the Illumina Infinium MethylationEPIC BeadChip array and pyrosequencing in *TNXB* were highly significantly and strongly correlated for cg10365886, cg14188106 and cg07524919 in both the dorsal raphe nuclei (DRN) and the locus coeruleus (LC) (**A** and **B**, respectively).

Cell type-specific analysis of *TNXB* in serotonergic and non-serotonergic cells in the DRN

In the second stage of our analysis, we aimed at validating our EWAS findings of *TNXB* in an independent patient cohort, an attempt which we combined with a novel cell type-specific and targeted epigenetic analysis in single DRN cells

isolated by LCM. This targeted approach was performed with the aim to define the cellular origin of the identified epigenetic signature. Given that bulk tissues were used in the DRN EWAS described above, it is expected that various cellular populations residing in this brainstem region, including serotonergic neurons and other parenchymal cells such as glia, contribute to the obtained epigenetic signals. As such, the cellular heterogeneity of these bulk tissue samples constitutes a major source of noise in epigenetic profiling approaches that represents a reoccurring challenge in AD EWAS [36]. Moreover, since AD is characterized by neuronal loss and increased microglia activation, differences in cellular composition between (brainstem) tissue samples obtained at varying stages of the disease can interfere with the interpretation of the acquired epigenetic data. In order to tackle this issue and based on the hypothesized role of the serotonergic neurotransmitter system in mediating affective symptoms linked to AD, we therefore first isolated single serotonergic neurons from the DRN by means of LCM, using an immunohistochemical labeling of serotonin, *i.e.* 5-HT-positive (5-HT+) cells . For this purpose, post-mortem DRN tissue derived from 12 subjects without- (Control cases; Braak I-IV) and 12 subjects with a clinical diagnosis of AD (AD cases; Braak V-VI) were utilized (Table 1), which were all part of the Banner Sun Health Research Institute Brain and Body Donation Program (BBDP; [22]).

Per individual, a total of 150 serotonergic cells divided in three pools of 50 cells were isolated with LCM from the DRN tissue. Limiting dilution bisulfite pyrosequencing (LDBSP) was then applied on each pool of cells to analyze the bisulfite methylation status of the Braak-associated region in *TNXB* on single alleles [34]. This type of adapted pyrosequencing allows for assessing DNA methylation patterns in only a single or a few cells and hence is very suited for low cell numbers as acquired with LCM. As such, the DNA bisulfite methylation status in *TNXB* was quantified in the serotonergic neurons across a region of 140 bp (6:32063774-32063913) spanning eleven CpG sites. This region overlaps with the DUR identified in the DRN, both the DUR and DMR identified in the LC, as well as the region targeted during our EWAS pyrosequencing validation analysis. From the eleven CpG sites in the sequenced region, six represented Illumina probes that were also part of the Braak-associated differentially modified regions, *i.e.* cg10365886, cg07524919, cg14188106, cg26266427, cg01337207 and cg02989255. Considering the smaller sample size and the lack of access to multiple Braak 0, III and IV cases, the LDBSP-derived CpG methylation data were analyzed using a case-control model based on clinical diagnosis, rather than looking for an association with Braak stage.

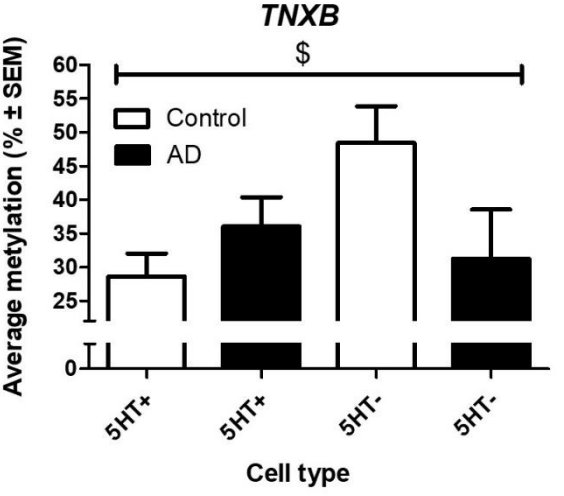


Figure 2. General linear model during the cell type-specific validation study in the dorsal raphe nuclei (DRN) of the Braak-associated region in Tenascin XB (*TNXB*). Average DNA bisulfite methylation levels quantified in both the serotonergic (5-HT+) neurons and non-serotonergic (5-HT-) cells derived from the DRN of Alzheimer's disease (AD) patients and healthy controls (Control) are shown for *TNXB*. A general linear model with experimental condition as a between-subject factor (AD/Control) and cell type (5-HT+/5-HT-) as a within-subject factor revealed a significant interaction effect between cell type and experimental condition. The AD-associated bisulfite methylation profiles were exactly opposite in the serotonergic neurons when compared to the non-serotonergic cells, the latter which resembled our EWAS data in the DRN. For the interaction effect: $ = $p < 0.05$.

Remarkably, by applying LDBSP, we observed a trend towards AD-associated bisulfite hypermethylation in the serotonergic neurons, *i.e.* a pattern opposite to that identified in the EWAS of the DRN (Figure 2). Accordingly, we then hypothesized that the EWAS-associated *TNXB* hypomethylation should originate from other, *i.e.* non-serotonergic, cells in the DRN bulk tissue. In order to confirm this hypothesis, we isolated additional, *i.e.* 5-HT-negative (5-HT-), cells by LCM from the same BBDP DRN tissue samples used in the 5-HT+ LDBSP analysis (AD = 10; C = 10), but now isolating tissue that was free of 5-HT immunoreactivity. As such, this isolated tissue most likely consisted of non-serotonergic parenchymal cellular populations, including glia cells, glutamatergic, GABAergic and dopaminergic neurons, amongst others [37]. Importantly, isolation of these cells based on the absence of 5-HT immunoreactivity in a limited amount of leftover DRN tissue sections did not allow us to determine the cell numbers that were isolated and, therefore, the average bisulfite methylation status over all residing non-serotonergic cells in the exact same region of *TNXB* was quantified by using regular bisulfite pyrosequencing with an extended pre-amplification protocol (see Materials and methods).

Subsequently, average *TNXB* bisulfite methylation levels were then compared between both experimental groups using a general linear model with experimental condition (AD/Control) as a between-subject and cell type (5-HT+/5-HT-) as a within-subject factor. Strikingly, when comparing the bisulfite methylation levels of *TNXB* in individually isolated serotonergic neurons with those of non-serotonergic cells in the DRN, we found a significant interaction effect between cell type and experimental condition ($p = 0.046$; Figure 2), with AD-associated bisulfite methylation profiles being exactly opposite in serotonergic neurons when compared to non-serotonergic cells. More specifically, while bisulfite methylation levels tended to be increased in serotonergic neurons derived from AD patients, bisulfite hypomethylation was observed in non-serotonergic cells from AD patients when compared to healthy controls – the latter of which resembled the initially acquired EWAS data in the DRN bulk tissue. Overall, these data corroborated our previous findings in the DRN-EWAS, indicating that epigenetic dysregulation in *TNXB* in AD is likely attributed to non-serotonergic cells within this brainstem nucleus, a finding that is indirectly supported by the notion that the LC, which is free of serotonergic cells, also displayed *TNXB* hypomethylation with advancing Braak stage.

Discussion

The present study reflects the first comprehensive epigenetic analysis in the brainstem of AD to date, targeting both the DRN and the LC simultaneously. In our AD discovery cohort, two independent EWAS on bulk tissue derived from both brainstem regions were performed, with the aim of examining potential epigenetic differences at the level of 5-mC, 5-hmC and 5-uC accompanying AD-associated pathology. In total, 1029 nominal significant DUPs, 647 DMPs and 831 DHPs were identified in the DRN, whereas for the LC, 611 DUPs, 777 DMPs and 741 DHPs were found. In a subsequent regional analysis we then identified significant overlapping, both previously annotated and novel, differentially modified regions that were associated with Braak stage in both the DRN and LC. We furthermore followed up the EWAS in both brainstem regions with a technical validation of the acquired array data using bisulfite pyrosequencing in a subset of samples derived from the same patient cohort. Moreover, our brainstem discovery EWAS was complemented with a novel cell type-specific and targeted analysis in the DRN, assessing methylation signatures in laser-captured serotonergic neurons and non-serotonergic cells derived from an independent AD patient cohort. As such, we aimed at validating our EWAS data and, simultaneously, to unravel potential cell-type specific contributions of the identified epigenetic signatures within the DRN.

The regional analysis in the DRN and the LC revealed a strong overlapping dysregulation at the level of 5-uC and 5-mC in a gene-coding region of *TNXB*. Subsequent bisulfite pyrosequencing of this locus in the same patient cohort confirmed the notion that active DNA demethylation in *TNXB* was associated with progressing AD neuropathology in both brainstem regions assessed. Interestingly, dysregulation at the level of DNA methylation in a similar region of *TNXB* has previously also been annotated in an AD EWAS targeting cortical tissues, including the entorhinal cortex, superior temporal gyrus and prefrontal cortex, which were derived from the same patient cohort as used in the present study [11]. Next to epigenetic variation, genetic variation in *TNXB* has also been directly associated with a risk of developing AD [38-40], as well as with progressive supranuclear palsy [41], which is a tau-related disorder. *TNXB* expresses a glycoprotein with anti-adhesive properties, but its exact physiological role in the brain is not yet clear [42]. In addition to AD, epigenetic dysregulation in a similar locus of *TNXB* has also been identified in EWAS targeting peripheral blood samples of patients and people at risk of developing several stress-related disorders, including anorexia nervosa, schizophrenia, remitted major depressive disorder and bipolar disorder [43-45]. These are interesting observations, given the established link between stress-related pathology and AD, especially when considering the physiological role of the DRN and the LC in the stress response. Of note, we also performed a methylation quantitative trait locus (mQTL) analysis in order to examine whether the previously identified genetic variation could explain the observed methylation differences in our EWAS, but could not identify such an effect (data not shown), suggesting that the observed alterations represent a true epigenetic AD-associated signature. Evidently, more research on the exact function of TNXB, the interplay between genetic and epigenetic variation in *TNXB*, and exploring the exact cause-effect relationship underlying this AD neuropathology-associated epigenetic signature, is vital in order to develop a better understanding of the exact role of this gene in the disease.

In addition to *TNXB*, a DUR in the DRN that has previously been nominated in other AD EWAS was identified for *PGLYRP1*. The same gene displayed differential methylation in the aforementioned cross-cortex analysis using the same AD cohort [11]. Furthermore, a similar locus has been reported to demonstrate differential methylation in peripheral blood samples derived from individuals with down syndrome, a patient population known to be at increased risk of developing AD [46]. *PGLYRP1* encodes an innate immunity protein that is a known activator of TREM-1 [15]. Interestingly, studies have demonstrated that overexpression of TREM-1 in APP/PSEN1 mice, a mouse model for AD, facilitates microglial-mediated amyloid beta clearance and restores AD-related cognitive impairment, emphasizing the importance of the PGLYRP1-TREM-1

interaction in the pathophysiology of AD [16]. However, whether there is a functional connection between the reported epigenetic alterations in *PGLYRP1* and TREM-1 efficacy remains to be elucidated. For the LC, a DMR was identified in *ZFYVE28* in a region that has previously been associated with AD in cortical tissue [11]. ZFYVE28 is a regulator of epidermal growth factor receptor (EGFR) signaling that acts by promoting EGFR degradation in endosomes when not mono-ubiquitinated [47]. Interestingly, several recent studies have demonstrated a beneficial effect of EGFR inhibitors in reducing amyloid beta peptides and improving memory performance both in *in vitro* and/or *in vivo* AD models [48, 49], pointing towards a crucial role for EGFR regulation in the disease. Whether there is a mechanistic link between epigenetic changes in *ZFYVE28* and altered EGFR signaling in AD remains to be investigated.

Next to the overlapping loci that were identified in previous reports, our brainstem EWAS also revealed novel genes displaying changes associated with AD pathology at the level of 5-uC and 5-mC. Interestingly, while representing new associations, three of these genes have been functionally linked to AD before, supporting the relevance of the current findings. These include *CAST* and *MALAT1*, for which we identified a DRN-specific DUR and DMR, respectively, as well as *LOC105377777;DLGAP2;DLGAP*, for which we identified a DMR within the LC. Calpastatin, the protein expressed by CAST, is known to protect against neuronal death induced by amyloid beta [50, 51]. Furthermore, CAST depletion has been shown to act upstream of calpains to activate a calpain-dependent cascade of protein kinase activation, cytoskeletal protein hyperphosphorylation, cytoskeletal proteolysis and neurodegeneration [52]. In another study using APP/PSEN1 mice, CAST downregulation was linked to hyperglycemia and the promotion AD pathological hallmarks, the former which is typically observed in diabetes, which is a known risk factor for developing AD [53]. The long noncoding RNA *MALAT1* on the other hand, has recently been shown to convey neuroprotective effects in AD by inhibiting apoptosis and inflammation while promoting neurite outgrowth [54, 55]. For *LOC105377777;DLGAP2;DLGAP2*, lower cortical *DLGAP2* expression has been observed in AD, associated with more plaques and tangles at autopsy and faster cognitive decline [56]. Furthermore, an association for this locus at the level of genetic variation, gene expression and protein expression, as well as altered methylation levels, have been associated with the disease [56, 57]. Further research is necessary to assess whether the identified epigenetic alterations in these loci also have functional implications and hence could mediate the effects observed in the aforementioned studies. For *GNAT1*, *RBMXL2* and *ANKRD2*, no direct or indirect association or functional connection has been identified with AD to date. Interestingly, however, is that other ankyrin repeat containing proteins, such as

ANK1, have previously been associated with AD neuropathology in other EWAS [11, 12, 16].

By conducting a GO term pathway analyses, we furthermore identified altered biological mechanisms in the DRN and LC, which were related to our top-ranked DUPs, DMPs and DHPs. Strikingly, genes related to 'homophilic cell adhesion via plasma membrane adhesion molecules' and 'calcium ion binding' were overrepresented in both brainstem regions. In the LC, we furthermore discovered two more pathways that were related to cell-cell interactions, *i.e.* 'cell-cell adhesion via plasma-membrane adhesion molecules' and 'cell-cell adhesion'. Interestingly, the cell adhesion molecules annotated to these aforementioned GO-terms have a major function in dendrite development, synaptic connectivity and neural circuit formation [58, 59], closely linked to key neuropathological features of AD. In the context of 'calcium ion binding', an increasing numbers of studies suggests that disruption of intracellular calcium ion homeostasis plays important roles in orchestrating the dynamic of the neuropathology of AD and associated memory loss, and cognitive dysfunction [60]. In fact, calcium dysregulation may even play an important role in the pathogenesis of AD, by inducing synaptic deficits and promoting the accumulation of amyloid beta plaques and neurofibrillary tangles [61]. All in all, these findings confirmed that the identified differentially modified genes are strongly linked to well-known AD-associated neuropathological processes, which further supports the relevance of epigenetic dysregulation in both brainstem regions in AD.

Finally, in the cell-type specific validation study using an independent patient cohort, where we targeted both laser-captured serotonergic and non-serotonergic cells derived from the DRN, we showed that epigenetic signatures in *TNXB* within this brainstem nucleus are strongly dependent upon the cell type analyzed. Whereas AD-associated hypermethylation was found for *TNXB* within serotonergic neurons, hypomethylation was identified in the patient-derived non-serotonergic cells. As such, these data suggest that epigenetic dysregulation in *TNXB* in AD is likely attributable to the non-serotonergic cells within the DRN, as the identified patterns in these cells overlap with the previously obtained bulk tissue EWAS data. This finding is further supported by the notion that the LC, which is free of serotonergic cells, also displayed *TNXB* hypomethylation with advancing Braak stage. These cell-specific findings also illustrate that a potential loss of serotonergic neurons, as commonly observed in AD [2, 62], possibly resulting in a different proportion of various cell types within the bulk tissues examined, is not able to explain our bulk tissue EWAS data, indicating a true epigenetic AD-specific signature. Thus, our results support the notion that epigenetic data derived from heterogeneous post-mortem bulk tissue should be

interpreted with caution, as changes in one cell type could negate or mask changes in another. Overall, this has crucial implications for future planned epigenetic studies in AD, as it warrants the need for single cell (-type) neuroepigenetic analyses, opposite to the more common bulk tissue analyses that have been performed to date. Evidently, interrogation of epigenetic marks is most informative when studied at a single-cell level, where intercellular differences can be dissected leading to a more refined understanding of their contribution to the disease. Taken together, the present study strongly implicates a role for *TNXB* dysregulation in the brainstem of AD in the development and course of the disease, and highlights potential cell-specific effects regarding this locus that emphasizes the need for future single cell(-type) neuroepigenetic studies in AD.

Acknowledgments

Funds have been provided by the Joint Program— Neurodegenerative Disease Research (JPND) for the EPI-AD consortium (http://www.epi-ad.eu/). The project is supported through the following funding organizations under the aegis of JPND; the Netherlands, The Netherlands Organization for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 643417.

References

1. Simic G, Stanic G, Mladinov M, Jovanov-Milosevic N, Kostovic I, Hof PR. Does Alzheimer's disease begin in the brainstem? Neuropathology and applied neurobiology. 2009;35(6):532-54.
2. Iatrou A, Kenis G, Rutten BP, Lunnon K, van den Hove DL. Epigenetic dysregulation of brainstem nuclei in the pathogenesis of Alzheimer's disease: looking in the correct place at the right time? Cell Mol Life Sci. 2017;74(3):509-23.
3. Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. Journal of Neuropathology & Experimental Neurology. 2011;70(11):960-9.
4. Grinberg L, Rüb U, Ferretti R, Nitrini R, Farfel J, Polichiso L, et al. The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset? Neuropathology and applied neurobiology. 2009;35(4):406-16.
5. Lee JH, Ryan J, Andreescu C, Aizenstein H, Lim HK. Brainstem morphological changes in Alzheimer's disease. Neuroreport. 2015;26(7):411.
6. Ji X, Wang H, Zhu M, He Y, Zhang H, Chen X, et al. Brainstem atrophy in the early stage of Alzheimer's disease: a voxel-based morphometry study. Brain Imaging and Behavior. 2021;15(1):49-59.
7. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131:21-64.

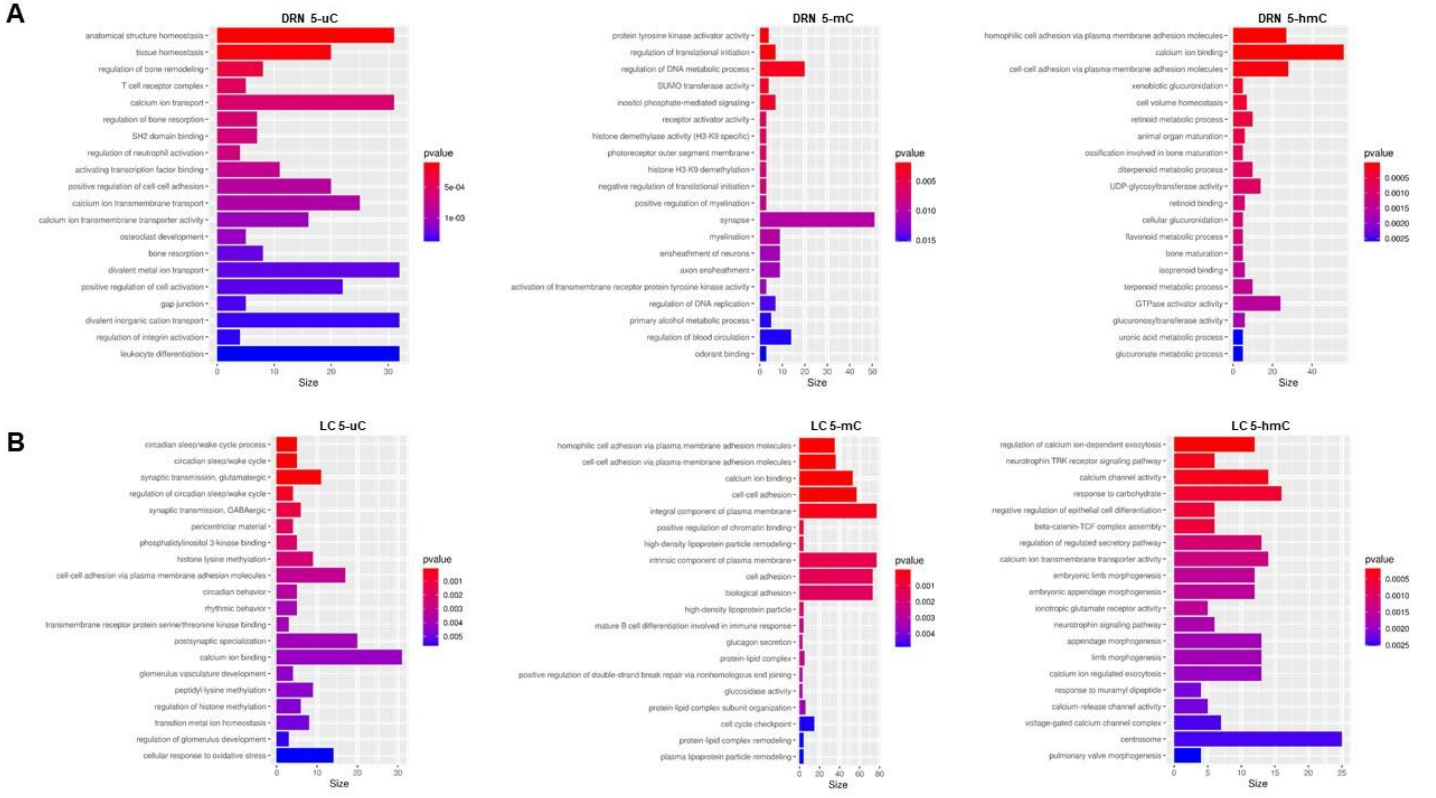
8. Lardenoije R, Pishva E, Lunnon K, van den Hove DL. Neuroepigenetics of Aging and Age-Related Neurodegenerative Disorders. Prog Mol Biol Transl Sci. 2018;158:49-82.
9. Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, et al. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. Journal of Alzheimer's Disease. 2012;29(3):571-88.
10. Sanchez-Mut JV, Aso E, Heyn H, Matsuda T, Bock C, Ferrer I, et al. Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. Hippocampus. 2014;24(4):363-8.
11. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.
12. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17(9):1156-63.
13. Smith AR, Smith RG, Condliffe D, Hannon E, Schalkwyk L, Mill J, et al. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. Neurobiology of aging. 2016;47:35-40.
14. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenetics. 2019;11(1):52.
15. Lardenoije R, Roubroeks JA, Pishva E, Leber M, Wagner H, Iatrou A, et al. Alzheimer's disease-associated (hydroxy) methylomic changes in the brain and blood. Clinical epigenetics. 2019;11(1):164.
16. Smith R, Pishva E, Shireby G, Smith AR, Roubroeks JA, Hannon E, et al. Meta-analysis of epigenome-wide association studies in Alzheime's disease highlights 220 differentially methylated loci across cortex. BioRxiv. 2020.
17. Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. Genome medicine. 2016;8(1):1-14.
18. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics & chromatin. 2018;11(1):41.
19. Smith RG, Hannon E, De Jager PL, Chibnik L, Lott SJ, Condliffe D, et al. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimer's & Dementia. 2018;14(12):1580-8.
20. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, et al. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. Acta neuropathologica. 2019;137(4):557-69.
21. Roubroeks JA, Smith AR, Smith RG, Pishva E, Ibrahim Z, Sattlecker M, et al. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. Neurobiology of Aging. 2020.
22. Beach TG, Sue LI, Walker DG, Roher AE, Lue L, Vedders L, et al. The Sun Health Research Institute Brain Donation Program: description and experience, 1987-2007. Cell Tissue Bank. 2008;9(3):229-45.
23. Charnay Y, Léger L. Brain serotonergic circuitries. Dialogues in clinical neuroscience. 2010;12(4):471.
24. Team RC. R: A language and environment for statistical computing. Vienna, Austria; 2013.
25. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80.

26. Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics. 2013;14(1):1-10.
27. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
28. Fortin J-P, Triche Jr TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. 2017;33(4):558-60.
29. Qu J, Zhou M, Song Q, Hong EE, Smith AD. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation. Bioinformatics. 2013;29(20):2645-6.
30. Kiihl SF, Martinez-Garrido MJ, Domingo-Relloso A, Bermudez J, Tellez-Plaza M. MLML2R: an R package for maximum likelihood estimation of DNA methylation and hydroxymethylation proportions. Statistical applications in genetics and molecular biology. 2019;18(1).
31. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics. 2012;28(22):2986-8.
32. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286-8.
33. Riemens RJM, Kenis G, Nolz J, Susano Chaves SC, Duroux D, Pishva E, et al. Targeted methylation profiling of single laser-capture microdissected post-mortem brain cells by adapted limiting dilution bisulfite pyrosequencing (LDBSP). 2021.
34. Hajj NE, Kuhtz J, Haaf T. Limiting Dilution Bisulfite Pyrosequencing(R): A Method for Methylation Analysis of Individual DNA Molecules in a Single or a Few Cells. Methods Mol Biol. 2015;1315:221-39.
35. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC genomics. 2019;20(1):1-15.
36. van den Hove DL, Riemens RJ, Koulousakis P, Pishva E. Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges. Brain Pathology. 2020.
37. McDevitt RA, Tiran-Cappello A, Shen H, Balderas I, Britt JP, Marino RA, et al. Serotonergic versus nonserotonergic dorsal raphe projection neurons: differential participation in reward circuitry. Cell reports. 2014;8(6):1857-69.
38. Chapuis J, Hot D, Hansmannel F, Kerdraon O, Ferreira S, Hubans C, et al. Transcriptomic and genetic studies identify IL-33 as a candidate gene for Alzheimer's disease. Molecular psychiatry. 2009;14(11):1004-16.
39. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. Journal of Alzheimer's Disease. 2011;23(2):349-59.
40. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. Alzheimer's research & therapy. 2019;11(1):5.
41. Chang T, editor Tau Network Genes in a Genome Wide Association Study of Progressive Supranuclear Palsy. MOVEMENT DISORDERS; 2017: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.
42. Valcourt U, Alcaraz LB, Exposito J-Y, Lethias C, Bartholin L. Tenascin-X: beyond the architectural function. Cell adhesion & migration. 2015;9(1-2):154-65.
43. Booij L, Casey KF, Antunes JM, Szyf M, Joober R, Israël M, et al. DNA methylation in individuals with anorexia nervosa and in matched normal-eater controls: A genome-wide study. International Journal of Eating Disorders. 2015;48(7):874-82.
44. Kesselmeier M, Pütter C, Volckmar A-L, Baurecht H, Grallert H, Illig T, et al. High-throughput DNA methylation analysis in anorexia nervosa confirms TNXB hypermethylation. The World Journal of Biological Psychiatry. 2018;19(3):187-99.

45. McCartney DL, Walker RM, Morris SW, Anderson SM, Duff BJ, Marioni RE, et al. Altered DNA methylation associated with a translocation linked to major mental illness. npj Schizophrenia. 2018;4(1):1-7.
46. Haertle L, Müller T, Lardenoije R, Maierhofer A, Dittrich M, Riemens RJ, et al. Methylomic profiling in trisomy 21 identifies cognition-and Alzheimer's disease-related dysregulation. Clinical epigenetics. 2019;11(1):1-11.
47. Mosesson Y, Chetrit D, Schley L, Berghoff J, Ziv T, Carvalho S, et al. Monoubiquitinylation regulates endosomal localization of Lst2, a negative regulator of EGF receptor signaling. Developmental cell. 2009;16(5):687-98.
48. Tavassoly O, Sato T, Tavassoly I. Inhibition of Brain Epidermal Growth Factor Receptor Activation: A Novel Target in Neurodegenerative Diseases and Brain Injuries. Molecular Pharmacology. 2020;98(1):13-22.
49. Wang L, Chiang H-C, Wu W, Liang B, Xie Z, Yao X, et al. Epidermal growth factor receptor is a preferred target for treating Amyloid-β–induced memory loss. Proceedings of the National Academy of Sciences. 2012;109(41):16743-8.
50. Nixon RA, Saito K-I, Grynspan F, Griffin WR, Katayama S, Honda T, et al. Calcium-activated neutral proteinase (calpain) system in aging and Alzheimer's disease. Annals of the New York Academy of Sciences. 1994;747:77.
51. Nakayama J, Yoshizawa T, Yamamoto N, Arinami T. Mutation analysis of the calpastatin gene (CAST) in patients with Alzheimer's disease. Neuroscience letters. 2002;320(1-2):77-80.
52. Rao MV, Mohan PS, Peterhoff CM, Yang D-S, Schmidt SD, Stavrides PH, et al. Marked calpastatin (CAST) depletion in Alzheimer's disease accelerates cytoskeleton disruption and neurodegeneration: neuroprotection by CAST overexpression. Journal of Neuroscience. 2008;28(47):12241-54.
53. Zhu L, Gong L, Yang T, Xiao X. Calpastatin Mediates Development of Alzheimer's Disease in Diabetes. Journal of Alzheimer's Disease. 2019;68(3):1051-9.
54. Ma P, Li Y, Zhang W, Fang F, Sun J, Liu M, et al. Long Non-coding RNA MALAT1 Inhibits Neuron Apoptosis and Neuroinflammation While Stimulates Neurite Outgrowth and Its Correlation With MiR-125b Mediates PTGS2, CDK5 and FOXQ1 in Alzheimer's Disease. Current Alzheimer Research. 2019;16(7):596-612.
55. Li L, Xu Y, Zhao M, Gao Z. Neuro-protective roles of long non-coding RNA MALAT1 in Alzheimer's disease with the involvement of the microRNA-30b/CNR1 network and the following PI3K/AKT activation. Experimental and Molecular Pathology. 2020;117:104545.
56. Ouellette AR, Neuner SM, Dumitrescu L, Anderson LC, Gatti DM, Mahoney ER, et al. Cross-Species Analyses Identify Dlgap2 as a Regulator of Age-Related Cognitive Decline and Alzheimer's Dementia. Cell reports. 2020;32(9):108091.
57. Chaudhry M, Wang X, Bamne MN, Hasnain S, Demirci FY, Lopez OL, et al. Genetic variation in imprinted genes is associated with risk of late-onset Alzheimer's disease. Journal of Alzheimer's Disease. 2015;44(3):989-94.
58. Friedman LG, Benson DL, Huntley GW. Cadherin-based transsynaptic networks in establishing and modifying neural connectivity. Current topics in developmental biology. 112: Elsevier; 2015. p. 415-65.
59. Peek SL, Mah KM, Weiner JA. Regulation of neural circuit formation by protocadherins. Cellular and molecular life sciences. 2017;74(22):4133-57.
60. Wang Y, Shi Y, Wei H. Calcium dysregulation in Alzheimer's disease: a target for new drug development. Journal of Alzheimer's disease & Parkinsonism. 2017;7(5).
61. Tong BC-K, Wu AJ, Li M, Cheung K-H. Calcium signaling in Alzheimer's disease & therapies. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2018;1865(11):1745-60.

62. Theofilas P, Dunlop S, Heinsen H, Grinberg LT. Turning on the light within: subcortical nuclei of the isodentritic core and their role in Alzheimer's disease pathogenesis. Journal of Alzheimer's Disease. 2015;46(1):17-34.

Supplementary material



Supplementary Figure 1. Gene Ontology (GO) terms enriched in the dorsal raphe nucleus (DRN) and locus coeruleus LC. Displayed are the top 20 enriched GO terms in the DRN (**A**) and LC (**B**) for each cytosine state: unmodified cytosine (5-uC), 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC). The x-axis displays the number of differentially modified genes in the pathway.

Supplementary Table 1. Top 100 ranked differentially unmodified positions (DUPs) in the dorsal raphe nuclei (DRN)

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg24088087	10	50747319	275.5560163	58.56392056	4.705218053	1.08E-05	ERCC6; ERCC6-PGBD3; ERCC6-PGBD3	TSS200; TSS200; 5'UTR	Island
2	cg21146221	22	30279857	-127.95102	26.66437096	-4.798576353	7.54E-06	MTMR3; MTMR3; MTMR3	5'UTR; 5'UTR; 5'UTR	Island
3	cg23248871	19	45004792	-160.4101653	36.77468014	-4.361973093	3.90E-05	ZNF180	TSS1500	Island
4	cg24066712	1	20982867	99.8026816	19.69174422	5.068249946	2.63E-06	DDOST	Body	
5	cg07379335	5	127211437	127.8666408	30.43421173	4.201411291	6.99E-05			
6	cg16256719	1	228613082	-200.9376863	48.62737944	-4.132192369	8.96E-05	HIST3H3	TSS200	Island
7	cg14225485	2	14772568	90.32632367	19.66975559	4.592142655	1.66E-05	FAM84A	TSS1500	Island
8	cg06714901	12	111283526	88.16537314	18.74654862	4.703018937	1.09E-05	CCDC63; CCDC63; CCDC63	TSS1500; TSS1500; TSS1500	
9	cg00109772	19	14638399	90.76773256	20.66696314	4.391924054	3.49E-05			Northern shore
10	cg25456440	5	132202607	-675.8455142	167.7148171	-4.029730502	1.29E-04	UQCRQ; GDF9; GDF9; GDF9; GDF9; GDF9	Body; TSS200; TSS200; TSS1500; TSS1500; TSS1500	Island
11	cg25459301	8	10941183	84.92135939	18.87164507	4.499944709	2.34E-05	XKR6	Body	
12	cg17867860	2	49054009	100.8966953	24.09625278	4.187235924	7.36E-05			
13	cg19037107	11	134126323	96.55238634	23.09082693	4.181417436	7.51E-05	ACAD8	Body	Southern shelf
14	cg21968868	1	203617631	108.1014362	26.42819217	4.09038331	1.04E-04	ATP2B4; ATP2B4	5'UTR; 5'UTR	
15	cg00351957	8	103571768	93.8311369	22.48197172	4.17361689	7.73E-05	ODF1	Body	
16	cg03948693	16	69125843	92.46522547	22.13797852	4.176769138	7.64E-05			
17	cg02938045	16	48222834	104.4106415	25.65293161	4.070125125	1.12E-04	ABCC11; ABCC11; ABCC11	Body; Body; Body	
18	cg08811130	1	110285277	78.65990855	18.15208081	4.333382457	4.33E-05			Southern shore
19	cg16008609	1	213123675	-210.7933164	53.15253954	-3.965818344	1.61E-04	VASH2; VASH2; VASH2	TSS1500; TSS1500; TSS1500	Island

Supplementary Table 1. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
20	cg10162019	5	146889701	-154.992089	39.2939016	-3.944431138	1.74E-04			Island
21	cg27196940	4	110736273	74.85598275	17.70653429	4.227590872	6.36E-05	GAR1; GAR1	TSS1500; TSS1500	Northern shore
22	cg22765964	11	75110494	118.4586417	30.02188075	3.945743528	1.73E-04	SNORD15A; RPS3	TSS1500; TSS200	Island
23	cg00258805	17	29876732	-205.9814617	52.72114233	-3.906999215	1.98E-04			Island
24	cg19455923	6	32016113	-110.1127	27.91765884	-3.944195343	1.74E-04	TNXB	Body	
25	cg11688683	11	120465256	116.2912615	29.57456197	3.932138085	1.81E-04	GRIK4; GRIK4; GRIK4	5'UTR; 5'UTR; 5'UTR	
26	cg06967232	1	247148831	-80.12879142	19.6627735	-4.075152034	1.10E-04	ZNF695; ZNF695; ZNF695; ZNF670-ZNF695	3'UTR; Body; Body; Body	
27	cg00836101	15	39876248	62.01610249	14.21858385	4.36162301	3.91E-05	THBS1	Body	Southern shelf
28	cg27243121	11	70281584	78.39347014	19.2108733	4.080682275	1.08E-04	CTTN; CTTN	3'UTR; 3'UTR	
29	cg13121728	16	10765186	87.86434487	21.81570605	4.027572826	1.30E-04	TEKT5	Body	
30	cg14811321	6	168084536	84.10081364	20.86573456	4.030570474	1.28E-04	LOC441178	TSS200	
31	cg00452094	21	45888218	75.82746543	18.49115926	4.100741568	1.00E-04			
32	cg06273744	22	40754965	89.09795242	22.3649374	3.983823019	1.51E-04	ADSL; ADSL	Body; Body	
33	cg00446758	7	132764863	106.5970183	27.36014775	3.896068811	2.05E-04	CHCHD3	Body	Northern shore
34	cg15670836	12	58936880	70.86932485	17.4295005	4.066055987	1.13E-04			
35	cg00527316	21	41356964	85.02610483	21.48417282	3.957615941	1.66E-04			
36	cg18215878	11	47152876	111.0448146	28.64147141	3.877063891	2.19E-04	C11orf49; C11orf49; C11orf49; C11orf49; C11orf49; C11orf49; C11orf49	Body; Body; Body; Body; Body; Body; Body	
37	cg13496359	9	109686439	74.20294736	18.40917329	4.030759349	1.28E-04	ZNF462; MIR548Q	Body; Body	
38	cg11235259	19	47526657	54.18839896	12.43817547	4.356619594	3.98E-05	NPAS1	Body	Southern shore

Supplementary Table 1. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
39	cg00411741	19	18208476	-273.1099286	71.23359628	-3.834004499	2.54E-04	MAST3	TSS200	Northern shore
40	cg15028904	5	140700449	-480.8187681	125.7546222	-3.823467954	2.63E-04	TAF7	TSS200	Island
41	cg22391883	1	16563727	-724.663516	189.7850801	-3.818337645	2.68E-04	C1orf89	TSS200	Island
42	cg23884626	6	31126149	-485.6867805	127.1112569	-3.820958051	2.66E-04	CCHCR1; CCHCR1; TCF19; TCF19; CCHCR1	TSS1500; TSS1500; TSS200; TSS200; TSS200	Island
43	cg26199493	1	226497589	87.19071962	22.16687096	3.933379672	1.80E-04	LIN9	TSS1500	Southern shore
44	cg16381797	5	95997312	55.51711834	13.1233352	4.230412277	6.30E-05	CAST; CAST	TSS1500; TSS1500	Northern shore
45	cg09316855	6	11806488	115.6832027	30.08550746	3.845147132	2.45E-04			
46	cg20218040	6	14369697	88.0586099	22.46007184	3.920673563	1.89E-04			
47	cg07505669	22	40754947	59.90658516	14.44264173	4.147896643	8.47E-05	ADSL; ADSL	Body; Body	
48	cg07481380	19	3834519	60.18075473	14.54253357	4.13825792	8.77E-05	ZFR2	Body	Island
49	cg26181763	17	9728527	68.52355579	16.99527538	4.031917945	1.28E-04	GLP2R	TSS1500	
50	cg11535971	10	33848245	58.40561915	14.07289439	4.150220809	8.40E-05			
51	cg01348938	19	2806850	-76.68129399	19.37073092	-3.958616446	1.65E-04	THOP1	Body	Northern shore
52	cg13713927	3	184534541	84.92912384	21.71081164	3.911835506	1.94E-04	VPS8; VPS8	5'UTR; 5'UTR	
53	cg11921048	3	156347867	94.5600643	24.43089211	3.870512131	2.24E-04			
54	cg20933293	17	56429700	-217.3370164	57.32660668	-3.791206718	2.94E-04	SUPT4H1	TSS200	Island
55	cg09806242	6	10303745	51.72917379	12.30798687	4.202894784	6.96E-05			
56	cg09856611	3	73403300	82.40861533	21.09117666	3.907255469	1.97E-04			
57	cg23404137	18	36370257	-46.54148333	10.18884378	-4.567886635	1.81E-05			
58	cg22666350	17	4052392	-56.21297095	13.61494022	-4.128771045	9.07E-05	CYB5D2; CYB5D2; CYB5D2	5'UTR; Body; 5'UTR	

Supplementary Table 1. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
59	cg08951005	10	79803774	66.67400258	16.71811771	3.988128553	1.49E-04	RPS24	Body	
60	cg16930098	20	30489015	96.53368525	25.18198302	3.833442552	2.54E-04	TTLL9	Body	
61	cg21339445	21	45528892	-83.36058348	21.49482222	-3.87817041	2.18E-04	PWP2	Body	Southern shore
62	cg15067802	9	114860737	62.56184631	15.67283855	3.991736793	1.47E-04	SUSD1; SUSD1; SUSD1; MIR3134; SUSD1; SUSD1; SUSD1	Exon boundaries; Exon boundaries; Exon boundaries; Body; Body; Body; Body	
63	cg07139190	6	7162285	-44.87612134	10.18445594	-4.406334672	3.31E-05	RREB1; RREB1; RREB1; RREB1	5'UTR; 5'UTR; 5'UTR; 5'UTR	
64	cg23326220	19	51563756	42.73344241	8.450226732	5.057076427	2.75E-06	KLK13	Body	
65	cg16678626	6	160513496	-51.98256763	12.71608152	-4.087939163	1.05E-04	LOC729603; IGF2R	TSS1500; Body	Southern shore
66	cg18382831	3	138191900	-93.39579679	24.44405832	-3.820797495	2.66E-04	ESYT3	Body	
67	cg24985835	17	79520203	108.4714223	28.74296816	3.773842065	3.12E-04	C17orf70; C17orf70	TSS1500; TSS1500	Southern shore
68	cg02817292	14	23532175	89.3526646	23.35407459	3.825998938	2.61E-04	ACIN1; ACIN1; ACIN1; ACIN1; ACIN1	Body; Body; Body; Body; Body	Southern shelf
69	cg21157465	1	11020631	-63.12368098	16.02234775	-3.93972731	1.76E-04	C1orf127; C1orf127	Body; Body	
70	cg00312919	7	38217710	124.5972264	33.13914857	3.759819784	3.27E-04	STARD3NL	TSS1500	Island
71	cg15848350	3	128326938	43.76583376	10.19791614	4.291644796	5.04E-05			Island
72	cg03945836	16	19138344	68.4001632	17.53029129	3.90182696	2.01E-04			
73	cg03381007	16	88893573	-48.06168677	11.59680307	-4.144391043	8.58E-05	GALNS	Body	Northern shore
74	cg20766090	11	2159131	170.0046649	45.50090817	3.736291687	3.54E-04	INS-IGF2; IGF2; IGF2; IGF2	Body; Body; 5'UTR; 5'UTR	Island
75	cg22038207	4	152808142	44.81889916	10.78927735	4.154022341	8.29E-05			
76	cg10758875	15	56725876	-74.49128613	19.37770894	-3.844174065	2.45E-04	MNS1; TEX9	Body; 3'UTR	
77	cg10467427	6	1633242	83.3263001	21.82316227	3.818250492	2.68E-04	GMDS; GMDS	Body; Body	

Supplementary Table 1. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
78	cg23293975	20	45625848	-53.95847685	13.60248246	-3.966810984	1.61E-04	EYA2; EYA2	Body; Body	
79	cg18165313	10	109872152	-41.46865328	9.579098863	-4.329076656	4.40E-05			
80	cg25410105	6	150945640	51.63799732	12.95783295	3.98507972	1.51E-04	PLEKHG1	5'UTR	
81	cg07447350	4	16901824	77.08908552	20.15383794	3.825032521	2.62E-04	LDB2; LDB2; LDB2; LDB2; LDB2	TSS1500; TSS1500; TSS1500; TSS1500; TSS1500	
82	cg05935360	19	11639865	-951.585562	256.5448461	-3.709236714	3.88E-04	ECSIT; ECSIT; ECSIT; ECSIT	5'UTR; TSS200; 5'UTR; 5'UTR	Island
83	cg03554286	3	185497487	57.90586759	14.76696866	3.921310387	1.88E-04	IGF2BP2; IGF2BP2	Body; Body	
84	cg08947125	3	12802515	-89.44829738	23.67033157	-3.778920339	3.06E-04	TMEM40	Body	
85	cg01987776	7	138516319	47.15762815	11.62484985	4.056622559	1.17E-04	KIAA1549; KIAA1549	3'UTR; 3'UTR	
86	cg05075268	22	19718750	43.3302499	10.43278168	4.153278695	8.31E-05			Island
87	cg05533872	2	112908757	99.3940528	26.4913594	3.751942333	3.36E-04	FBLN7; FBLN7	Body; Body	
88	cg12807924	20	36013546	47.13650669	11.6638908	4.041233539	1.24E-04	SRC; SRC	Body; Body	Southern shore
89	cg09959585	14	103570855	49.33905839	12.35164686	3.994532789	1.46E-04	C14orf73	Body	Island
90	cg03346733	19	13987364	37.72311923	8.508473563	4.433594222	3.00E-05	NANOS3	TSS1500	Northern shore
91	cg03265360	6	16472398	69.37749066	18.0806776	3.837106783	2.51E-04	ATXN1; ATXN1	5'UTR; 5'UTR	
92	cg08968069	16	72903003	72.60675705	19.01125092	3.819146743	2.67E-04	ZFHX3; ZFHX3	Body; Body	
93	cg27289832	3	71592784	39.08494946	9.038586561	4.324232467	4.48E-05	FOXP1; FOXP1; FOXP1; FOXP1; FOXP1	TSS200; 5'UTR; 5'UTR; 5'UTR; 5'UTR	
94	cg08931356	1	237057546	60.81153062	15.69522343	3.874524685	2.21E-04	MTR; MTR; MTR	Body; Body; Body	
95	cg00002531	7	127721794	-57.32480574	14.71659136	-3.895250221	2.06E-04	SND1; MIR593	Body; TSS200	
96	cg24022821	6	28831886	-318.4817417	86.07149278	-3.700200048	3.99E-04			Island

Supplementary Table 1. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
97	cg23439947	22	28254426	103.4331577	27.74028526	3.728626319	3.63E-04	PITPNB; PITPNB; PITPNB	Body; Body; Body	
98	cg26770421	6	10735168	56.71910629	14.56518288	3.894156823	2.07E-04			
99	cg07151462	2	173179124	49.39524301	12.49516169	3.953149567	1.68E-04			
100	cg06562246	17	34309577	49.66977334	12.57309363	3.950481466	1.70E-04	CCL16	TSS1500	

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the *t*-statistics, the accompanying *p* values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined *p* value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 2. Top 100 ranked differentially methylated positions (DMPs) in the dorsal raphe nuclei (DRN)

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg03442014	16	23607821	152.1853638	34.86627573	4.364829928	3.72E-05	NDUFAB1	TSS200	Island
2	cg00339300	1	61508924	463.6149796	111.3774606	4.162556565	7.80E-05			Island
3	cg21540765	1	200008933	203.0088517	49.27401443	4.119998218	9.09E-05	NR5A2; NR5A2	Body; Body	Island
4	cg21336876	12	100536815	350.8025324	86.38446546	4.060944645	1.12E-04	UHRF1BP1L; UHRF1BP1L	TSS200; TSS200	Southern shore
5	cg24671939	18	3593798	81.05029979	19.03812481	4.257262761	5.53E-05	FLJ35776; DLGAP1; DLGAP1	TSS1500; Body; Body	
6	cg02827328	9	100881403	640.1800163	163.3235977	3.919703124	1.84E-04	TRIM14; TRIM14; TRIM14; TRIM14	1st exon; 1st exon; 1st exon; 1st exon	Island
7	cg14182420	7	135242446	69.89612255	16.24393268	4.302906441	4.68E-05	NUP205	TSS1500	
8	cg16636721	21	47920571	67.46546117	15.77041216	4.277977043	5.13E-05	DIP2A; DIP2A; DIP2A; DIP2A; DIP2A; DIP2A; DIP2A	Body; Body; Body; Body; Body; Body; Body	Southern shore
9	cg11013544	22	42062791	182.3850332	45.91988316	3.971809609	1.54E-04			Island

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
10	cg15647861	19	50320776	245.3163197	62.16994693	3.945898812	1.68E-04	MED25	TSS1500	Northern shore
11	cg17255342	2	29181736	79.43439409	19.05463694	4.16876975	7.63E-05			
12	cg23493016	10	106440778	43.86415027	9.864989461	4.446446745	2.75E-05	SORCS3	Body	
13	cg14702927	19	46366445	112.1124942	28.63488895	3.915241101	1.87E-04	FOXA3; SYMPK; SYMPK	TSS1500; 5'UTR; 1st exon	Island
14	cg18957501	22	24246964	36.57113877	7.689921251	4.755723443	8.48E-06			
15	cg10873891	20	35806913	79.68493421	19.98882318	3.986474516	1.46E-04	RPN2; RPN2; MROH8; MROH8; MROH8	TSS1500; TSS1500; Body; Body; Body	Northern shore
16	ch.6.93211F	6	3310225	36.95110859	8.25544884	4.475966032	2.46E-05	SLC22A23; SLC22A23	Body; Body	
17	cg13966567	17	882342	-79.69936368	20.0321764	-3.978567385	1.50E-04	NXN	Body	Island
18	cg25298664	15	74753253	187.2760544	48.45659852	3.864820482	2.23E-04	UBL7; UBL7; UBL7; UBL7; UBL7-AS1; UBL7-AS1; UBL7-AS1; UBL7; UBL7; UBL7	5'UTR; 5'UTR; 5'UTR; 5'UTR; TSS1500; TSS1500; 1st exon; 5'UTR; 5'UTR	Northern shore
19	cg18165313	10	109872152	39.57038389	9.375501545	4.220615153	6.32E-05			
20	cg18547866	7	100464803	372.2324804	97.11222044	3.833013792	2.49E-04	MIR6875; TRIP6	TSS1500; TSS200	Southern shore
21	cg06768385	1	214557399	39.01433032	9.1791741	4.250309439	5.67E-05	PTPN14	Body	Island
22	cg21146221	22	30279857	99.72576701	25.61191578	3.8937254	2.02E-04	MTMR3; MTMR3; MTMR3	5'UTR; 5'UTR; 5'UTR	Island
23	cg00723431	1	234743527	43.06059566	10.33512297	4.166432832	7.69E-05	IRF2BP2; IRF2BP2	Body; Body	Northern shore
24	cg10348596	3	128045033	34.62400727	7.676044373	4.510657519	2.16E-05	EEFSEC	Body	
25	cg27098574	16	75285489	36.60567415	8.476274894	4.31860394	4.42E-05	BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1	Body; 5'UTR; 1st exon; Body; Body; 5'UTR; Body; Body; 1st exon	Island

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
26	cg00683895	19	58739944	137.7404989	35.8319653	3.844067657	2.40E-04	ZNF544	TSS200	Island
27	cg24654765	5	1887193	373.581911	97.88310553	3.816612775	2.63E-04	CTD-2194D22.4; IRX4; IRX4; IRX4; IRX4; IRX4; IRX4	TSS1500; TSS200; TSS200; 1st exon; 1st exon; 5'UTR; 5'UTR	Island
28	ch.11.991049F	11	47494477	67.15339201	17.03399975	3.942314958	1.70E-04	CUGBP1; CUGBP1; CUGBP1	Body; Body; Body	
29	cg25017194	12	6566966	36.8227229	8.776647222	4.195534122	6.92E-05	TAPBPL	Body	
30	cg21655969	15	90792609	33.71533815	7.65162803	4.406296021	3.19E-05	TTLL13	TSS200	Island
31	cg09618400	12	6723242	74.62394869	19.05984643	3.915243965	1.87E-04			
32	cg21943376	11	118800953	-267.0819365	70.01489024	-3.814644793	2.65E-04			
33	cg18397528	17	7816510	37.67267269	9.052912955	4.161386824	7.83E-05			Northern shelf
34	cg22985466	12	57992676	-31.2626363	6.593505247	-4.741428895	8.96E-06	PIP4K2C; PIP4K2C; PIP4K2C; PIP4K2C	Body; Body; Body; Body	
35	cg20420007	1	173399241	55.01190558	13.88665799	3.961493515	1.59E-04			
36	cg01021976	12	58424937	-31.3314203	6.775418073	-4.624278526	1.41E-05			
37	cg03498048	14	99733089	131.8760599	34.54060866	3.818000463	2.62E-04	BCL11B; BCL11B	Body; Body	Northern shelf
38	ch.9.2787281F	9	2797281	36.63973261	8.825941689	4.151368081	8.12E-05			
39	cg26856965	10	102802736	54.35432138	13.87078317	3.918619497	1.85E-04			Island
40	cg04378603	11	65266494	47.83080887	12.10651624	3.950831759	1.65E-04	MALAT1	Body	Southern shore
41	cg18174834	1	212740460	64.13793797	16.53423815	3.879098475	2.12E-04	ATF3	5'UTR	
42	cg05423144	4	139792229	38.55556721	9.677648381	3.983980993	1.47E-04			
43	ch.12.1175666F	12	56599430	34.99294825	8.579612914	4.078616204	1.05E-04	RNF41; RNF41; RNF41	3'UTR; 3'UTR; 3'UTR	
44	cg04563422	14	101532902	-28.38472525	6.425042179	-4.417827068	3.06E-05	MIR656	TSS200	Southern shore
45	cg05393131	5	156755714	35.44503864	8.889427546	3.987325219	1.46E-04	CYFIP2; CYFIP2; CYFIP2; CYFIP2	Body; Body; Body; Body	

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
46	cg03286076	16	1264599	-28.12710604	6.48051822	-4.340255684	4.08E-05	CACNA1H; CACNA1H	Body; Body	Northern shore
47	cg27000511	9	96216109	42.94419276	11.05766048	3.883659914	2.09E-04	FAM120AOS; FAM120A; FAM120A; FAM120A; FAM120A	TSS1500; Body; Body; Body; Body	Southern shore
48	cg23602181	18	9913801	509.6958059	136.8458696	3.724597662	3.60E-04	VAPA; VAPA	TSS200; TSS200	Island
49	cg09635586	16	11836025	117.1990439	31.21537548	3.754529368	3.26E-04	TXNDC11	Body	Northern shore
50	cg04141970	18	3594060	85.9012493	22.77358175	3.771969217	3.07E-04	DLGAP1-AS1; DLGAP1; DLGAP1; DLGAP1; DLGAP1; DLGAP1; DLGAP1; DLGAP1; DLGAP1	TSS200; Body; Body; Body; Body; Body; Body; Body; Body	
51	cg00782260	1	16142290	48.09635946	12.53029112	3.838407184	2.44E-04			
52	cg25256175	22	30828305	-28.05954413	6.710663983	-4.181336481	7.29E-05			
53	cg21638110	10	126430141	84.15061473	22.34248259	3.766395	3.13E-04	FAM53B	5'UTR	Northern shore
54	cg27180153	13	25320559	28.45169113	6.900294915	4.123257264	8.98E-05			Island
55	cg17780956	4	156297616	446.4218229	120.3289059	3.710013146	3.79E-04	MAP9	5'UTR	Island
56	cg02796548	1	92372096	36.02741566	9.224692226	3.905541212	1.94E-04	TGFBR3	TSS1500	
57	cg24506767	6	52264762	49.40710329	12.94496458	3.816704401	2.63E-04	PAQR8	5'UTR	Northern shelf
58	cg03500459	6	2876783	117.0590911	31.34792252	3.734189755	3.49E-04	SERPINB9P1	TSS200	Southern shore
59	cg22268510	6	32118420	-22.67295459	4.82899409	-4.695171328	1.07E-05	PRRT1	Body	Island
60	cg10487619	16	81238825	55.08978757	14.57953091	3.778570648	3.00E-04	PKD1L2; PKD1L2; PKD1L2	Body; Body; Body	
61	cg24754949	9	36555992	91.36845698	24.48622543	3.731422683	3.52E-04			
62	cg25531478	6	111279493	48.05566253	12.66050881	3.795713366	2.83E-04	GTF3C6	TSS1500	Northern shore

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
63	cg04031129	1	43638029	788.5918523	215.0072816	3.667744861	4.36E-04	WDR65; WDR65; EBNA1BP2; EBNA1BP2; WDR65; WDR65; WDR65; WDR65	5'UTR; 1st exon; TSS200; 1st exon; 1st exon; 5'UTR; Body; 1st exon; 5'UTR	Island
64	cg07362810	10	131525915	36.21010635	9.411936759	3.84725347	2.37E-04	MGMT	Body	
65	cg25809301	3	71179811	98.34103026	26.4392203	3.719513252	3.67E-04	FOXP1	Body	
66	cg23420697	3	58038212	39.36234094	10.33816143	3.807479814	2.72E-04	FLNB; FLNB; FLNB; FLNB	Body; Body; Body; Body	
67	cg20080247	8	133887910	48.05603662	12.73050224	3.774873584	3.04E-04	TG	Body	
68	cg10851774	8	20113188	-22.16864687	5.047526899	-4.391981918	3.37E-05	LZTS1	TSS1500	
69	cg09861346	5	176037510	-61.01543644	16.29153172	-3.745224051	3.36E-04	GPRIN1	TSS1500	Island
70	cg12565250	10	72165370	30.49527024	7.842399139	3.888512902	2.06E-04	EIF4EBP2	Body	Southern shore
71	cg07141231	21	26980614	85.20797302	22.93540065	3.715129041	3.72E-04	MRPL39; MRPL39	TSS1500; TSS1500	Southern shore
72	cg12210890	8	110988071	-79.63693624	21.3853459	-3.723902181	3.61E-04	KCNV1	TSS1500	Island
73	cg07214473	17	46048181	150.1196675	40.73978692	3.684841744	4.12E-04	CDK5RAP3; CDK5RAP3; CDK5RAP3; CDK5RAP3; CDK5RAP3	TSS200; TSS200; TSS200; TSS200; Body	Island
74	cg24012595	7	99150098	80.11085462	21.54370486	3.718527297	3.68E-04	C7orf38	TSS1500	Southern shore
75	cg14336803	9	19046431	28.21685175	7.196913137	3.920688109	1.84E-04			Northern shelf
76	cg10310275	19	16296904	94.97975814	25.66221484	3.701152014	3.90E-04	FAM32A	Body	Southern shore
77	cg05229927	7	3278961	-22.02421653	5.23065052	-4.210607542	6.55E-05			
78	cg09989688	11	45929369	21.04627539	4.771390095	4.41093161	3.14E-05	C11orf94	TSS1500	
79	cg05491695	11	65266512	36.90968286	9.749004103	3.785995212	2.93E-04	MALAT1	Body	Southern shore

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
80	cg00340102	20	34129456	-80.45830404	21.71169071	-3.705759496	3.84E-04	ERGIC3; ERGIC3	TSS1500; TSS1500	Northern shore
81	cg11267619	7	29604021	425.1724341	116.680822	3.64389303	4.73E-04	PRR15	5'UTR	Island
82	cg22387174	7	155302770	80.39647644	21.69975937	3.704947833	3.85E-04	CNPY1	5'UTR	Island
83	cg10840389	10	72141924	214.8809623	58.86160571	3.650613328	4.62E-04	LRRC20; LRRC20; LRRC20	5'UTR; TSS1500; 5'UTR	Island
84	cg00924357	7	131012552	629.4688207	173.4674714	3.6287427	4.97E-04	MKLN1; MKLN1	Body; TSS200	Island
85	cg25280938	7	6543803	31.10601753	8.121998129	3.829847907	2.52E-04	GRID2IP	Body	Island
86	cg03260785	9	100395621	517.7361213	142.6715692	3.628866805	4.97E-04	TSTD2; NCBP1; TSTD2	5'UTR; TSS200; 1st exon	Island
87	cg10488854	1	37078857	-23.03226696	5.686554111	-4.050302963	1.17E-04			
88	cg17581200	17	80040464	-19.59343406	4.246315422	-4.614220119	1.46E-05	FASN	Body	Northern shore
89	cg19632574	1	26324332	90.98432801	24.7440935	3.67701197	4.23E-04	PAFAH2	5'UTR	Northern shore
90	cg07822469	1	182585393	49.39184414	13.25277755	3.726905092	3.58E-04			Southern shore
91	cg27532721	8	17534102	26.72290135	6.8684157	3.890693651	2.04E-04	MTUS1; MTUS1; MTUS1; MTUS1; MTUS1	Body; Body; Body; Body; Body	
92	cg03062389	22	42606263	19.51902066	4.246595496	4.596392729	1.56E-05	TCF20; TCF20	1st exon; 1st exon	
93	cg06245023	9	130635140	21.19490644	5.06361794	4.185723861	7.17E-05	AK1; AK1	Body; Exon boundaries	
94	cg08047546	7	151170250	-24.81417581	6.305280809	-3.93545927	1.75E-04	RHEB	Body	
95	cg14080129	9	90439478	-37.13297659	9.90828275	-3.74767026	3.33E-04			Northern shore
96	cg04972065	12	53591766	32.06649437	8.476723514	3.782887848	2.96E-04	ITGB7	Body	Island
97	cg23571456	6	26383374	37.39747851	9.990448631	3.743323237	3.38E-04	BTN2A2; BTN2A2; BTN2A2; BTN2A2	5'UTR; 1st exon; 1st exon; 5'UTR	
98	cg08891047	11	56309888	-41.94519342	11.26349893	-3.723993198	3.61E-04	OR8U8; OR5M11	Body; 1st exon	
99	cg16307912	17	32078405	30.18433579	7.97194747	3.786318951	2.92E-04	ACCN1	Body	

Supplementary Table 2. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
100	cg08960352	12	68043561	190.7151927	52.78633164	3.612965456	5.24E-04	DYRK2; DYRK2	Body; 5'UTR	Island

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the t-statistics, the accompanying p values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined p value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 3. Top 100 ranked differentially hydroxymethylated positions (DHPs) in the dorsal raphe nuclei (DRN)

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg05707690	7	157541334	-36.66215402	8.105786538	-4.522960708	2.06E-05	PTPRN2; PTPRN2; PTPRN2; PTPRN2; PTPRN2	Body; Body; Body; Body; Body	
2	cg10348596	3	128045033	-34.3697842	7.816315427	-4.397184904	3.30E-05	EEFSEC	Body	
3	cg04884072	3	123459525	30.01401977	5.910145844	5.078389022	2.39E-06	MYLK; MYLK; MYLK; MYLK	Body; Body; Body; Body	
4	cg00501931	9	110193168	-37.15526598	8.739751862	-4.251295297	5.65E-05	LINC01509; LINC01509	Exon boundaries; Body	
5	cg12081455	16	67465561	28.62176198	5.900182669	4.850995908	5.86E-06	HSD11B2	Body	Southern shore
6	cg15782451	13	34820887	29.58266848	6.814181954	4.341338209	4.06E-05			
7	cg02167713	17	79630115	-51.48768476	12.65312127	-4.06916868	1.09E-04			Northern shelf
8	cg11813198	12	57572413	-27.58812391	6.051050557	-4.55922879	1.80E-05	LRP1	Body	Southern shelf
9	cg15650274	9	91150624	-32.4440358	7.810020334	-4.154155101	8.04E-05	NXNL2; NXNL2	1st exon; 1st exon	Island
10	cg16540086	2	234615059	-25.91380715	4.690522873	-5.524716082	3.89E-07	UGT1A8; UGT1A10; UGT1A9; UGT1A7; UGT1A6; UGT1A6	Body; Body; Body; Body; Body; Body	
11	cg10526478	16	85936211	-33.78388656	8.205845228	-4.117051397	9.18E-05	IRF8	5'UTR	Southern shelf
12	cg00574649	19	45753669	-31.83553479	7.698596626	-4.135238711	8.60E-05	MARK4; MARK4	TSS1500; TSS1500	Northern shore

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
13	cg23584647	11	61717191	-35.47514755	8.664485171	-4.094316841	9.96E-05	BEST1; BEST1	TSS200; TSS200	
14	cg10914523	15	52393162	-28.73694965	6.777540913	-4.240025994	5.89E-05			
15	cg02466397	19	38714092	-39.36269183	9.796185767	-4.018165107	1.31E-04	DPF1; DPF1; DPF1; DPF1	Body; Body; Body; 5'UTR	Southern shore
16	cg03382278	2	221012092	-44.49218159	11.14523708	-3.992035458	1.43E-04			
17	cg17370413	3	10421792	-26.9420358	6.233656036	-4.32202798	4.36E-05	ATP2B2; ATP2B2	Body; Body	
18	cg04440551	2	25051151	-25.37643412	5.632099216	-4.505679525	2.20E-05	ADCY3	Body	
19	cg16504854	2	96333656	24.3399803	5.155615275	4.721062182	9.70E-06			
20	cg06521368	1	151483322	-42.78689454	10.83891991	-3.94752382	1.67E-04	CGN	TSS1500	Northern shore
21	cg05346689	5	1719318	-46.88283608	11.94148142	-3.926048573	1.80E-04			
22	cg05920424	19	51139819	26.44503221	6.22773156	4.246334633	5.76E-05	SYT3; SYT3; SYT3	Body; Body; Body	Northern shelf
23	cg07712411	4	175395859	-25.84753263	6.095559879	-4.24038696	5.88E-05			
24	cg13715631	5	134366914	-27.49750226	6.700862899	-4.103576311	9.64E-05	PITX1	Body	Island
25	cg17409356	8	39779075	25.51746845	6.076299234	4.199508198	6.82E-05	IDO1	Body	
26	cg03898845	1	117114009	30.85131892	7.789245742	3.960758196	1.60E-04	CD58; CD58; CD58	TSS1500; TSS1500; TSS1500	Southern shore
27	cg10851774	8	20113188	22.37352804	4.548244296	4.919157059	4.48E-06	LZTS1	TSS1500	
28	cg08372227	2	220254277	27.35175547	6.67606804	4.096985728	9.87E-05			Southern shore
29	cg23381867	2	46622124	-22.92751886	5.194477864	-4.413825501	3.11E-05			
30	cg12537728	11	13309596	31.48968711	7.978359118	3.946887655	1.68E-04	ARNTL; ARNTL; ARNTL	5'UTR; 5'UTR; 5'UTR	
31	cg24627820	8	108826547	26.96193379	6.577686539	4.098999494	9.80E-05			
32	cg14889758	17	75775462	-34.81991388	8.934151121	-3.897394773	1.99E-04			

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
33	cg08667991	3	49021370	-27.28989027	6.726348231	-4.057162867	1.14E-04	ARIH2	3'UTR	
34	cg17521120	19	51961666	-28.77728024	7.271047813	-3.957789988	1.61E-04	SIGLEC8; SIGLEC8	1st exon; 5'UTR	
35	cg16579593	2	137814280	-44.59440897	11.60932298	-3.841258361	2.42E-04	THSD7B	Body	
36	cg22985466	12	57992676	27.26376541	6.856342623	3.976429842	1.51E-04	PIP4K2C; PIP4K2C; PIP4K2C; PIP4K2C	Body; Body; Body; Body	
37	cg12284761	2	103216004	34.28469686	8.860534844	3.869371032	2.20E-04			
38	cg03016859	16	48239712	-22.03734527	5.076348542	-4.341180494	4.06E-05	ABCC11; ABCC11; ABCC11	Body; Body; Body	
39	cg23556358	1	203738697	-29.44793686	7.534359597	-3.908485715	1.92E-04	LAX1; LAX1; LAX1	TSS200; Body; Body	
40	cg16446162	15	46021526	-43.97915809	11.52056545	-3.817447875	2.63E-04			
41	cg15385345	15	40779019	-39.34211807	10.2879095	-3.824111989	2.57E-04			
42	cg13734707	2	10540304	24.70851942	6.043880751	4.088187779	1.02E-04	HPCAL1; HPCAL1; HPCAL1; HPCAL1; HPCAL1	5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR	
43	cg13510812	3	45429626	24.0223799	5.817536226	4.129304738	8.79E-05	LARS2	TSS1500	Northern shore
44	cg07162198	20	45338753	24.56927606	6.009699624	4.08827023	1.02E-04	SLC2A10	Body	Southern shore
45	cg15892763	11	120262862	-21.72620963	5.045249915	-4.306270253	4.62E-05	ARHGEF12	Body	
46	cg12838765	1	29779135	-22.27268121	5.350374161	-4.162826849	7.79E-05			
47	cg19558628	14	24801616	-30.62569993	7.967257317	-3.843945126	2.40E-04	ADCY4	Body	Northern shore
48	cg22573917	12	49371628	44.83445444	11.87292421	3.776193097	3.02E-04	WNT1	TSS1500	Northern shore
49	cg05880390	20	46983876	-38.91221036	10.26394387	-3.791155801	2.87E-04			
50	cg09548897	16	341277	-21.59215123	5.136883472	-4.203356246	6.73E-05	AXIN1; AXIN1	Body; Body	Island
51	cg00316080	3	124866646	20.04210791	4.327396068	4.631447548	1.37E-05	SLC12A8	Body	
52	cg06768385	1	214557399	-39.20666814	10.36612355	-3.782191864	2.96E-04	PTPN14	Body	Island

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
53	cg01866308	2	128391781	-25.06138622	6.355676433	-3.943150108	1.70E-04	MYO7B	Body	
54	cg17451760	11	65327156	-26.51947905	6.830356991	-3.882590484	2.10E-04	LTBP3; LTBP3; LTBP3	TSS1500; TSS1500; TSS1500	Southern shore
55	cg19144019	17	26875307	-23.33236682	5.808182009	-4.0171549	1.31E-04	UNC119; UNC119	Body; Body	Northern shelf
56	cg05295536	6	30749628	-24.06442958	6.034063227	-3.988097021	1.45E-04			
57	cg01182430	18	75680550	-26.43874272	6.804857729	-3.885274868	2.08E-04			
58	cg05851148	5	179225977	20.85077308	4.855026731	4.294677298	4.82E-05	MGAT4B; MIR1229; MGAT4B	Body; TSS1500; Body	Southern shore
59	cg11148307	10	134671663	-42.09361869	11.15819711	-3.772439067	3.06E-04			Northern shore
60	cg18857951	14	94395516	-22.16405248	5.400848258	-4.103809517	9.63E-05	FAM181A	3'UTR	
61	cg25923056	19	41306455	-21.36144968	5.111019778	-4.179488753	7.34E-05	EGLN2; EGLN2; EGLN2	1st exon; 5'UTR; 5'UTR	Northern shore
62	cg25077777	8	144676885	-20.7454948	4.876850258	-4.253871597	5.60E-05	EEF1D; EEF1D; EEF1D; EEF1D; EEF1D; EEF1D; EEF1D	5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR	Southern shore
63	cg24321566	5	171844324	-24.23758473	6.131810631	-3.952761457	1.64E-04	SH3PXD2B; SH3PXD2B	Body; Body	
64	cg11125714	19	46916944	33.27436746	8.773679216	3.792521546	2.86E-04	CCDC8	TSS200	Southern shore
65	cg02676315	12	15154181	23.17193408	5.80433155	3.992179613	1.43E-04	LINC01489	TSS1500	
66	cg01432450	19	39201446	-31.02206426	8.160326411	-3.801571493	2.77E-04	ACTN4	Body	
67	cg13295950	1	208361419	-29.9268106	7.843503912	-3.815489982	2.64E-04	PLXNA2	Body	
68	cg19150929	4	22714279	-26.72965564	6.918043889	-3.863759188	2.24E-04	GBA3; GBA3; GBA3; GBA3; GBA3; GBA3	TSS1500; TSS1500; Body; Body; Body; Body	
69	cg19727360	20	45522404	-21.57934303	5.230599918	-4.125596177	8.91E-05	EYA2; EYA2	TSS1500; TSS1500	Northern shore
70	cg02672229	10	101594256	-19.72796808	4.396588932	-4.487107708	2.36E-05	ABCC2	Body	
71	cg18260869	4	187619486	-28.90731256	7.582830499	-3.812206083	2.67E-04	FAT1	Body	

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
72	cg01707361	17	77079031	-22.47224711	5.657361376	-3.972213478	1.54E-04	ENGASE	Body	
73	cg03286076	16	1264599	26.20664879	6.790164784	3.85950115	2.27E-04	CACNA1H; CACNA1H	Body; Body	Northern shore
74	cg08781611	14	69132299	-30.10637268	7.96069927	-3.781875394	2.97E-04			
75	cg24954895	5	140800761	-25.47475091	6.599701576	-3.859985277	2.27E-04	PCDHGA4; PCDHGA9; PCDHGA1; PCDHGB1; PCDHGB6; PCDHGA11; PCDHGA11; PCDHGA11; PCDHGB3; PCDHGA11; PCDHGB7; PCDHGA6; PCDHGA8; PCDHGA11; PCDHGA10; PCDHGA5; PCDHGB4; PCDHGA3; PCDHGA11; PCDHGA2; PCDHGA7; PCDHGB2; PCDHGB5	Body; Body; Body; Body; Body; 5'UTR; 5'UTR; 5'UTR; Body; 1st exon; Body; Body; Body; 1st exon; Body; Body; Body; Body; 1st exon; Body; Body; Body; Body	Island
76	cg10408430	11	73979900	30.22629812	8.015247072	3.771099986	3.08E-04	P4HA3	Body	
77	cg15349661	10	4169941	26.13778312	6.84312829	3.819566433	2.61E-04			
78	cg13364082	10	46994040	-19.06062844	4.32693165	-4.405114288	3.21E-05	GPRIN2	5'UTR	Southern shore
79	cg15536864	22	47532672	-21.16380802	5.248676563	-4.032217983	1.24E-04	TBC1D22A; TBC1D22A; TBC1D22A; TBC1D22A; TBC1D22A	Body; Body; Body; Body; Body	

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
80	cg15115810	5	167753517	-34.8252674	9.325059349	-3.734589357	3.48E-04	WWC1; WWC1; WWC1	Body; Body; Body	
81	cg18734051	9	136255344	-25.36535319	6.620480035	-3.831346527	2.50E-04	STKLD1; STKLD1	Body; Body	
82	cg15206586	9	5550375	-24.29828091	6.303737929	-3.854582976	2.31E-04	PDCD1LG2	Body	
83	cg16181011	2	20417099	-34.47473581	9.236102696	-3.732606376	3.51E-04	SDC1; SDC1	Body; Body	
84	cg17930897	10	8170650	-23.35602889	6.034970672	-3.87011473	2.19E-04			
85	cg11941811	3	132927140	21.34275729	5.388453881	3.960831392	1.60E-04	TMEM108; TMEM108; TMEM108	5'UTR; 5'UTR; 5'UTR	
86	cg06521247	1	3209787	-19.68928439	4.736641323	-4.156802901	7.96E-05	PRDM16; PRDM16	Body; Body	
87	cg03571308	12	52477022	-24.23636307	6.299233317	-3.847509983	2.37E-04	OR7E47P	Body	Southern shelf
88	cg11611536	17	74907612	-32.2254732	8.635104312	-3.731914756	3.52E-04	MGAT5B; MGAT5B; MGAT5B	Body; Body; Body	
89	cg13533070	2	152032584	-30.13641786	8.054754137	-3.741444785	3.40E-04			
90	cg18534396	16	962918	-22.68766397	5.867671901	-3.866552928	2.22E-04	LMF1; LMF1; LMF1	Body; Body; Body	Island
91	cg24427716	9	124466639	-18.61244123	4.254597423	-4.374665658	3.59E-05	DAB2IP	Body	
92	cg19008088	10	102890038	21.65625937	5.542613487	3.907228859	1.93E-04	TLX1; TLX1NB	TSS1500; 5'UTR	Northern shore
93	cg07695566	17	61525112	18.37725397	4.166146776	4.411091342	3.14E-05	CYB561	TSS1500	Southern shore
94	cg09177486	1	145451575	18.85408827	4.458612325	4.228689758	6.14E-05	NBPF20; NBPF10; NBPF10	Body; Body; Body	
95	cg04391313	16	52908117	-17.9708614	3.599773852	-4.99221955	3.36E-06			
96	cg26656815	12	1906835	18.47320387	4.267929464	4.328376095	4.26E-05	CACNA2D4	Body	Southern shore
97	cg00503321	17	7297230	26.03678435	6.909922757	3.768028279	3.11E-04	PLSCR3	Body	Island
98	cg04864149	1	95006838	31.63506996	8.571561743	3.690700821	4.04E-04	F3	Body	Island
99	cg03858460	16	75569852	-21.55830264	5.558067719	-3.878740551	2.13E-04	CHST5	TSS1500	

Supplementary Table 3. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
100	cg02549588	8	80992699	17.88142706	3.851855693	4.642288934	1.31E-05	TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52; TPD52	1st exon; 1st exon; 1st exon; 1st exon; 5'UTR; 5'UTR; 5'UTR; 5'UTR; Body; Body; Body; Body; Body; Body; Body; Body; Body	

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the t-statistics, the accompanying p values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined p value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 4. Top 100 ranked differentially unmodified positions (DUPs) in the locus coeruleus (LC)

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg05775895	3	12838266	277.1749287	61.90650359	4.477315186	2.91E-05	CAND2; CAND2	1st exon; 1st exon	Island
2	cg05224975	20	57049253	-117.2800978	23.19068893	-5.057206288	3.35E-06	APCDD1L; APCDD1L; APCDD1L	Body; Body; Body	
3	cg11267619	7	29604021	491.3782777	116.4528028	4.219548742	7.32E-05	PRR15	5'UTR	Island
4	cg11959007	13	76432094	-160.2558812	37.57580802	-4.264868533	6.24E-05	LMO7; LMO7	Body; 3'UTR	
5	cg01486300	11	71885949	-115.0633189	26.33007239	-4.370034279	4.29E-05			
6	cg15531185	3	167415859	-91.81059444	19.25808254	-4.76737984	1.00E-05	PDCD10; PDCD10; PDCD10	Body; Body; Body	
7	cg22068764	1	170501645	-253.1292187	60.80758413	-4.162790257	8.93E-05	GORAB; GORAB; GORAB	Body; Body; Body	
8	cg00604202	6	99272987	85.02374214	15.77736084	5.388971135	9.30E-07			Island
9	cg07330172	12	59932731	-125.2908759	30.42753168	-4.117681223	1.05E-04			
10	cg07209547	7	16873191	-79.45576479	15.75975264	-5.041688574	3.56E-06			

Supplementary Table 4. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
11	cg04582364	5	52778441	-109.2618288	26.41805025	-4.135877848	9.81E-05	FST; FST	Body; Body	Southern shore
12	cg05254590	8	55382640	-166.06681	40.88034884	-4.062264994	1.27E-04			Island
13	cg01061283	8	63147364	-109.2061951	26.44831182	-4.129042181	1.00E-04			
14	cg06793461	12	103987955	-79.82935514	17.56174358	-4.54563949	2.27E-05	STAB2	Body	
15	cg04988171	7	86847564	-87.931112	20.29767409	-4.332078228	4.91E-05	TMEM243	Body	Northern shore
16	cg10312966	6	121717017	-99.8590694	23.88837359	-4.180237262	8.40E-05			
17	cg21610556	4	181452599	-104.6540147	25.51927941	-4.100978443	1.11E-04			
18	cg02450124	15	58429751	-118.4688347	29.52053808	-4.013098759	1.50E-04	AQP9	TSS1500	
19	cg16965936	21	43255477	69.06390774	14.1476788	4.881642334	6.53E-06	PRDM15; PRDM15	Body; Body	Island
20	cg21839449	7	157956042	-76.43803553	17.37056591	-4.400434385	3.84E-05	PTPRN2; PTPRN2; PTPRN2	Body; Body; Body	Northern shore
21	cg12413852	18	55820185	-105.8824946	26.24354546	-4.034610901	1.39E-04	NEDD4L; NEDD4L; NEDD4L; NEDD4L; NEDD4L; NEDD4L; NEDD4L; NEDD4L	Body; Body; Body; 5'UTR; 5'UTR; 5'UTR; Body; Body	
22	cg00050133	4	178167583	-74.14812874	16.82202973	-4.407799174	3.74E-05			
23	cg12866769	15	67041828	-68.96777812	15.03064302	-4.588478218	1.94E-05	SMAD6; SMAD6; SMAD6	Body; Body; Body	
24	cg21373732	14	99886511	-90.52854458	22.106226	-4.095160548	1.13E-04	SETD3; SETD3	Body; Body	
25	cg18675840	1	245703471	79.07185283	19.01162987	4.159130668	9.05E-05	KIF26B	Body	
26	cg19712963	1	24069298	-86.28693736	21.5032579	-4.012737872	1.50E-04	TCEB3	TSS1500	Northern shore
27	ch.17.206577 3F	17	76200219	-57.63521382	12.1632779	-4.73846066	1.12E-05	AFMID; AFMID	Body; Body	
28	cg13482990	2	140582839	-62.95580061	14.70937482	-4.279978001	5.91E-05			
29	cg12746706	6	169276508	71.52095017	17.41755883	4.106255697	1.09E-04			

Supplementary Table 4. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
30	cg11383275	2	67623023	-57.42414781	12.91303668	-4.446990218	3.25E-05	ETAA1	TSS1500	Northern shore
31	cg17690975	6	76564689	-84.5572518	21.22702405	-3.983471804	1.66E-04	MYO6; MYO6	Body; Body	
32	cg06538616	14	36402316	-61.97513982	14.69834924	-4.216469402	7.40E-05			
33	cg03550821	6	143474125	-100.6463181	25.79086158	-3.902402321	2.19E-04	AIG1	Body	
34	cg00875467	3	76875348	-90.08720841	22.94001344	-3.927077403	2.01E-04			
35	cg25113483	8	28321280	55.31737607	12.56368615	4.402957494	3.81E-05	FBXO16; FBXO16	Body; Body	
36	cg06587475	8	26279631	108.4225369	27.98107071	3.874853038	2.40E-04			
37	cg23106652	13	47368091	-82.50245959	20.92656325	-3.942475341	1.91E-04	ESD	5'UTR	Northern shelf
38	cg27163126	8	103953809	-76.18693736	19.16915885	-3.974453858	1.71E-04	AZIN1-AS1; AZIN1-AS1	Body; Body	
39	cg04340928	16	55067613	-229.9522895	60.14561477	-3.82325944	2.85E-04			
40	cg04169380	6	161516638	-65.4055128	16.24186798	-4.026969858	1.43E-04	MAP3K4; MAP3K4	Body; Body	
41	cg13851625	7	26676500	-91.05879487	23.32289206	-3.904266874	2.17E-04	C7orf71	TSS1500	
42	cg00733324	14	67845244	-90.26208611	23.15314562	-3.898480474	2.22E-04	EIF2S1	Body	
43	cg01907875	18	41650428	-50.80036739	11.42383769	-4.446874052	3.25E-05			
44	cg22665099	10	97055150	-63.47384816	15.83203598	-4.009203126	1.52E-04			Island
45	cg21524032	13	50030794	-84.7167459	21.68803867	-3.906150629	2.16E-04	SETDB2; SETDB2	Body; Body	
46	cg12545993	8	119964636	84.35871185	21.58487097	3.908233316	2.14E-04	TNFRSF11B	TSS1500	Island
47	cg00335757	8	80767166	-50.17742896	11.39872657	-4.402020579	3.82E-05	LOC101927040	Body	
48	cg15057359	2	77750866	-58.13929618	14.29377957	-4.06745437	1.24E-04	LRRTM4; LRRTM4	TSS1500; TSS1500	
49	cg14947281	10	101501169	-93.56885004	24.28207497	-3.853412452	2.58E-04	CUTC	Body	
50	cg25664353	6	114019348	69.73878629	17.78683879	3.920808364	2.06E-04			

Supplementary Table 4. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
51	cg08854791	5	179407590	-88.8828497	23.10663808	-3.846637031	2.64E-04	RNF130; RNF130	Body; Body	
52	cg09406933	16	75254017	75.04587899	19.30226984	3.88793026	2.30E-04	CTRB1	Body	
53	cg21735115	12	9800259	67.48839544	17.24501171	3.913502441	2.11E-04	LOC374443	TSS1500	Northern shore
54	cg06126581	17	16229132	-94.55382406	24.81114161	-3.810942098	2.97E-04	PIGL	Body	
55	cg12299771	14	50012007	-66.08017425	16.86871037	-3.917322238	2.08E-04			
56	cg07331152	5	7983756	-99.83838214	26.42823866	-3.777716079	3.32E-04			
57	cg13503413	10	60937257	-56.23722473	14.07702682	-3.994964666	1.60E-04	PHYHIPL; PHYHIPL; PHYHIPL	5'UTR; Body; 1st exon	Southern shore
58	cg11030597	9	7537075	-52.43686252	12.96970653	-4.043026138	1.35E-04			
59	cg03304589	7	63734144	81.79287041	21.34492253	3.831959113	2.77E-04			
60	cg21339170	3	176763896	-43.80496222	9.834346548	-4.454282956	3.17E-05	TBL1XR1	Body	
61	cg06346857	1	21877579	-132.1273416	35.4614307	-3.725945035	3.94E-04	ALPL; ALPL	5'UTR; TSS1500	
62	cg17428843	1	221845800	-54.21065635	13.52522309	-4.008115504	1.53E-04			
63	cg20066716	7	49815181	107.8203954	28.79033256	3.745020841	3.70E-04	VWC2	Body	Island
64	cg07267974	8	62539945	-86.51643464	22.83793272	-3.788277849	3.21E-04	ASPH; ASPH; ASPH; ASPH; ASPH; ASPH; ASPH; ASPH; ASPH	Body; Body; Body; Body; Body; Body; Body; Body; Body	
65	cg03783777	12	2822034	56.88671872	14.43537675	3.940785177	1.92E-04			
66	cg01603981	5	80043300	63.72244227	16.40009695	3.885491804	2.32E-04	MSH3	Body	
67	cg19688999	10	59961765	-91.63601944	24.35863463	-3.761952212	3.50E-04	IPMK	Body	
68	cg22071224	2	171572348	-76.97768465	20.17434666	-3.815622184	2.93E-04	LOC440925; SP5	TSS1500; Body	Island
69	cg07715503	4	170121522	-43.66400595	10.24622022	-4.261474475	6.31E-05	SH3RF1	Body	
70	cg03882355	2	30939430	-48.0056749	11.82659959	-4.05912744	1.28E-04			
71	cg00632007	6	1598984	-79.77615041	21.0988738	-3.78106202	3.29E-04			Northern shore
72	cg09445207	22	18341142	-62.07136883	16.02901612	-3.872437857	2.42E-04	MICAL3; MICAL3	Body; Body	

Supplementary Table 4. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
73	cg04724720	11	65640939	-107.1747446	28.83984975	-3.716203292	4.07E-04	EFEMP2	TSS1500	Island
74	cg00213182	4	36718783	43.93339074	10.47226617	4.195213338	7.97E-05			
75	cg23276399	14	102427771	40.15556584	8.58706961	4.676282791	1.41E-05			Northern shelf
76	cg13065537	3	193853752	193.2726711	52.42873304	3.686388357	4.49E-04	HES1	TSS200	Northern shore
77	cg15006851	6	134830131	-42.69305687	10.05008621	-4.248028919	6.62E-05			
78	cg23228887	2	190543482	54.48384069	13.84150759	3.936264915	1.95E-04	ANKAR	Body	Southern shelf
79	cg07016184	15	96884187	361.2182177	98.47901323	3.667971539	4.77E-04			Island
80	cg22598496	3	190980644	-41.36553681	9.538835854	-4.336539327	4.83E-05			
81	cg11162637	15	101818008	-67.38524497	17.63315699	-3.821507685	2.87E-04	SELS; SELS	TSS1500; TSS1500	Southern shore
82	cg19830670	4	75097154	-38.57295556	8.104093787	-4.759687705	1.03E-05	MTHFD2L	Body	
83	cg24284976	3	73862941	-45.64727152	11.26959354	-4.050480736	1.32E-04			
84	cg11932767	6	32864607	-87.2345496	23.35711691	-3.73481667	3.83E-04	LOC100294145; LOC100294145	Body; Body	Southern shore
85	cg25969802	1	98088314	-42.40963059	10.09823013	-4.199709262	7.85E-05	DPYD	Body	
86	cg00913912	2	153282173	-49.48829303	12.49126144	-3.961833099	1.79E-04	FMNL2	Body	
87	cg26545307	2	141958753	-117.4407146	31.9717525	-3.67326485	4.69E-04	LRP1B	Body	
88	cg16596716	2	75825456	65.37845586	17.20697215	3.799532846	3.09E-04			
89	cg03122917	3	130073646	-40.10117636	9.356188932	-4.286058848	5.78E-05	COL6A5; COL6A5; COL6A5	5'UTR; 5'UTR; Body	
90	cg07522285	3	14929139	-47.46746089	11.93946621	-3.975676974	1.71E-04	FGD5	Body	
91	cg26033238	2	121231363	-37.2249777	7.906584494	-4.708098387	1.25E-05			
92	cg12914151	17	62309147	-72.954197	19.39169508	-3.762136147	3.50E-04	TEX2; TEX2; TEX2	5'UTR; 5'UTR; TSS1500	
93	cg01039664	16	69027363	60.69854053	15.8090411	3.839482744	2.70E-04	TANGO6	Body	
94	cg11976592	6	76078187	-42.50241174	10.32768232	-4.115387209	1.05E-04	FILIP1	Body	
95	cg23501964	20	19866810	-47.76817827	12.05803804	-3.961521609	1.79E-04			

Supplementary Table 4. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t* value	*p* value	UCSC annotation	Gene feature	CpG island feature
96	cg11375836	3	43896529	65.62270014	17.40865841	3.769543786	3.41E-04			
97	cg07274676	6	28232729	-58.94546216	15.39157061	-3.829723662	2.79E-04			Northern shelf
98	cg15865113	14	85853478	-45.55454077	11.43277763	-3.984555832	1.65E-04			
99	cg06698686	13	49796489	-61.54837155	16.1837455	-3.803098087	3.05E-04	MLNR	Body	Southern shore
100	cg03964554	17	59521084	-41.31063879	9.970708664	-4.143199865	9.56E-05			

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the t-statistics, the accompanying p values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined p value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 5. Top 100 ranked differentially methylated positions (DMPs) in the locus coeruleus (LC)

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg19265143	1	28907637	-235.7018337	46.86845394	-5.029008083	3.74E-06	SNHG12; SNORA44; SNORA16A; SNORA61	Body; TSS1500; TSS200; TSS1500	Northern shore
2	cg06793267	3	57993846	-367.1921366	80.20683065	-4.578065654	2.02E-05	FLNB; FLNB; FLNB; FLNB	TSS1500; TSS1500; TSS1500; TSS1500	Island
3	cg27254871	2	206750458	-193.7961418	41.95441751	-4.619207065	1.73E-05			
4	cg08946575	1	85931072	-84.90253179	18.79217286	-4.517973116	2.51E-05	DDAH1; DDAH1	TSS200; 5'UTR	Island
5	cg19108372	1	183605267	-224.8127272	53.59715679	-4.194489795	7.99E-05	ARPC5; ARPC5; RGL1; RGL1; RGL1; RGL1; RGL1; RGL1	TSS200; TSS200; 1st exon; 1st exon; 1st exon; 5'UTR; 5'UTR; 5'UTR	Island
6	cg04845871	1	207996319	-64.61917125	14.19566479	-4.552035582	2.22E-05			
7	cg00930873	15	58357973	-64.81381417	14.51794997	-4.464391619	3.05E-05	ALDH1A2; ALDH1A2	TSS200; TSS200	Island
8	cg01805869	2	26915772	-172.2016798	41.58645181	-4.140812025	9.64E-05	KCNK3	1st exon	Island

Supplementary Table 5. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
9	cg27277034	3	40575533	-59.99136991	13.45651713	-4.45816472	3.12E-05	ZNF621; ZNF621	3'UTR; 3'UTR	
10	cg20670037	19	58639808	-62.11879931	14.39166923	-4.316302598	5.19E-05	ZNF329	Body	
11	cg09279544	19	8674388	-279.4581039	69.29420427	-4.032921755	1.40E-04	ADAMTS10	5'UTR	Island
12	cg24396624	3	25469860	-66.56853228	15.8708946	-4.19437807	7.99E-05	RARB; RARB; RARB; RARB	5'UTR; 1st exon; 1st exon; 5'UTR	
13	ch.17.2065773F	17	76200219	54.81949247	12.70372863	4.315228548	5.21E-05	AFMID; AFMID	Body; Body	
14	cg07499372	19	57631232	53.70540771	12.50503602	4.294702358	5.61E-05	USP29	TSS1500	Southern shore
15	cg25376657	17	16229039	54.96947487	12.99957845	4.228558264	7.09E-05	PIGL	Body	
16	cg12862167	14	91752171	49.77703955	11.53716524	4.31449481	5.23E-05	CCDC88C	Body	Northern shelf
17	cg16474684	4	2470049	70.66200586	17.17760787	4.113611533	1.06E-04	RNF4	TSS1500	Northern shore
18	cg15054586	19	50194696	-329.2565526	84.40657547	-3.90084008	2.20E-04	CPT1C; CPT1C; CPT1C; CPT1C; CPT1C	5'UTR; 5'UTR; 5'UTR; 5'UTR; Body	Island
19	cg12790874	13	100310405	-211.1199738	53.73915257	-3.928606308	2.00E-04	CLYBL	Body	Island
20	cg23968184	2	196438113	67.56868886	16.71459583	4.042496125	1.36E-04			
21	cg00558156	8	23261618	-53.37163384	12.90505376	-4.135715732	9.82E-05	LOXL2; LOXL2	5'UTR; 1st exon	Island
22	cg00785478	1	120604399	-40.41689339	9.457338232	-4.273601345	6.05E-05	NOTCH2; NOTCH2	Body; Body	
23	cg09732027	6	29984811	62.57170533	15.38955171	4.06585627	1.25E-04	NCRNA00171	Body	
24	cg22541679	13	78493666	-322.4609998	83.45555127	-3.863865194	2.49E-04	EDNRB; EDNRB; EDNRB	TSS1500; 5'UTR; TSS1500	Southern shore
25	cg06059849	1	109656930	-92.65139737	23.5783908	-3.929504695	2.00E-04	KIAA1324; C1orf194	1st exon; TSS1500	Southern shore
26	cg18643191	17	20465437	55.24290442	13.70432539	4.031056098	1.41E-04			
27	cg18572198	5	5050553	70.39127589	17.80383805	3.953713559	1.84E-04	LINC01020	Body	

Supplementary Table 5. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
28	cg21295467	11	123986166	-294.3159433	76.5772297	-3.843387184	2.67E-04	VWA5A; VWA5A; VWA5A; VWA5A; VWA5A; VWA5A	1st exon; 5'UTR; 1st exon; 5'UTR; 5'UTR; 1st exon	
29	cg08410015	10	119862094	-55.33454187	13.77927176	-4.015781302	1.49E-04	CASC2; CASC2	Body; Body	
30	cg02184543	10	7226492	-32.95182337	7.299555856	-4.51422306	2.55E-05	SFMBT2; SFMBT2	Body; Body	
31	cg02759617	8	88746918	53.88857599	13.42575357	4.013821325	1.50E-04			
32	cg00242449	17	7340872	-33.75850906	7.795128071	-4.330718976	4.93E-05	FGF11; TMEM102; FGF11	TSS1500; 3'UTR; TSS1500	Island
33	cg21484213	2	142888868	-98.14687522	25.23726021	-3.888967123	2.29E-04	LRP1B; LRP1B	5'UTR; 1st exon	Southern shore
34	cg16739342	10	12221327	-66.21527713	16.80114786	-3.941116267	1.92E-04	NUDT5	Body	
35	cg00573948	7	101330867	-32.8866926	7.491535447	-4.389846759	3.99E-05			Northern shore
36	cg23128263	1	66797473	-69.74425212	17.84475114	-3.908390293	2.14E-04	PDE4B; PDE4B; PDE4B; PDE4B; PDE4B; PDE4B	TSS1500; Body; Body; Body; Body; Body	
37	cg13497069	6	29426119	36.02036865	8.650016905	4.16419633	8.89E-05	OR2H1	TSS200	
38	cg01072259	2	37946508	47.18252359	11.79960913	3.998651401	1.58E-04			
39	cg25533997	18	35144259	-29.25234088	6.447789117	-4.53680174	2.35E-05	BRUNOL4; BRUNOL4; BRUNOL4; BRUNOL4	Body; Body; Body; Body	Northern shore
40	cg21211367	2	162094118	-143.3632098	37.53075096	-3.819886522	2.89E-04			Northern shore
41	cg25699851	15	68624713	28.10427811	5.69458373	4.935264707	5.33E-06	ITGA11	Body	
42	cg14417711	11	6478126	184.2694812	48.45989271	3.802515252	3.06E-04	TRIM3; TRIM3	Body; Body	Island
43	cg21032945	5	5652728	56.07724425	14.32439628	3.914806821	2.10E-04			
44	cg10960253	1	92296137	-47.43755101	11.97918937	-3.959996754	1.80E-04	TGFBR3; TGFBR3; TGFBR3; TGFBR3	Body; Body; Body; Body	
45	cg25565383	2	17699773	-311.8402925	82.43573141	-3.782829207	3.27E-04	RAD51AP2	TSS200	
46	cg14678430	7	128494150	36.45977698	8.930030378	4.082827878	1.18E-04	FLNC; FLNC	Body; Body	Island

Supplementary Table 5. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
47	cg02091786	11	62415525	29.13592149	6.650703005	4.380878453	4.12E-05	GANAB; GANAB; INTS5	TSS1500; TSS1500; Body	Southern shore
48	cg27408178	4	108974899	35.17052961	8.524264859	4.125931114	1.02E-04	LEF1; LEF1; LEF1; LEF1	3'UTR; 3'UTR; 3'UTR; Body	
49	cg04331204	8	143835180	-38.0223288	9.38773439	-4.050213525	1.32E-04	LYPD2	TSS1500	
50	cg25840838	11	47488535	-32.2135785	7.689176559	-4.189470518	8.13E-05	CELF1; CELF1; CELF1; CELF1; CELF1	3'UTR; 3'UTR; 3'UTR; 3'UTR; 3'UTR	
51	cg00298218	1	83955790	27.65290416	5.605105241	4.933520956	5.37E-06			
52	cg10516989	16	2868780	50.4374971	12.88219073	3.915288801	2.09E-04	PRSS21; PRSS21; PRSS21; PRSS21; PRSS21	Body; Body; Body; Body; Body	Southern shore
53	cg17191367	1	50570612	28.94258485	6.656804725	4.347819419	4.64E-05	ELAVL4; ELAVL4; ELAVL4	TSS1500; Body; Body	
54	cg17085352	12	54332026	-62.96593443	16.33749869	-3.854074337	2.57E-04	HOXC13	TSS1500	Northern shore
55	cg09484243	5	134463743	43.387776	10.97064548	3.954897283	1.83E-04	C5orf66	5'UTR	
56	cg14881567	19	16583117	129.2161761	34.0487605	3.795033187	3.14E-04	EPS15L1	TSS1500	Island
57	cg12816995	18	76699288	27.70430241	6.102116927	4.540113331	2.32E-05			
58	cg02853268	11	1969721	-31.42722217	7.518185669	-4.180160421	8.40E-05	MRPL23	Body	Southern shore
59	cg16871328	16	57622905	-36.33740557	9.009604709	-4.033185333	1.40E-04			
60	cg09377423	6	37751308	-57.32538259	14.87334707	-3.854235522	2.57E-04			
61	cg00171785	11	6044373	-109.3159005	28.79988089	-3.795706687	3.13E-04			
62	cg01645899	12	2996093	-27.64786899	6.080748878	-4.546786842	2.26E-05	RHNO1; RHNO1; RHNO1; RHNO1; RHNO1; RHNO1	Body; Body; Body; Body; Body; Body; Body	Northern shelf
63	cg00789116	10	120761718	27.17464154	5.720713246	4.750219138	1.07E-05			
64	cg10639981	19	50762343	-28.40608855	6.540712145	-4.342965708	4.72E-05	MYH14; MYH14; MYH14	Body; Body; Body	Island
65	cg01961069	16	67189629	-241.9642251	64.3185657	-3.761965499	3.50E-04	TRADD	Body	Island

Supplementary Table 5. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
66	cg25813748	13	103518224	34.72788897	8.537891523	4.06750178	1.24E-04	BIVM-ERCC5; ERCC5	Body; Body	
67	cg21931174	14	20977318	36.06546074	8.952903048	4.02835377	1.42E-04	RNASE10	TSS1500	
68	cg03783039	4	184401720	-39.36036334	9.9391724	-3.960124823	1.80E-04			Northern shelf
69	cg05106686	12	68738010	-31.83845792	7.709244681	-4.129906267	1.00E-04			
70	cg20740028	7	28450406	-264.5634476	70.82354554	-3.735529555	3.82E-04	CREB5	5'UTR	Southern shore
71	cg08351753	20	43408177	32.70812488	8.003467729	4.086744145	1.16E-04	RIMS4; RIMS4	Body; Body	
72	cg18371836	10	122553700	-25.81279625	4.673452504	-5.523282033	5.48E-07			
73	cg05889864	6	122972445	33.22023588	8.159772074	4.071221056	1.23E-04	PKIB; PKIB; PKIB; PKIB; PKIB; PKIB	TSS1500; 5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR	
74	cg09591921	16	24126487	26.60000141	5.82456101	4.566868021	2.10E-05	PRKCB; PRKCB	Body; Body	
75	cg14695920	9	94997513	-44.76674805	11.50936992	-3.889591554	2.28E-04	IARS; IARS; IARS	Body; Body; Body	
76	cg18537222	3	12435731	35.10770195	8.79050441	3.993821095	1.60E-04	PPARG; PPARG; PPARG; PPARG	Body; Body; Body; Body	
77	cg09295258	11	119274583	27.03969421	6.224598626	4.344006069	4.71E-05	USP2-AS1	Body	
78	cg08114611	5	94172463	48.2995411	12.5810682	3.839065201	2.71E-04	MCTP1; MCTP1; MCTP1	Body; Body; Body	
79	cg04393637	17	75790091	33.39864697	8.333363452	4.007823151	1.53E-04			Southern shore
80	cg16093863	3	144393272	37.7990054	9.686043352	3.902419597	2.19E-04			
81	cg07615572	17	7405889	-30.85115139	7.587855595	-4.065859057	1.25E-04	POLR2A	Body	
82	cg18754985	3	98237750	24.79688456	5.168837552	4.79738129	8.96E-06	CLDND1; CLDND1; CLDND1; CLDND1; CLDND1; CLDND1	Body; Body; Body; Body; Body; Body	Northern shelf
83	cg23845609	11	47609357	-25.42658417	5.674457124	-4.480884007	2.87E-05	FAM180B	Body	Northern shore
84	cg12289867	10	1623710	-25.42325971	5.673408822	-4.481125988	2.87E-05	ADARB2	Body	
85	cg10434137	13	44032630	27.7377438	6.66412008	4.1622515	8.95E-05	ENOX1-AS2; ENOX1; ENOX1; ENOX1	Body; 5'UTR; 5'UTR; 5'UTR	

Supplementary Table 5. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
86	cg13063967	12	12418943	-98.24067657	26.41868415	-3.718605969	4.04E-04	LRP6	Body	Northern shore
87	cg26911948	12	2157278	66.72515328	17.72825916	3.76377357	3.48E-04	CACNA1C-IT2	TSS1500	
88	cg25046661	3	151839485	24.50125959	4.935566696	4.964224191	4.78E-06			
89	cg03002059	8	11402647	-30.9327288	7.717597682	-4.008077393	1.53E-04	BLK	Body	
90	cg13221907	6	170755442	36.24851567	9.313280975	3.892131652	2.26E-04			Northern shore
91	cg17587467	11	64652643	41.45384106	10.80850257	3.83529918	2.74E-04			
92	cg04572695	6	100441722	-166.5267694	45.06486369	-3.69526846	4.36E-04	MCHR2-AS1; MCHR2; MCHR2	TSS200; 5'UTR; 5'UTR	Island
93	cg22917346	13	44453235	-224.195738	60.77127164	-3.689173057	4.45E-04	CCDC122; C13orf31; C13orf31	5'UTR; TSS200; TSS1500	Island
94	cg15052158	10	131570747	-28.36596078	6.955801982	-4.078028796	1.20E-04			
95	cg18451156	12	133424373	-24.3256532	5.241503931	-4.640968225	1.60E-05	CHFR; CHFR; CHFR; CHFR; CHFR	Body; Body; Body; Body; Body	
96	cg25396488	6	29641118	33.03594372	8.394943278	3.935219408	1.96E-04	ZFP57	Body	
97	cg21737243	7	99012691	-26.66290258	6.366083733	-4.188273938	8.17E-05	BUD31	Body	
98	cg23164993	1	65613294	-408.3883075	111.4238386	-3.665178948	4.81E-04	AK3L1; AK3L1; AK3L1; AK3L1	5'UTR; TSS1500; TSS1500; 1st exon	Island
99	cg12898370	12	24737779	31.58801994	8.005029959	3.946021452	1.89E-04	C12orf67	TSS1500	
100	cg08554039	15	40262416	-25.67337457	6.074517404	-4.226405632	7.14E-05	EIF2AK4	Body	

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the t-statistics, the accompanying p values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined *p* value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 6. Top 100 ranked differentially hydroxymethylated positions (DHPs) in the locus coeruleus (LC)

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
1	cg00053361	8	127250180	-62.14971407	14.05520712	-4.421828406	3.56E-05			
2	cg23968184	2	196438113	-66.51323794	15.31426896	-4.343219914	4.72E-05			
3	cg19591003	14	94392757	-47.05792749	9.696387631	-4.853140085	7.27E-06	FAM181A; C14orf86	Body; TSS200	
4	cg04994405	1	65731836	57.21702512	13.038981	4.388151584	4.02E-05	DNAJC6	Body	Island
5	cg00347713	15	81680130	-45.2902628	9.862030362	-4.592387281	1.91E-05	TMC3-AS1	Body	
6	cg13020390	1	88837555	64.27111371	15.11960286	4.250846686	6.55E-05			
7	cg13137032	4	5904897	61.45395872	14.65213362	4.194198629	8.00E-05			
8	cg18133284	3	63032108	-44.69904089	10.10122842	-4.425109404	3.52E-05			
9	cg27408178	4	108974899	-40.90539965	8.854027427	-4.619976613	1.73E-05	LEF1; LEF1; LEF1; LEF1	3'UTR; 3'UTR; 3'UTR; Body	
10	cg03872724	17	60881608	40.38068035	8.723718777	4.628837929	1.67E-05	MARCH10; MARCH10	5'UTR; 5'UTR	Northern shelf
11	cg18643191	17	20465437	-50.04115337	11.99579853	-4.171556671	8.66E-05			
12	cg08351753	20	43408177	-41.20634115	9.629172143	-4.279323346	5.92E-05	RIMS4; RIMS4	Body; Body	
13	cg18308985	5	5239937	51.71313803	12.69859474	4.07235124	1.22E-04	ADAMTS16	Body	
14	cg06715134	8	62357092	60.08016235	15.01863943	4.000373178	1.57E-04	CLVS1	Body	
15	cg25913233	5	151066683	39.54313266	9.468754256	4.176170549	8.52E-05	SPARC	TSS200	
16	cg07049329	19	52209605	43.56052162	10.77715067	4.041933064	1.36E-04			Southern shore
17	cg17709286	8	32852907	-39.83607065	9.737896245	-4.090829235	1.15E-04			
18	cg04191452	10	29528675	-36.50578834	8.70631771	-4.193022763	8.03E-05			
19	cg02184543	10	7226492	32.40413226	7.232208286	4.480530839	2.88E-05	SFMBT2; SFMBT2	Body; Body	
20	cg24911283	4	24780369	-37.7015075	9.225220948	-4.086786399	1.16E-04			

Supplementary Table 6. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
21	cg08706907	11	46696158	33.60668714	8.036142247	4.181942791	8.35E-05	KIAA0652; KIAA0652; KIAA0652; KIAA0652; KIAA0652	Body; 3'UTR; Body; Body; 3'UTR	
22	cg08410015	10	119862094	44.90897724	11.33739371	3.961137665	1.79E-04	CASC2; CASC2	Body; Body	
23	cg00034101	1	40098595	30.40626474	7.01380785	4.335200706	4.86E-05	HEYL	Body	
24	cg23036065	15	55972380	34.36221073	8.393401393	4.093955374	1.13E-04	PRTG	Body	
25	cg14695920	9	94997513	40.30693139	10.11286175	3.985709723	1.65E-04	IARS; IARS; IARS	Body; Body; Body	
26	cg12588208	3	119547934	28.62084821	5.96043224	4.801807496	8.81E-06	GSK3B; GSK3B	Body; Body	
27	cg16502726	16	75282608	40.85075805	10.29850026	3.966670585	1.76E-04	BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1; BCAR1	TSS1500; Body; Body; Body; Body; Body; Body; Body	Northern shelf
28	cg01420298	5	88324824	-32.09452702	7.774233964	-4.128320188	1.01E-04	MEF2C-AS1	Body	
29	cg02475777	4	1388615	45.81881161	11.71131765	3.912353246	2.11E-04	CRIPAK	1st exon	Island
30	cg19258310	11	13951237	42.16608493	10.67024838	3.951743524	1.85E-04			
31	cg26746027	4	94124976	28.36632269	5.95462526	4.763746071	1.02E-05	GRID2; GRID2	Body; Body	
32	cg12289867	10	1623710	30.56563658	7.306479245	4.183360488	8.31E-05	ADARB2	Body	
33	cg03879902	19	46171872	59.86239595	15.46346373	3.871215208	2.43E-04	GIPR	5'UTR	Northern shelf
34	cg01890546	7	884588	-30.90224935	7.488220324	-4.126781533	1.01E-04	UNC84A; UNC84A	Body; Body	Northern shelf
35	cg05693680	19	3022636	48.39310715	12.43730292	3.890964743	2.27E-04	TLE2; TLE2; TLE2; TLE2	Body; Body; Body; Body	
36	cg02031359	5	178367423	28.87092462	6.644622264	4.345006152	4.69E-05	ZNF454	TSS1500	Northern shore
37	cg01624229	2	153252785	-29.58032349	7.028077393	-4.20887845	7.60E-05	FMNL2	Body	
38	cg11750696	1	7074500	-28.09926793	6.18275001	-4.544784745	2.28E-05	CAMTA1	Body	
39	cg02600434	2	30924426	43.96823488	11.27687382	3.898973739	2.21E-04			

Supplementary Table 6. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
40	cg23417011	5	93077375	-34.12626414	8.502965648	-4.013454312	1.50E-04	FAM172A; FAM172A; POU5F2; FAM172A; FAM172A	Body; Body; TSS200; Body; Body	
41	cg21737243	7	99012691	27.25505045	5.827143062	4.677257819	1.40E-05	BUD31	Body	
42	cg06428091	16	3340853	-41.94721416	10.77411988	-3.893330928	2.26E-04	ZNF263	3'UTR	
43	cg13083079	5	55075960	28.23617338	6.580469458	4.29090562	5.69E-05	DDX4; DDX4; DDX4; DDX4	Body; Body; Body; Body	
44	cg07499372	19	57631232	-55.63457183	14.61732749	-3.806070013	3.02E-04	USP29	TSS1500	Southern shore
45	cg15924582	14	102356845	-26.59707986	5.852220919	-4.544783976	2.28E-05	PPP2R5C; PPP2R5C; PPP2R5C; PPP2R5C; PPP2R5C	Body; Body; Body; Body; Body	Northern shore
46	cg00318643	16	89105494	-29.05324115	7.038288921	-4.127884132	1.01E-04			Northern shelf
47	cg17232722	19	3178607	27.10900971	6.201796245	4.371154524	4.27E-05	S1PR4	TSS200	Northern shore
48	cg05380156	17	76141571	27.45102034	6.363901012	4.313552377	5.25E-05	C17orf99	TSS1500	
49	cg20774846	8	105479420	25.59855616	5.021375655	5.09791697	2.87E-06	DPYS	TSS200	Southern shore
50	cg04950623	22	28102227	34.1688803	8.622267932	3.962864593	1.78E-04			
51	cg07615572	17	7405889	28.99391814	7.080709016	4.094776112	1.13E-04	POLR2A	Body	
52	cg03002059	8	11402647	38.2502259	9.829393519	3.891412611	2.27E-04	BLK	Body	
53	cg14596205	11	19232715	-34.06238384	8.617291264	-3.95279477	1.84E-04	CSRP3	TSS1500	
54	cg25183883	13	70654280	31.42976161	7.891096091	3.982939917	1.66E-04	KLHL1; KLHL1	Body; Body	
55	cg04224811	3	55709442	29.07370373	7.156614987	4.06249376	1.27E-04	ERC2	3'UTR	
56	cg08904079	9	126181029	27.46584665	6.614950015	4.152086801	9.27E-05	DENND1A; DENND1A	Body; Body	
57	cg06614044	18	42324631	44.07209995	11.58015661	3.805829355	3.03E-04	SETBP1; SETBP1	Body; Body	Island
58	cg02162815	14	89305821	80.51471431	21.43933125	3.75546762	3.58E-04	TTC8; TTC8; TTC8; TTC8; TTC8; TTC8	5'UTR; 5'UTR; Body; Body; Body; Body	

Supplementary Table 6. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
59	cg10269110	19	55873515	31.062356	7.830237617	3.966974889	1.76E-04	FAM71E2	Body	Northern shore
60	cg09362918	7	147704677	31.95881912	8.0874878	3.951637382	1.85E-04	CNTNAP2; MIR548T	Body; Body	
61	cg23326197	7	99382370	-28.22927427	6.919401279	-4.079727873	1.19E-04	CYP3A4	TSS1500	
62	cg08935914	14	69387818	26.17251357	6.130816553	4.269009412	6.14E-05	ACTN1; ACTN1; ACTN1	Body; Body; Body	
63	cg09225071	17	67980864	55.74253259	14.83106253	3.758498925	3.54E-04			
64	cg05871135	3	151159930	-40.03770931	10.47061321	-3.823817049	2.85E-04	IGSF10; IGSF10; IGSF10	Body; TSS1500; TSS1500	
65	cg17250812	3	141460046	28.51927466	7.054347302	4.042794243	1.35E-04	RNF7; RNF7	Body; Body	Southern shelf
66	cg18371836	10	122553700	24.22751075	4.568081571	5.303651079	1.30E-06			
67	cg25699851	15	68624713	-24.79701928	5.705778277	-4.345948629	4.67E-05	ITGA11	Body	
68	cg02313829	11	75136574	-65.49407814	17.56923038	-3.727771607	3.92E-04	KLHL35	Body	Island
69	cg16009352	6	126912363	65.02458051	17.44575751	3.727243169	3.93E-04			
70	cg03238482	17	75418673	-27.53293062	6.82710136	-4.032887337	1.40E-04	SEPT9; SEPT9; SEPT9; SEPT9; SEPT9; SEPT9; SEPT9	Body; Body; Body; Body; Body; Body; Body	
71	cg05131483	16	23706242	-35.25830208	9.218119597	-3.824890935	2.84E-04	ERN2	Body	
72	cg26684226	22	46858421	30.28447474	7.737291781	3.914092372	2.10E-04	CELSR1	Body	Northern shore
73	cg19880831	15	33159516	28.21903344	7.083255908	3.983907091	1.66E-04	FMN1; FMN1	Body; Body	
74	cg04399083	6	82461423	40.73365998	10.82162284	3.764099025	3.48E-04	FAM46A	Body	Island
75	cg08365892	10	5693584	42.11976815	11.21931419	3.754219503	3.59E-04	ASB13; ASB13; ASB13	Body; Body; Body	
76	cg25376657	17	16229039	-55.21371718	14.84015476	-3.720562088	4.01E-04	PIGL	Body	
77	cg21248322	14	97645376	-24.22809232	5.635945149	-4.298851688	5.53E-05			

Supplementary Table 6. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
78	cg13907034	2	67487793	-29.74326805	7.608306194	-3.909315331	2.14E-04	LOC102800447	Body	
79	cg26121193	5	135328777	36.20919474	9.531532236	3.79888499	3.10E-04			
80	cg10079875	7	77304699	-23.57108862	5.264004865	-4.477786253	2.91E-05			
81	cg08542751	5	170221963	-27.97488997	7.052409287	-3.966713903	1.76E-04	GABRP	Body	
82	cg20120218	3	85071786	-34.1232843	8.965406169	-3.806105787	3.02E-04	CADM2; CADM2	Body; Body	
83	cg05702638	6	43034789	-35.68319919	9.429670289	-3.784140706	3.25E-04	KLC4; KLC4; KLC4; KLC4; KLC4; KLC4	Body; Body; Body; Body; Body; Body	
84	cg06185703	10	104874603	23.34719258	5.307936165	4.398544342	3.87E-05	NT5C2; NT5C2	Body; Body	
85	cg16992440	12	53777383	-28.5830216	7.280979014	-3.925711302	2.02E-04	SP1; SP1; SP1	Body; Body; Body	Southern shelf
86	cg22821677	15	21941178	-25.47128569	6.299856707	-4.043153182	1.35E-04	LOC646214	TSS1500	
87	cg24639246	13	29390312	-36.9606517	9.82428817	-3.762170965	3.50E-04			Northern shelf
88	cg02061596	6	168351335	-24.3252482	5.847737972	-4.159770549	9.02E-05	MLLT4; MLLT4; MLLT4	Body; Body; Body	Northern shore
89	cg27473061	2	127793088	49.13915865	13.24120014	3.711080426	4.14E-04			
90	cg11535648	11	45744794	-27.14975668	6.840459275	-3.968996173	1.74E-04	LOC100507384	Body	
91	cg09417399	11	11971576	36.12749774	9.586741905	3.768485487	3.43E-04	USP47; USP47	Body; Body	
92	cg25264393	11	134176598	-23.24809644	5.394599414	-4.309513024	5.32E-05	GLB1L3	Body	
93	cg01501775	12	54095918	-25.6172033	6.405309747	-3.999369947	1.57E-04			
94	cg14251442	5	167864068	30.63794458	8.009466804	3.825216501	2.84E-04	WWC1; WWC1; WWC1	Body; Body; Body	
95	cg16604801	17	2718310	34.57163757	9.167183109	3.771238903	3.39E-04	RAP1GAP2; RAP1GAP2	Body; Body	
96	cg13604697	18	18700867	33.94605263	8.987566151	3.7770017	3.33E-04			
97	cg08762290	8	123711460	27.45568681	6.993080368	3.926122019	2.02E-04			

Supplementary Table 6. *(Continued)*

Rank	Probe	Chromosome	Position	RE	SE	*t*	*p* value	UCSC annotation	Gene feature	CpG island feature
98	cg24155515	6	15387512	22.49029681	4.993715482	4.503720104	2.65E-05	JARID2; JARID2	Body; 5'UTR	
99	cg07622648	8	94039594	-37.9979702	10.17744972	-3.733545363	3.85E-04			
100	cg01072259	2	37946508	-44.88264092	12.11445408	-3.704883491	4.23E-04			

Displayed for each ranked probe is the chromosomal location and position (Ensembl GRCh37 assembly), the regression estimate (RE) for the Braak stage-associated analysis, the standard error (SE), the t-statistics, the accompanying p values, the Illumina gene annotation (UCSC annotation), the gene feature (TSS1500, 200 to 1500 nucleotides (nt) upstream of transcription start site (TSS); TSS200, up to 200 nt upstream of TSS; 5′UTR, 5′untranslated region; Body, gene body; 3′UTR, 3′ untranslated region) and the cytosine-phosphate-guanine (CpG) island feature. Probes are ranked based on a combined *p* value (cut-off = $p < 0.001$) and regression estimate ranking.

Supplementary Table 7. Top 50 enriched Gene Ontology (GO) terms for unmodified cytosine (5-uC) in the dorsal raphe nucleus (DRN)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Anatomical structure homeostasis	GO:0060249	Biological process	397	31	1.01E-04	5.61E-01	CCT4; USH1G; CNGB1; COL2A1; PARP1; CTSH; ADRB2; NXNL2; EPAS1; ERCC1; ERCC6; FEN1; NCS1; WWTR1; STK39; SIGLEC15; ANKRD11; INPP5D; CLDN18; GAR1; FGGY; PRKCA; MAP2K7; BBS4; SRC; SYK; TFRC; XRCC5; CACNB2; HIST3H3; LDB2
2	Tissue homeostasis	GO:0001894	Biological process	209	20	1.29E-04	5.61E-01	USH1G; CNGB1; COL2A1; CTSH; ADRB2; NXNL2; EPAS1; ERCC6; WWTR1; STK39; SIGLEC15; ANKRD11; INPP5D; CLDN18; PRKCA; BBS4; SRC; SYK; TFRC; LDB2
3	Regulation of bone remodeling	GO:0046850	Biological process	43	8	3.50E-04	5.78E-01	SIGLEC15; INPP5D; CLDN18; PRKCA; SRC; SYK; TFRC; SYT7
4	T cell receptor complex	GO:0042101	Cellular component	17	5	5.09E-04	5.78E-01	SYK; ZAP70; SKAP1; CD6; CD8A
5	Calcium ion transport	GO:0006816	Biological process	396	31	5.84E-04	5.78E-01	LILRB1; PHB2; TPCN2; NCS1; STAC3; ABL1; NALCN; ATP2C1; GRIN2C; GRM6; HRC; LETM1; TRPM1; ATP2B4; P2RX5; PDE4D; PIK3CG; PLCG2; PSEN2; CACHD1; CXCL12; TRPM2; FAM155A; CACNA1C; CACNA1E; CACNB2; CACNB4; ORAI2; CUL5; CAMK2B; CASQ1
6	Regulation of bone resorption	GO:0045124	Biological process	37	7	6.12E-04	5.78E-01	SIGLEC15; INPP5D; CLDN18; PRKCA; SRC; SYK; TFRC
7	SH2 domain binding	GO:0042169	Molecular function	32	7	6.70E-04	5.78E-01	LILRB1; ABL1; SRC; SYK; RUFY1; SKAP1; SQSTM1
8	Regulation of neutrophil activation	GO:1902563	Biological process	11	4	6.85E-04	5.78E-01	ITGAM; ITGB2; PTAFR; SYK

Supplementary Table 7. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
9	Activating transcription factor binding	GO:0033613	Molecular function	74	11	7.73E-04	5.78E-01	NEK6; ATF2; HIPK2; IFI27; SMAD2; CIITA; PRDM16; LDB2; NCOR1; ZNF516; HDAC4
10	Positive regulation of cell-cell adhesion	GO:0022409	Biological process	242	20	8.79E-04	5.78E-01	FSTL3; CXCL13; LILRB1; CYLD; CD55; EPB49; IGF2; IL15; CD46; PRKCZ; PTAFR; RPS3; SRC; SYK; TFRC; ZAP70; RUNX1; SKAP1; IL1RL2; CD6
11	Calcium ion transmembrane transport	GO:0070588	Biological process	289	25	9.24E-04	5.78E-01	PHB2; TPCN2; NCS1; STAC3; ABL1; NALCN; ATP2C1; GRIN2C; HRC; LETM1; TRPM1; ATP2B4; PDE4D; PIK3CG; PLCG2; CACHD1; TRPM2; FAM155A; CACNA1C; CACNA1E; CACNB2; CACNB4; ORAI2; CUL5; CASQ1
12	Calcium ion transmembrane transporter activity	GO:0015085	Molecular function	133	16	1.04E-03	5.78E-01	TPCN2; NCS1; ATP2C1; GRIN2C; LETM1; TRPM1; ATP2B4; CACHD1; TRPM2; FAM155A; CACNA1C; CACNA1E; CACNB2; CACNB4; ORAI2; CUL5
13	Osteoclast development	GO:0036035	Biological process	18	5	1.06E-03	5.78E-01	LILRB1; FOXP1; SIGLEC15; CLDN18; SRC
14	Bone resorption	GO:0045453	Biological process	57	8	1.25E-03	5.78E-01	ADRB2; SIGLEC15; INPP5D; CLDN18; PRKCA; SRC; SYK; TFRC
15	Divalent metal ion transport	GO:0070838	Biological process	440	32	1.27E-03	5.78E-01	SLC30A9; LILRB1; PHB2; TPCN2; NCS1; STAC3; ABL1; NALCN; ATP2C1; GRIN2C; GRM6; HRC; LETM1; TRPM1; ATP2B4; P2RX5; PDE4D; PIK3CG; PLCG2; PSEN2; CACHD1; CXCL12; TRPM2; FAM155A; CACNA1C; CACNA1E; CACNB2; CACNB4; ORAI2; CUL5; CAMK2B; CASQ1
16	Positive regulation of cell activation	GO:0050867	Biological process	301	22	1.29E-03	5.78E-01	LILRB1; CYLD; CD55; EPB49; IGF2; IL15; INPP5D; ITGAM; ITGB2; CD46; PRKCZ; PTAFR; RPS3; SRC; SYK; TFRC; THBS1; TNFRSF4; ZAP70; RUNX1; IL1RL2; CD6
17	Gap junction	GO:0005921	Cellular component	30	5	1.31E-03	5.78E-01	GJB6; PANX3; GJA5; SGSM3; PANX2
18	Divalent inorganic cation transport	GO:0072511	Biological process	443	32	1.34E-03	5.78E-01	SLC30A9; LILRB1; PHB2; TPCN2; NCS1; STAC3; ABL1; NALCN; ATP2C1; GRIN2C; GRM6; HRC; LETM1; TRPM1; ATP2B4; P2RX5; PDE4D; PIK3CG; PLCG2; PSEN2; CACHD1; CXCL12; TRPM2; FAM155A; CACNA1C; CACNA1E; CACNB2; CACNB4; ORAI2; CUL5; CAMK2B; CASQ1
19	Regulation of integrin activation	GO:0033623	Biological process	12	4	1.35E-03	5.78E-01	CDH17; CXCL13; SRC; SKAP1
20	Leukocyte differentiation	GO:0002521	Biological process	479	32	1.38E-03	5.78E-01	CDH17; FSTL3; LILRB1; PARP1; CYLD; ADGRG3; ACIN1; ABL1; FOXP1; SIGLEC15; TMEM176B; IL15; INPP5D; LFNG; CD46; CLDN18; PLCG2; POU2F2; PRKCA; PRKCZ; SRC; SYK; TFRC; TRPM2; ZAP70; ZC3H12A; LY6D; RUNX1; IL1RL2; PGLYRP1; CD8A; HDAC4

Supplementary Table 7. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
21	Positive regulation of leukocyte cell-cell adhesion	GO:1903039	Biological process	205	17	1.40E-03	5.78E-01	LILRB1; CYLD; CD55; IGF2; IL15; CD46; PRKCZ; PTAFR; RPS3; SRC; SYK; TFRC; ZAP70; RUNX1; SKAP1; IL1RL2; CD6
22	Integrin activation	GO:0033622	Biological process	21	5	1.57E-03	6.21E-01	CDH17; CXCL13; CXCL12; SRC; SKAP1
23	Positive regulation of leukocyte activation	GO:0002696	Biological process	290	21	1.83E-03	6.48E-01	LILRB1; CYLD; CD55; IGF2; IL15; INPP5D; ITGAM; ITGB2; CD46; PRKCZ; PTAFR; RPS3; SRC; SYK; TFRC; THBS1; TNFRSF4; ZAP70; RUNX1; IL1RL2; CD6
24	Positive regulation of T cell proliferation	GO:0042102	Biological process	92	9	1.94E-03	6.48E-01	CD55; IGF2; IL15; CD46; RPS3; SYK; TFRC; ZAP70; CD6
25	Regulation of leukocyte differentiation	GO:1902105	Biological process	257	20	1.96E-03	6.48E-01	FSTL3; LILRB1; CYLD; ACIN1; ABL1; FOXP1; SIGLEC15; TMEM176B; IL15; INPP5D; CD46; CLDN18; PRKCA; PRKCZ; SYK; ZAP70; ZC3H12A; RUNX1; IL1RL2; PGLYRP1
26	Detection of light stimulus	GO:0009583	Biological process	65	8	2.04E-03	6.48E-01	CNGB1; GNAT1; GRK4; GRM6; GUCY2D; EYS; ACCN1; ARRB1
27	Protein kinase binding	GO:0019901	Molecular function	616	42	2.21E-03	6.48E-01	NEK6; FRS2; NISCH; ATF2; PARP1; CYLD; DNM2; TPCN2; MAPRE3; NCS1; ABL1; NBEA; STK39; GNAT1; NR3C1; RACGAP1; ITGB2; ARRB1; ATP2B4; TPRKB; PRKAR1B; CENPJ; PRKCZ; MAP2K7; PTAFR; RPTOR; PTPRR; RPS3; SRC; SYK; TRAF3; PRKRIP1; SIKE1; CDK5RAP3; PITPNM3; FAM83A; CADPS; SKAP1; SQSTM1; CD8A; MAPKAPK2; HDAC4
28	Regulation of cytokine production	GO:0001817	Biological process	619	35	2.21E-03	6.48E-01	SPON2; IGF2BP2; LILRB1; PANX3; ATF2; CYLD; LILRA4; ABL1; FOXP1; IL15; INPP5D; LTBP1; ARRB1; CD46; PDE4D; PLCG2; RNF216; PRKCZ; PANX2; PTAFR; SIGIRR; RPS3; SRC; SYK; THBS1; TRAF3; TNFRSF4; XRCC5; ZC3H12A; RUNX1; IL1RL2; PGLYRP1; CD6; MAPKAPK2; IL27RA
29	Zinc ion binding	GO:0008270	Molecular function	764	45	2.22E-03	6.48E-01	SEC23A; PITRM1; CPXM2; PARP1; PRICKLE1; CYLD; PIKFYVE; MYT1L; TRIM35; ZZEF1; ZFR2; TRIM2; UBR4; BHMT2; ALPI; NR5A1; NR3C1; NANOS3; MDM2; MMP1; MNAT1; MTR; ZFHX3; PEX10; ZCCHC10; ZCWPW1; PRKCA; GATAD2B; RXRA; RXRB; ZSWIM4; SUPT4H1; TCEA3; THRB; TNP2; TRAF3; ZNF22; NEIL1; ADAM33; PDXK; SQSTM1; PGLYRP1; CPA5; GDA; HDAC4
30	Regulation of tissue remodeling	GO:0034103	Biological process	82	9	2.36E-03	6.48E-01	SIGLEC15; IL15; INPP5D; CLDN18; PRKCA; SRC; SYK; TFRC; SYT7

Supplementary Table 7. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
31	Lymphocyte activation involved in immune response	GO:0002285	Biological process	161	13	2.42E-03	6.48E-01	CDH17; LILRB1; ERCC1; ABL1; FOXP1; LFNG; CD46; PLCG2; PRKCZ; TFRC; ZC3H12A; PGLYRP1; IL27RA
32	Supramolecular fiber	GO:0099512	Cellular component	903	51	2.60E-03	6.48E-01	KRTAP16-1; CCT4; NEK6; KIF3A; AKAP13; IFFO2; MYOM3; COL2A1; DNAH2; CYLD; TTLL9; DNAH8; DNM2; CTTN; EPB49; MAPRE3; NINL; PALLD; TTLL5; CLASP1; KRT23; STAU2; TEKT2; IGSF22; RACGAP1; HRC; NCKAP5; KRTAP10-9; KRTAP12-4; KRTAP10-12; LMNB1; LTBP1; ATP2B4; ODF1; MNS1; CENPJ; PSEN2; MTUS1; SLAIN2; SHROOM3; SRC; CACNA1C; CDK5RAP3; DNAL1; OBSCN; CASQ1; GAS7; SKAP1; SQSTM1; NEXN; HDAC4
33	Cytokine production	GO:0001816	Biological process	687	38	2.61E-03	6.48E-01	SPON2; IGF2BP2; LILRB1; PANX3; ATF2; CYLD; CD55; PLA2R1; LILRA4; ABL1; FOXP1; IL15; INPP5D; LTBP1; ARRB1; CD46; PDE4D; PIK3CG; PLCG2; RNF216; PRKCZ; PANX2; PTAFR; SIGIRR; RPS3; SRC; SYK; THBS1; TRAF3; TNFRSF4; XRCC5; ZC3H12A; RUNX1; IL1RL2; PGLYRP1; CD6; MAPKAPK2; IL27RA
34	Positive regulation of cell adhesion	GO:0045785	Biological process	384	28	2.71E-03	6.48E-01	FSTL3; AGR2; CXCL13; LILRB1; TRIOBP; CYLD; CD55; DNM2; EPB49; ABL1; IGF2; IL15; CD46; ZFHX3; PRKCA; PRKCZ; PTAFR; RPS3; RREB1; CXCL12; SRC; SYK; TFRC; ZAP70; RUNX1; SKAP1; IL1RL2; CD6
35	Adaptive immune response	GO:0002250	Biological process	368	22	2.73E-03	6.48E-01	CXCL13; LILRB1; CTSH; CD55; ERCC1; LAMP3; INPP5D; CD46; PIK3CG; POU2F2; PRKCZ; SYK; TAP1; TFRC; ZAP70; ZC3H12A; STX7; SKAP1; CD6; CD7; CD8A; IL27RA
36	Detection of light stimulus involved in visual perception	GO:0050908	Biological process	15	4	2.84E-03	6.48E-01	CNGB1; GNAT1; GRM6; EYS
37	Detection of light stimulus involved in sensory perception	GO:0050962	Biological process	15	4	2.84E-03	6.48E-01	CNGB1; GNAT1; GRM6; EYS
38	Supramolecular polymer	GO:0099081	Cellular component	908	51	3.10E-03	6.48E-01	KRTAP16-1; CCT4; NEK6; KIF3A; AKAP13; IFFO2; MYOM3; COL2A1; DNAH2; CYLD; TTLL9; DNAH8; DNM2; CTTN; EPB49; MAPRE3; NINL; PALLD; TTLL5; CLASP1; KRT23; STAU2; TEKT2; IGSF22; RACGAP1; HRC; NCKAP5; KRTAP10-9; KRTAP12-4; KRTAP10-12; LMNB1; LTBP1; ATP2B4; ODF1; MNS1; CENPJ; PSEN2; MTUS1; SLAIN2; SHROOM3; SRC; CACNA1C; CDK5RAP3; DNAL1; OBSCN; CASQ1; GAS7; SKAP1; SQSTM1; NEXN; HDAC4

Supplementary Table 7. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
39	Transcriptional repressor complex	GO:0017053	Cellular component	78	10	3.11E-03	6.48E-01	CHD5; LIN9; DEPDC1; RCOR3; GATAD2B; PHF12; PRDM16; TBX15; NCOR1; HDAC4
40	Lymphocyte activation	GO:0046649	Biological process	616	36	3.18E-03	6.48E-01	CDH17; LILRB1; CYLD; CD55; DDOST; ERCC1; ADGRG3; ABL1; FOXP1; IGF2; IL15; INPP5D; ITGB2; LFNG; CD46; IL21R; PIK3CG; PLCG2; POU2F2; PRKCZ; RPS3; SRC; SYK; TFRC; TNFRSF4; ZAP70; ZC3H12A; LY6D; RUNX1; IL1RL2; PGLYRP1; CD6; CD7; CD8A; IL27RA; HDAC4
41	Supramolecular complex	GO:0099080	Cellular component	909	51	3.19E-03	6.48E-01	KRTAP16-1; CCT4; NEK6; KIF3A; AKAP13; IFFO2; MYOM3; COL2A1; DNAH2; CYLD; TTLL9; DNAH8; DNM2; CTTN; EPB49; MAPRE3; NINL; PALLD; TTLL5; CLASP1; KRT23; STAU2; TEKT2; IGSF22; RACGAP1; HRC; NCKAP5; KRTAP10-9; KRTAP12-4; KRTAP10-12; LMNB1; LTBP1; ATP2B4; ODF1; MNS1; CENPJ; PSEN2; MTUS1; SLAIN2; SHROOM3; SRC; CACNA1C; CDK5RAP3; DNAL1; OBSCN; CASQ1; GAS7; SKAP1; SQSTM1; NEXN; HDAC4
42	Desensitization of g protein-coupled receptor signaling pathway	GO:0002029	Biological process	16	4	3.22E-03	6.48E-01	ADRB2; DNM2; GRK4; ARRB1
43	Negative adaptation of signaling pathway	GO:0022401	Biological process	16	4	3.22E-03	6.48E-01	ADRB2; DNM2; GRK4; ARRB1
44	G protein-coupled receptor signaling pathway involved in heart process	GO:0086103	Biological process	16	4	3.28E-03	6.48E-01	AKAP13; ATP2B4; PDE4D; SRC
45	Microtubule plus-end	GO:0035371	Cellular component	20	5	3.67E-03	7.08E-01	MAPRE3; CLASP1; NCKAP5; SLAIN2; SKAP1
46	Adaptation of signaling pathway	GO:0023058	Biological process	17	4	3.80E-03	7.19E-01	ADRB2; DNM2; GRK4; ARRB1
47	Kinase binding	GO:0019900	Molecular function	693	45	3.92E-03	7.26E-01	NEK6; FRS2; NISCH; AKAP13; ATF2; PARP1; CYLD; DNM2; TPCN2; MAPRE3; NCS1; ABL1; NBEA; STK39; GNAT1; NR3C1; RACGAP1; ITGB2; ARRB1; ATP2B4; TPRKB; PFKL; TOLLIP; PRKAR1B; CENPJ; PRKCZ; MAP2K7; PTAFR; RPTOR; PTPRR; RPS3; SRC; SYK; TRAF3; PRKRIP1; SIKE1; CDK5RAP3; PITPNM3; FAM83A; CADPS; SKAP1; SQSTM1; CD8A; MAPKAPK2; HDAC4

Supplementary Table 7. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
48	T cell proliferation	GO:0042098	Biological process	170	13	4.09E-03	7.29E-01	LILRB1; CD55; ABL1; IGF2; IL15; CD46; PIK3CG; RPS3; SYK; TFRC; TNFRSF4; ZAP70; CD6
49	Calcium ion transport into cytosol	GO:0060402	Biological process	145	14	4.11E-03	7.29E-01	TPCN2; ABL1; GRIN2C; HRC; LETM1; TRPM1; ATP2B4; P2RX5; PDE4D; PLCG2; TRPM2; FAM155A; CACNA1C; CASQ1
50	Bone remodeling	GO:0046849	Biological process	82	9	4.28E-03	7.34E-01	ADRB2; SIGLEC15; INPP5D; CLDN18; PRKCA; SRC; SYK; TFRC; SYT7

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value.

Supplementary Table 8. Top 50 enriched Gene Ontology (GO) terms for 5-methylcytosine (5-mC) in the dorsal raphe nucleus (DRN)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Protein tyrosine kinase activator activity	GO:0030296	Molecular function	16	4	1.85E-03	1.00	NRG3; EFNA5; NRG1; AFAP1L2
2	Regulation of translational initiation	GO:0006446	Biological process	71	7	2.00E-03	1.00	EIF4EBP2; EIF2AK1; EIF2C2; NCBP1; NPM1; RPS6KB1; EIF2B2
3	Regulation of DNA metabolic process	GO:0051052	Biological process	389	20	2.80E-03	1.00	EHMT2; CDAN1; C20orf196; DNMT1; ERCC4; EYA2; WAPAL; SMG6; ANKRD1; ENPP7; MGMT; NPAS2; NPM1; SIRT6; XRN1; PPP2R1A; PPP4C; RFC2; TNFAIP1; SPIRE2
4	Sumo transferase activity	GO:0019789	Molecular function	23	4	3.20E-03	1.00	NSMCE2; MDM2; TRIM27; HDAC4
5	Inositol phosphate-mediated signaling	GO:0048016	Biological process	56	7	3.25E-03	1.00	CAMTA1; NRG1; HRH1; INPP5A; ITPR1; NFATC3; DYRK2
6	Receptor activator activity	GO:0030546	Molecular function	10	3	6.05E-03	1.00	NRG3; EFNA5; NRG1
7	Histone demethylase activity (H3-K9 specific)	GO:0032454	Molecular function	12	3	6.08E-03	1.00	KDM4B; KDM3B; HR
8	Photoreceptor outer segment membrane	GO:0042622	Cellular component	14	3	6.53E-03	1.00	NAPEPLD; GNAT1; DHRS3
9	Histone H3-K9 demethylation	GO:0033169	Biological process	13	3	7.39E-03	1.00	KDM4B; KDM3B; HR
10	Negative regulation of translational initiation	GO:0045947	Biological process	17	3	7.40E-03	1.00	EIF4EBP2; EIF2AK1; EIF2C2

Supplementary Table 8. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
11	Positive regulation of myelination	GO:0031643	Biological process	13	3	8.72E-03	1.00	WASF3; S100B; C11orf9
12	Synapse	GO:0045202	Cellular component	1102	51	1.01E-02	1.00	NRG3; WASF3; LZTS1; CHRM2; CTNND2; DPYSL3; EFNA5; EIF4EBP2; SORCS3; ERC1; KIAA1107; PIP5K1C; SCRIB; MAPK8IP2; GABBR1; GABRG3; DNM3; RGS17; CYFIP2; SYPL2; NRG1; ITPR1; KCND2; GRID2IP; ACCN1; LRP4; MDM2; MKLN1; ATP2A2; NTRK3; RAB6B; PLD1; XRN1; PPP2R1A; PPP1R9A; ANKS1B; PTK2; RHEB; RPS6KB1; RPS21; SNTB1; CAD; PRRT1; MAP1LC3A; STON2; SYNGAP1; DLGAP1; AKAP7; NOS1AP; HDAC4; MAGI2
13	Myelination	GO:0042552	Biological process	111	9	1.01E-02	1.00	WASF3; NRG1; NTRK3; S100B; SCN8A; TG; C11orf9; EIF2B2; ARHGEF10
14	Ensheathment of neurons	GO:0007272	Biological process	113	9	1.12E-02	1.00	WASF3; NRG1; NTRK3; S100B; SCN8A; TG; C11orf9; EIF2B2; ARHGEF10
15	Axon ensheathment	GO:0008366	Biological process	113	9	1.12E-02	1.00	WASF3; NRG1; NTRK3; S100B; SCN8A; TG; C11orf9; EIF2B2; ARHGEF10
16	Activation of transmembrane receptor protein tyrosine kinase activity	GO:0007171	Biological process	13	3	1.13E-02	1.00	NRG3; EFNA5; NRG1
17	Regulation of DNA replication	GO:0006275	Biological process	103	7	1.44E-02	1.00	EHMT2; CDAN1; WAPAL; ENPP7; PPP2R1A; RFC2; TNFAIP1
18	Primary alcohol metabolic process	GO:0034308	Biological process	73	5	1.50E-02	1.00	AKR1C4; NAPEPLD; ALDH1A2; CACNA1H; DHRS3
19	Regulation of blood circulation	GO:1903522	Biological process	250	14	1.51E-02	1.00	CHRM2; SMTNL1; HRH1; ITPR1; KCNJ12; ACCN1; MDM2; ATP2A2; SPTBN4; SCN4B; TNNT2; CACNA1H; NOS1AP; HDAC4
20	Odorant binding	GO:0005549	Molecular function	82	3	1.52E-02	1.00	OR5M11; OR8K1; OR8U8
21	Ion channel inhibitor activity	GO:0008200	Molecular function	36	4	1.55E-02	1.00	NEDD4L; KCNV1; ITPR1; STX8
22	Regulation of ossification	GO:0030278	Biological process	185	11	1.61E-02	1.00	NPNT; LRP4; DDR2; ATRAID; ANKH; FAM20C; PTK2; ZBTB16; LIMD1; DHRS3; HDAC4
23	Action potential	GO:0001508	Biological process	127	9	1.65E-02	1.00	NEDD4L; KCND2; ATP2A2; NTRK3; SCN4B; SCN8A; CACNB3; CACNA1H; NOS1AP
24	Oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	GO:0016655	Molecular function	52	4	1.67E-02	1.00	AKR1C4; NDUFA5; NDUFAB1; CBR1
25	Organellar large ribosomal subunit	GO:0000315	Cellular component	57	4	1.79E-02	1.00	MRPL22; NDUFAB1; MRPL30; MRPL39

Supplementary Table 8. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
26	Mitochondrial large ribosomal subunit	GO:0005762	Cellular component	57	4	1.79E-02	1.00	MRPL22; NDUFAB1; MRPL30; MRPL39
27	Channel inhibitor activity	GO:0016248	Molecular function	37	4	1.89E-02	1.00	NEDD4L; KCNV1; ITPR1; STX8
28	Nucleoside monophosphate biosynthetic process	GO:0009124	Biological process	153	9	1.94E-02	1.00	DUT; AK1; AKD1; ENTPD8; SIRT6; ATP5G1; PFKP; CAD; HDAC4
29	Plasma membrane raft	GO:0044853	Cellular component	101	8	1.94E-02	1.00	EFNA5; MYOF; ITGAM; KCND2; LRP4; MYO1D; CD177; NOS1AP
30	Adherens junction assembly	GO:0034333	Biological process	86	8	1.95E-02	1.00	EFNA5; LIMCH1; PIP5K1C; BCAS3; PTK2; TBCD; TRIP6; PHLDB2
31	Transcription elongation from RNA polymerase ii promoter	GO:0006368	Biological process	83	6	1.96E-02	1.00	TCEB3C; MLLT1; NCBP1; POLR2C; ELP2; TCEB1
32	Synaptic vesicle endocytosis	GO:0048488	Biological process	48	5	2.02E-02	1.00	KIAA1107; PIP5K1C; SCRIB; DNM3; STON2
33	Presynaptic endocytosis	GO:0140238	Biological process	48	5	2.02E-02	1.00	KIAA1107; PIP5K1C; SCRIB; DNM3; STON2
34	Heart contraction	GO:0060047	Biological process	233	13	2.09E-02	1.00	CHRM2; NEDD4L; ITPR1; KCNJ12; MDM2; ATP2A2; SPTBN4; SCN4B; SGCD; TNNT2; CACNA1H; NOS1AP; HDAC4
35	Regulation of membrane repolarization	GO:0060306	Biological process	35	4	2.15E-02	1.00	NEDD4L; SCN4B; CACNB3; NOS1AP
36	Calcineurin-mediated signaling	GO:0097720	Biological process	46	5	2.19E-02	1.00	CAMTA1; NR5A2; NRG1; NFATC3; DYRK2
37	Tertiary granule membrane	GO:0070821	Cellular component	69	5	2.22E-02	1.00	MCEMP1; ITGAM; PLD1; CD177; CD59
38	Response to starvation	GO:0042594	Biological process	176	10	2.30E-02	1.00	EHMT2; EIF2AK1; FOXA3; ATF3; BCAS3; RPTOR; SSTR3; CAD; MAP1LC3A; MTMR3
39	Negative regulation of DNA metabolic process	GO:0051053	Biological process	139	8	2.30E-02	1.00	CDAN1; C20orf196; ERCC4; WAPAL; SMG6; ANKRD1; ENPP7; XRN1
40	Tertiary granule	GO:0070820	Cellular component	158	8	2.42E-02	1.00	MCEMP1; ITGAM; PLD1; CD177; STXBP2; TIMP2; DOK3; CD59
41	Epithelial cell morphogenesis	GO:0003382	Biological process	30	4	2.44E-02	1.00	FLNB; NOTCH4; BCL11B; COL23A1
42	Regulation of cell shape	GO:0008360	Biological process	145	10	2.45E-02	1.00	WASF3; ARHGEF18; GNA12; MKLN1; PLXNA1; PLXNA2; PTK2; SHROOM3; S100B; LIMD1

Supplementary Table 8. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
43	Positive regulation of plasma membrane bounded cell projection assembly	GO:0120034	Biological process	95	7	2.61E-02	1.00	DPYSL3; COBL; DNM3; AUTS2; BCAS3; MARK4; HDAC4
44	Dendrite extension	GO:0097484	Biological process	31	4	2.63E-02	1.00	NEDD4L; AUTS2; CYFIP2; CPNE6
45	5'-nucleotidase activity	GO:0008253	Molecular function	12	2	2.66E-02	1.00	NT5C3L; NT5DC1
46	Positive regulation of DNA repair	GO:0045739	Biological process	57	5	2.67E-02	1.00	C20orf196; EYA2; MGMT; NPAS2; SPIRE2
47	Regulation of myelination	GO:0031641	Biological process	34	4	2.78E-02	1.00	WASF3; S100B; TG; C11orf9
48	Modulation by virus of host process	GO:0019054	Biological process	22	3	2.90E-02	1.00	SCRIB; NTRK3; VAPA
49	Surfactant homeostasis	GO:0043129	Biological process	11	2	2.94E-02	1.00	SFTPA1; LPCAT1
50	Heart process	GO:0003015	Biological process	241	13	3.01E-02	1.00	CHRM2; NEDD4L; ITPR1; KCNJ12; MDM2; ATP2A2; SPTBN4; SCN4B; SGCD; TNNT2; CACNA1H; NOS1AP; HDAC4

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value.

Supplementary Table 9. Top 50 enriched Gene Ontology (GO) terms for 5-hydroxymethylcytosine (5-hmC) in the dorsal raphe nucleus (DRN)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Homophilic cell adhesion via plasma membrane adhesion molecules	GO:0007156	Biological process	167	27	1.31E-07	1.16E-03	CDH7; CDH15; FAT1; PALLD; DCHS2; PCDHGB7; PCDHGB6; PCDHGB5; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA11; PCDHGA10; PCDHGA9; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; DSCAML1; PVRL4; PCDHGB4; CELSR1; PCDHGA8; CLSTN3

Supplementary Table 9. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
2	Calcium ion binding	GO:0005509	Molecular function	690	56	2.47E-06	1.09E-02	CDH7; CDH15; CGREF1; FBLN7; NCAN; DGKB; FBLN1; FAT1; KCNIP3; EHD4; ANXA11; HPCAL1; ITPR1; AGRN; LRP1; LTBP3; PLA2G2A; DCHS2; PRF1; STAB2; PCDHGB7; PCDHGB6; PCDHGB5; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA11; PCDHGA10; PCDHGA9; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PLSCR3; MAN1C1; HEG1; SYT13; EPS15L1; RYR1; SLIT3; DST; SPARC; TPD52; COLEC11; CUBN; ACTN4; DYSF; SYT3; PCDHGB4; CELSR1; PCDHGA8; RAB11FIP3; CLSTN3
3	Cell-cell adhesion via plasma-membrane adhesion molecules	GO:0098742	Biological process	270	28	7.01E-05	2.07E-01	CDH7; CDH15; FAT1; PALLD; CLDN18; DCHS2; PCDHGB7; PCDHGB6; PCDHGB5; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA11; PCDHGA10; PCDHGA9; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; DSCAML1; PVRL4; PCDHGB4; CELSR1; PCDHGA8; CLSTN3
4	Xenobiotic glucuronidation	GO:0052697	Biological process	11	5	2.62E-04	5.62E-01	UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
5	Cell volume homeostasis	GO:0006884	Biological process	28	7	3.17E-04	5.62E-01	ANO6; LRRC8B; KCNMA1; LRRC8D; SLC12A9; SLC12A4; SLC12A8
6	Retinoid metabolic process	GO:0001523	Biological process	86	10	4.53E-04	6.68E-01	NAPEPLD; AGRN; LRP1; UGT1A8; UGT1A7; UGT1A9; SDC1; STRA6; ALDH1A2; DHRS3
7	Animal organ maturation	GO:0048799	Biological process	23	6	5.31E-04	6.72E-01	ANO6; IGF1; RHOA; RYR1; XYLT1; ALDH1A2
8	Ossification involved in bone maturation	GO:0043931	Biological process	17	5	7.90E-04	7.50E-01	ANO6; IGF1; RHOA; RYR1; XYLT1
9	Diterpenoid metabolic process	GO:0016101	Biological process	92	10	8.22E-04	7.50E-01	NAPEPLD; AGRN; LRP1; UGT1A8; UGT1A7; UGT1A9; SDC1; STRA6; ALDH1A2; DHRS3
10	UDP-glycosyltransferase activity	GO:0008194	Molecular function	145	14	8.62E-04	7.50E-01	MGAT4B; MGAT4A; MGAT5B; C3orf21; ALG14; EXT1; GALNT2; B4GALNT3; UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9; XYLT1
11	Retinoid binding	GO:0005501	Molecular function	35	6	1.06E-03	7.50E-01	IGF2R; UGT1A8; UGT1A7; UGT1A9; STRA6; ALDH1A2
12	Cellular glucuronidation	GO:0052695	Biological process	19	5	1.06E-03	7.50E-01	UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
13	Flavonoid metabolic process	GO:0009812	Biological process	15	5	1.10E-03	7.50E-01	UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
14	Bone maturation	GO:0070977	Biological process	19	5	1.26E-03	7.88E-01	ANO6; IGF1; RHOA; RYR1; XYLT1
15	Isoprenoid binding	GO:0019840	Molecular function	37	6	1.42E-03	7.88E-01	IGF2R; UGT1A8; UGT1A7; UGT1A9; STRA6; ALDH1A2
16	Terpenoid metabolic process	GO:0006721	Biological process	104	10	1.42E-03	7.88E-01	NAPEPLD; AGRN; LRP1; UGT1A8; UGT1A7; UGT1A9; SDC1; STRA6; ALDH1A2; DHRS3

Supplementary Table 9. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
17	GTPase activator activity	GO:0005096	Molecular function	273	24	1.53E-03	7.96E-01	FAM13A; ADAP1; AGAP2; AGAP1; DAB2IP; DOCK2; PGAM5; ARHGAP27; RASA3; ARHGEF12; ACAP2; TBC1D22A; ABR; GRLF1; AGFG2; C20orf95; ASAP1; RIN2; ARHGAP15; RASA1; SMAP2; SYNGAP1; TRIP10; SRGAP3
18	Glucuronosyltransferase activity	GO:0015020	Molecular function	32	6	1.75E-03	8.61E-01	EXT1; UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
19	Uronic acid metabolic process	GO:0006063	Biological process	24	5	2.59E-03	1.00	UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
20	Glucuronate metabolic process	GO:0019585	Biological process	24	5	2.59E-03	1.00	UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9
21	Regulation of catecholamine secretion	GO:0050433	Biological process	61	7	2.65E-03	1.00	ADRA2A; SYT13; CXCL12; P2RY12; SNCG; VIP; SYT3
22	Cardiac conduction	GO:0061337	Biological process	110	12	2.76E-03	1.00	GJC1; CTNNA3; ITPR1; KCNK1; KCNQ1; MIR328; ATP2B2; ATP2B4; RYR1; TRPC1; CACNA1C; CACNA2D1
23	Positive regulation of mitochondrial translation	GO:0070131	Biological process	17	3	2.83E-03	1.00	MRPS27; CCDC56; NSUN4
24	Primary alcohol metabolic process	GO:0034308	Biological process	76	7	3.11E-03	1.00	ALDH2; NAPEPLD; DKK3; IGF1; ALDH1A2; CACNA1H; DHRS3
25	Isoprenoid metabolic process	GO:0006720	Biological process	123	10	3.55E-03	1.00	NAPEPLD; AGRN; LRP1; UGT1A8; UGT1A7; UGT1A9; SDC1; STRA6; ALDH1A2; DHRS3
26	Catecholamine secretion	GO:0050432	Biological process	63	7	3.81E-03	1.00	ADRA2A; SYT13; CXCL12; P2RY12; SNCG; VIP; SYT3
27	SH2 domain binding	GO:0042169	Molecular function	33	6	3.96E-03	1.00	KHDRBS2; SH3PXD2B; LAX1; SRC; AFAP1L2; SQSTM1
28	Gtpase regulator activity	GO:0030695	Molecular function	303	24	4.73E-03	1.00	FAM13A; ADAP1; AGAP2; AGAP1; DAB2IP; DOCK2; PGAM5; ARHGAP27; RASA3; ARHGEF12; ACAP2; TBC1D22A; ABR; GRLF1; AGFG2; C20orf95; ASAP1; RIN2; ARHGAP15; RASA1; SMAP2; SYNGAP1; TRIP10; SRGAP3
29	Calcium ion import across plasma membrane	GO:0098703	Biological process	13	4	4.82E-03	1.00	ATP2B4; TRPV6; FAM155A; CACNA2D1
30	Calcium ion import into cytosol	GO:1902656	Biological process	13	4	4.82E-03	1.00	ATP2B4; TRPV6; FAM155A; CACNA2D1
31	Positive regulation of epidermal growth factor receptor signaling pathway	GO:0045742	Biological process	31	5	4.92E-03	1.00	ADRA2A; DOK1; HIP1; PDE6G; AFAP1L2

Supplementary Table 9. *(Continued)*								
Rank	**GO term**	**ID**	**Ontology**	**Number of genes**	**Differentially modified**	***p* value**	**FDR adjusted *p***	**Genes**
32	Cell-cell adhesion	GO:0098609	Biological process	810	49	4.94E-03	1.00	DNAJB6; CDH7; CDH15; FAT1; PALLD; CTNNA3; HLA-DMB; IGF1; IDO1; ITPKB; RHOA; NPHP1; CLDN18; FAM49B; DCHS2; LAX1; SIRPG; PRKAR1A; PCDHGB7; PCDHGB6; PCDHGB5; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA11; PCDHGA10; PCDHGA9; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; DSCAML1; IL21; CXCL12; P2RY12; SRC; TNR; TSTA3; WNT1; PDCD1LG2; PVRL4; PCDHGB4; CELSR1; CD58; PCDHGA8; CLSTN3
33	Nucleoside-triphosphatase regulator activity	GO:0060589	Molecular function	347	25	6.31E-03	1.00	DNAJB6; FAM13A; ADAP1; AGAP2; AGAP1; DAB2IP; DOCK2; PGAM5; ARHGAP27; RASA3; ARHGEF12; ACAP2; TBC1D22A; ABR; GRLF1; AGFG2; C20orf95; ASAP1; RIN2; ARHGAP15; RASA1; SMAP2; SYNGAP1; TRIP10; SRGAP3
34	Positive regulation of ERBB signaling pathway	GO:1901186	Biological process	33	5	7.40E-03	1.00	ADRA2A; DOK1; HIP1; PDE6G; AFAP1L2
35	Positive regulation of GTPase activity	GO:0043547	Biological process	399	28	7.52E-03	1.00	FAM13A; ADAP1; AGAP2; AGAP1; DAB2IP; DOCK2; PGAM5; ARHGAP27; RASA3; ARHGEF12; ACAP2; TBC1D22A; ABR; GRLF1; AGFG2; C20orf95; AGRN; ASAP1; RIN2; BCAS3; ARHGAP15; GRHL3; RASA1; SMAP2; SYNGAP1; TRIP10; ARHGEF10; SRGAP3
36	Inorganic anion transmembrane transporter activity	GO:0015103	Molecular function	157	13	7.62E-03	1.00	ABCC5; SLC26A7; ABCC2; ANO6; LRRC8B; SLC26A4; LRRC8D; SLC12A9; ANO2; SLC12A4; BEST1; SLC12A8; ABCC11
37	Transferase activity, transferring glycosyl groups	GO:0016757	Molecular function	270	17	8.18E-03	1.00	MGAT4B; MGAT4A; MGAT5B; C3orf21; PARP15; ALG14; EXT1; GALNT2; B4GALNT3; ART3; UGT1A10; UGT1A8; UGT1A7; UGT1A6; UGT1A9; XYLT1; TNKS
38	Retinoic acid binding	GO:0001972	Molecular function	19	4	8.24E-03	1.00	IGF2R; UGT1A8; UGT1A7; UGT1A9
39	Retinoic acid metabolic process	GO:0042573	Biological process	22	4	8.30E-03	1.00	UGT1A8; UGT1A7; UGT1A9; ALDH1A2
40	Negative regulation of cell-substrate adhesion	GO:0010812	Biological process	66	8	8.60E-03	1.00	FBLN1; RHOA; BCAS3; AJAP1; RASA1; SRC; WNT1; ACTN4
41	Negative regulation of alcohol biosynthetic process	GO:1902931	Biological process	17	3	8.62E-03	1.00	SCAP; DKK3; NFKB1
42	Protein ADP-ribosylation	GO:0006471	Biological process	31	4	8.63E-03	1.00	PARP15; ART3; XRCC1; TNKS
43	Acetylglucosaminyltransferase activity	GO:0008375	Molecular function	52	6	9.05E-03	1.00	MGAT4B; MGAT4A; MGAT5B; ALG14; EXT1; XYLT1

Supplementary Table 9. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
44	Enzyme activator activity	GO:0008047	Molecular function	520	31	9.22E-03	1.00	DNAJB6; FAM13A; ADAP1; AGAP2; AGAP1; DAB2IP; DOCK2; PGAM5; ARHGAP27; FBLN1; RASA3; ARHGEF12; ACAP2; TBC1D22A; SH3PXD2B; ABR; GRLF1; GUCA2A; AGFG2; C20orf95; ASAP1; RIN2; BCAS3; ARHGAP15; GPRC5C; RASA1; SMAP2; AFAP1L2; SYNGAP1; TRIP10; SRGAP3
45	Negative regulation of steroid metabolic process	GO:0045939	Biological process	30	4	9.28E-03	1.00	SCAP; DKK3; NFKB1; UGT1A8
46	Bleb assembly	GO:0032060	Biological process	11	3	9.29E-03	1.00	ANO6; EMP2; MYLK
47	Positive regulation of podosome assembly	GO:0071803	Biological process	11	3	9.55E-03	1.00	RHOA; ASAP1; SRC
48	Positive regulation of voltage-gated calcium channel activity	GO:1901387	Biological process	12	3	9.83E-03	1.00	STAC3; STAC2; CACNA2D1
49	Podosome assembly	GO:0071800	Biological process	19	4	9.86E-03	1.00	SH3PXD2B; RHOA; ASAP1; SRC
50	Transmembrane transporter activity	GO:0022857	Molecular function	1044	53	9.98E-03	1.00	GJC1; ABCC5; SLC26A7; SLC5A11; ABCC2; SLC7A13; ANO6; SLC29A1; RASA3; SLC35A3; SLC44A1; ABCB9; LRRC8B; NNT; GRIK1; KCNIP3; ITPR1; KCNK1; KCNMA1; KCNQ1; ATP2B2; ATP2B4; SLC45A1; ATP5G2; SLC26A4; LRRC8D; SLC38A7; TRPV6; PRF1; SLC39A4; PANX2; SLC12A9; ANO2; SLC44A2; JPH3; RYR1; STRA6; SLC12A4; SLC14A1; SLC22A2; SLC22A5; TRPC1; FAM155A; BEST1; CACNA1C; CACNA2D1; SLC2A10; SLC14A2; SLC12A8; ABCC11; CACNA1H; SLC28A1; CACNA2D4

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value and those that are significantly enriched after multiple testing correction (FDR adjusted $p < 0.05$) are seperated from the other terms by a dashed line.

Supplementary Table 10. Top 50 enriched Gene Ontology (GO) terms for unmodified cytosine (5-uC) in the locus coeruleus (LC)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Circadian sleep/wake cycle process	GO:0022410	Biological process	23	5	9.36E-05	4.84E-01	BTBD9; ADORA1; NLGN1; HCRTR2; NPS
2	Circadian sleep/wake cycle	GO:0042745	Biological process	25	5	1.50E-04	4.84E-01	BTBD9; ADORA1; NLGN1; HCRTR2; NPS

Supplementary Table 10. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
3	Synaptic transmission, glutamatergic	GO:0035249	Biological process	88	11	1.67E-04	4.84E-01	ADORA1; NLGN1; GRID1; GRIN1; GRM3; GRM4; PARK2; NPS; DGKI; NRXN1; CLSTN3
4	Regulation of circadian sleep/wake cycle	GO:0042749	Biological process	20	4	5.60E-04	1.00	ADORA1; NLGN1; HCRTR2; NPS
5	Synaptic transmission, GABAergic	GO:0051932	Biological process	43	6	1.20E-03	1.00	ADORA1; NLGN1; GABRA1; NPS; BAIAP3; CLSTN3
6	Pericentriolar material	GO:0000242	Cellular component	21	4	1.80E-03	1.00	CEP152; LCK; NEK1; TNKS
7	Phosphatidylinositol 3-kinase binding	GO:0043548	Molecular function	30	5	2.15E-03	1.00	DAB1; RASD2; LCK; AXL; CBL
8	Histone lysine methylation	GO:0034968	Biological process	104	9	2.41E-03	1.00	KDM4C; AUTS2; WHSC1L1; KDM4D; PRDM16; RDBP; SETDB2; SETD3; EED
9	Cell-cell adhesion via plasma-membrane adhesion molecules	GO:0098742	Biological process	259	17	3.13E-03	1.00	DAB1; NLGN1; PALLD; ODZ4; PCDH7; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; PTPRS; NRXN1; CLSTN3; PCDHA9
10	Circadian behavior	GO:0048512	Biological process	41	5	3.52E-03	1.00	BTBD9; ADORA1; NLGN1; HCRTR2; NPS
11	Rhythmic behavior	GO:0007622	Biological process	42	5	3.95E-03	1.00	BTBD9; ADORA1; NLGN1; HCRTR2; NPS
12	Transmembrane receptor protein serine/threonine kinase binding	GO:0070696	Molecular function	14	3	4.06E-03	1.00	SMAD6; BMP7; MAGI2
13	Postsynaptic specialization	GO:0099572	Cellular component	324	20	4.08E-03	1.00	ADORA1; CTNND2; DAB1; NLGN1; GABRA1; DISC1; GRID1; GRIN1; GRM3; ITPR1; ANKS1B; PTPRS; RAPSN; SPTBN1; PKP4; MPDZ; LIN7A; DGKI; CLSTN3; MAGI2
14	Calcium ion binding	GO:0005509	Molecular function	672	31	4.30E-03	1.00	EDIL3; MICU1; DGKG; ZZEF1; PLCB1; CRB2; GRIN1; EFEMP2; ITPR1; MMP13; ASPH; PCDH7; LRP1B; STAB2; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; PLSCR2; TRPM2; CBL; DLK1; BAIAP3; NRXN1; PLCH2; CLSTN3; PCDHA9
15	Glomerulus vasculature development	GO:0072012	Biological process	22	4	4.42E-03	1.00	HES1; BMP7; PDGFD; CFLAR
16	Peptidyl-lysine methylation	GO:0018022	Biological process	118	9	4.53E-03	1.00	KDM4C; AUTS2; WHSC1L1; KDM4D; PRDM16; RDBP; SETDB2; SETD3; EED
17	Regulation of histone methylation	GO:0031060	Biological process	59	6	4.69E-03	1.00	KDM4C; AUTS2; WHSC1L1; KDM4D; RDBP; EED
18	Transition metal ion homeostasis	GO:0055076	Biological process	127	8	4.95E-03	1.00	BTBD9; LCK; SLC11A2; PARK2; CUTC; SLC30A5; SFXN5; ABCG2

Supplementary Table 10. *(Continued)*

GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes	
19	Regulation of glomerulus development	GO:0090192	Biological process	12	3	5.55E-03	1.00	BMP7; PDGFD; CFLAR
20	Cellular response to oxidative stress	GO:0034599	Biological process	276	14	5.67E-03	1.00	PDCD10; EIF2S1; ANKRD2; ARNT; PARK2; AXL; SELS; BMP7; STAT6; TPM1; TRPM2; VRK2; PDGFD; CFLAR
21	Regulation of synaptic transmission, glutamatergic	GO:0051966	Biological process	65	7	6.05E-03	1.00	ADORA1; NLGN1; GRM3; GRM4; NPS; DGKI; NRXN1
22	Renal system vasculature development	GO:0061437	Biological process	24	4	6.27E-03	1.00	HES1; BMP7; PDGFD; CFLAR
23	Kidney vasculature development	GO:0061440	Biological process	24	4	6.27E-03	1.00	HES1; BMP7; PDGFD; CFLAR
24	Regulation of cell proliferation involved in kidney development	GO:1901722	Biological process	12	3	6.42E-03	1.00	BMP7; PDGFD; CFLAR
25	Heparan sulfate proteoglycan binding	GO:0043395	Molecular function	14	3	7.29E-03	1.00	FST; PTPRC; PTPRS
26	Adult locomotory behavior	GO:0008344	Biological process	77	7	7.37E-03	1.00	BTBD9; DAB1; FGF12; HIPK2; GRIN1; PARK2; PUM1
27	Ubiquitin-specific protease binding	GO:1990381	Molecular function	15	3	7.83E-03	1.00	MARCH6; PARK2; SELS
28	Histone methylation	GO:0016571	Biological process	129	9	8.28E-03	1.00	KDM4C; AUTS2; WHSC1L1; KDM4D; PRDM16; RDBP; SETDB2; SETD3; EED
29	Homophilic cell adhesion via plasma membrane adhesion molecules	GO:0007156	Biological process	161	12	8.79E-03	1.00	PALLD; PCDH7; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; CLSTN3; PCDHA9
30	Receptor serine/threonine kinase binding	GO:0033612	Molecular function	20	3	8.85E-03	1.00	SMAD6; BMP7; MAGI2
31	Lamin binding	GO:0005521	Molecular function	15	3	8.89E-03	1.00	TMEM201; PLCB1; LBR
32	Glomerular mesangium development	GO:0072109	Biological process	13	3	8.90E-03	1.00	BMP7; PDGFD; CFLAR

Supplementary Table 10. *(Continued)*

GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes	
33	Cell adhesion	GO:0007155	Biological process	1337	52	9.28E-03	1.00	EDIL3; TESK2; SPINK5; LILRB4; COL5A1; RC3H1; CTNND2; DAB1; NLGN1; PALLD; COL29A1; ODZ4; SIGLEC8; HES1; VWC2; LCK; LMO7; SMAD6; MLLT4; MUC4; MYO10; PCDH7; KIF26B; STAB2; LRRC16A; PRKCA; AXL; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; PTK2; PTPRC; PTPRS; TNN; BMP7; TECTA; TNXB; TPM1; PKP4; ADAM19; MPDZ; AIMP1; ADGRG1; NRXN3; NRXN1; CLSTN3; PCDHA9
34	Postsynaptic specialization membrane	GO:0099634	Cellular component	98	8	9.92E-03	1.00	NLGN1; GABRA1; GRID1; GRIN1; PTPRS; RAPSN; DGKI; CLSTN3
35	Biological adhesion	GO:0022610	Biological process	1344	52	1.01E-02	1.00	EDIL3; TESK2; SPINK5; LILRB4; COL5A1; RC3H1; CTNND2; DAB1; NLGN1; PALLD; COL29A1; ODZ4; SIGLEC8; HES1; VWC2; LCK; LMO7; SMAD6; MLLT4; MUC4; MYO10; PCDH7; KIF26B; STAB2; LRRC16A; PRKCA; AXL; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; PTK2; PTPRC; PTPRS; TNN; BMP7; TECTA; TNXB; TPM1; PKP4; ADAM19; MPDZ; AIMP1; ADGRG1; NRXN3; NRXN1; CLSTN3; PCDHA9
36	Histone-lysine n-methyltransferase activity	GO:0018024	Molecular function	44	5	1.06E-02	1.00	WHSC1L1; PRDM16; SETDB2; SETD3; EED
37	Postsynaptic density	GO:0014069	Cellular component	302	18	1.06E-02	1.00	ADORA1; CTNND2; DAB1; NLGN1; DISC1; GRID1; GRIN1; GRM3; ITPR1; ANKS1B; PTPRS; SPTBN1; PKP4; MPDZ; LIN7A; DGKI; CLSTN3; MAGI2
38	Regulation of embryonic development	GO:0045995	Biological process	123	9	1.07E-02	1.00	COL5A1; PLCB1; NIPBL; ODZ4; CRB2; HES1; BMP7; GORAB; SEPT7
39	Regulation of antigen receptor-mediated signaling pathway	GO:0050854	Biological process	56	5	1.09E-02	1.00	LILRB4; RC3H1; ELF2; LCK; PTPRC
40	Bmp signaling pathway	GO:0030509	Biological process	141	9	1.11E-02	1.00	GDF1; PDCD4; CRB2; HIPK2; HES1; VWC2; SMAD6; SFRP4; BMP7
41	Neuron to neuron synapse	GO:0098984	Cellular component	326	19	1.14E-02	1.00	ADORA1; CTNND2; DAB1; NLGN1; DISC1; GRID1; GRIN1; GRM3; ITPR1; PRKAR1B; ANKS1B; PTPRS; SPTBN1; PKP4; MPDZ; LIN7A; DGKI; CLSTN3; MAGI2
42	Positive T cell selection	GO:0043368	Biological process	34	4	1.17E-02	1.00	DOCK2; PTPRC; BCL11B; STAT6
43	Chaperone cofactor-dependent protein refolding	GO:0051085	Biological process	29	3	1.17E-02	1.00	GAK; HSPA9; DNAJB13

Supplementary Table 10. *(Continued)*

GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes	
44	Asymmetric synapse	GO:0032279	Cellular component	305	18	1.18E-02	1.00	ADORA1; CTNND2; DAB1; NLGN1; DISC1; GRID1; GRIN1; GRM3; ITPR1; ANKS1B; PTPRS; SPTBN1; PKP4; MPDZ; LIN7A; DGKI; CLSTN3; MAGI2
45	Activation of mapkk activity	GO:0000186	Biological process	49	5	1.21E-02	1.00	FRS2; ADORA1; MAP3K4; ZAK; MAP3K14
46	Proteoglycan binding	GO:0043394	Molecular function	32	4	1.23E-02	1.00	FST; COL5A1; PTPRC; PTPRS
47	Social behavior	GO:0035176	Biological process	47	5	1.25E-02	1.00	GRID1; GRIN1; ATXN1; NRXN3; NRXN1
48	Intraspecies interaction between organisms	GO:0051703	Biological process	47	5	1.25E-02	1.00	GRID1; GRIN1; ATXN1; NRXN3; NRXN1
49	Glomerulus development	GO:0032835	Biological process	58	6	1.25E-02	1.00	HES1; BMP7; PDGFD; CFLAR; PROM1; MAGI2
50	Regulation of behavior	GO:0050795	Biological process	64	5	1.26E-02	1.00	ADORA1; NLGN1; HCRTR2; NPS; NRXN1

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value.

Supplementary Table 11. Top 50 enriched Gene Ontology (GO) terms for 5-methylcytosine (5-mC) in the locus coeruleus (LC)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Homophilic cell adhesion via plasma membrane adhesion molecules	GO:0007156	Biological process	161	35	8.34E-15	7.25E-11	CDH2; CDH13; SDK1; PLXNB2; CDH20; DCHS2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; KIRREL3; PCDHGB4; CELSR1; PCDHGA8; PCDHA9
2	Cell-cell adhesion via plasma-membrane adhesion molecules	GO:0098742	Biological process	259	36	2.74E-10	1.19E-06	CDH2; CDH13; SDK1; PLXNB2; CLDN15; CDH20; DCHS2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; KIRREL3; PCDHGB4; CELSR1; PCDHGA8; PCDHA9

Supplementary Table 11. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
3	Calcium ion binding	GO:0005509	Molecular function	672	53	4.04E-07	1.17E-03	CDH2; CDH13; MYL9; CAPN11; MEGF6; NINL; CDH20; EHD4; LOXL2; MYO5A; NOTCH2; NOTCH3; NIN; LRP1B; DCHS2; PPP3R2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; CD248; EPS15L1; ELTD1; MCTP1; PITPNM3; NCALD; RAB11FIP4; PCDHGB4; DLK1; SYT7; CELSR1; PCDHGA8; PCDHA9
4	Cell-cell adhesion	GO:0098609	Biological process	782	57	1.26E-06	2.75E-03	CDH2; CDH13; MYL9; MAP3K8; CTNND1; DLG1; SDK1; PLXNB2; CLDN15; ALOX15; GNAS; CDH20; RNASE10; LAG3; LCK; LRP6; MLLT4; LEF1; PDCD1; PIK3R1; DCHS2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; METTL3; ZMIZ1; PTK2; TRPV4; MAD1L1; KIRREL3; SKAP1; RUNX3; PCDHGB4; CELSR1; PCDHGA8; PCDHA9
5	Integral component of plasma membrane	GO:0005887	Cellular component	1521	77	2.99E-04	5.20E-01	CDH2; B3GNT3; TSPAN9; GRIN3A; TMC1; OLFM3; CPT1C; CNTFR; CSPG4; DLG1; GPR183; EDNRB; ITGA11; PLXNB2; GABRP; EPHA6; GRIK5; TRHDE; IGF1R; IL11RA; TMPRSS9; KCNK3; SHISA6; LIFR; NOTCH2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; SEMA3G; TRPC7; SLC24A3; BAI1; KIAA1324; TRPV4; SCNN1G; ELTD1; SEMA4A; HHIP; SLC1A2; SLC6A12; SLC14A1; TACR1; TGFBR3; TRAF5; GPR172A; LRRC8C; MCHR2; CNTNAP4; PCDHGB4; SYT7; SLC28A1; KL; STX8; CELSR1; PCDHGA8; PCDHA9
6	Positive regulation of chromatin binding	GO:0035563	Biological process	13	4	5.97E-04	8.64E-01	DTX3L; CDT1; MED25; PARP9
7	High-density lipoprotein particle remodeling	GO:0034375	Biological process	17	4	9.87E-04	1	ABCA5; APOC1; LIPC; APOM

Supplementary Table 11. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
8	Intrinsic component of plasma membrane	GO:0031226	Cellular component	1594	77	1.13E-03	1	CDH2; B3GNT3; TSPAN9; GRIN3A; TMC1; OLFM3; CPT1C; CNTFR; CSPG4; DLG1; GPR183; EDNRB; ITGA11; PLXNB2; GABRP; EPHA6; GRIK5; TRHDE; IGF1R; IL11RA; TMPRSS9; KCNK3; SHISA6; LIFR; NOTCH2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; SEMA3G; TRPC7; SLC24A3; BAI1; KIAA1324; TRPV4; SCNN1G; ELTD1; SEMA4A; HHIP; SLC1A2; SLC6A12; SLC14A1; TACR1; TGFBR3; TRAF5; GPR172A; LRRC8C; MCHR2; CNTNAP4; PCDHGB4; SYT7; SLC28A1; KL; STX8; CELSR1; PCDHGA8; PCDHA9
9	Cell adhesion	GO:0007155	Biological process	1337	73	1.33E-03	1	CDH2; CDH13; MYL9; COL6A2; COL12A1; MAP3K8; CTNND1; DLG1; ENG; SDK1; ITGA11; PLXNB2; CLDN15; ALOX15; CADM2; AATF; GNAS; CDH20; TMEM102; GSK3B; RHOD; APLP1; RNASE10; LAG3; LCK; LOXL2; LRP6; MLLT4; NTM; LEF1; PDCD1; PIK3R1; DCHS2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; METTL3; ZMIZ1; PTK2; BAI1; TRPV4; TNXB; C4orf31; MAD1L1; KIRREL3; CNTNAP4; SKAP1; RUNX3; PCDHGB4; CELSR1; PCDHGA8; PCDHA9
10	Biological adhesion	GO:0022610	Biological process	1344	73	1.51E-03	1	CDH2; CDH13; MYL9; COL6A2; COL12A1; MAP3K8; CTNND1; DLG1; ENG; SDK1; ITGA11; PLXNB2; CLDN15; ALOX15; CADM2; AATF; GNAS; CDH20; TMEM102; GSK3B; RHOD; APLP1; RNASE10; LAG3; LCK; LOXL2; LRP6; MLLT4; NTM; LEF1; PDCD1; PIK3R1; DCHS2; PCDHGB3; PCDHGB2; PCDHGB1; PCDHGA7; PCDHGA6; PCDHGA5; PCDHGA4; PCDHGA3; PCDHGA2; PCDHGA1; PCDHAC2; PCDHAC1; PCDHA13; PCDHA12; PCDHA11; PCDHA10; PCDHA8; PCDHA7; PCDHA6; PCDHA5; PCDHA4; PCDHA3; PCDHA2; PCDHA1; METTL3; ZMIZ1; PTK2; BAI1; TRPV4; TNXB; C4orf31; MAD1L1; KIRREL3; CNTNAP4; SKAP1; RUNX3; PCDHGB4; CELSR1; PCDHGA8; PCDHA9
11	High-density lipoprotein particle	GO:0034364	Cellular component	25	4	1.68E-03	1	HDLBP; APOC1; LIPC; APOM
12	Mature B cell differentiation involved in immune response	GO:0002313	Biological process	17	4	2.32E-03	1	C17orf99; GPR183; NOTCH2; ITFG2
13	Glucagon secretion	GO:0070091	Biological process	10	3	2.42E-03	1	PASK; UCN; SYT7
14	Protein-lipid complex	GO:0032994	Cellular component	39	5	2.53E-03	1	BIN1; HDLBP; APOC1; LIPC; APOM

Supplementary Table 11. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
15	Positive regulation of double-strand break repair via nonhomologous end joining	GO:2001034	Biological process	10	3	2.83E-03	1	DTX3L; SETMAR; PARP9
16	Glucosidase activity	GO:0015926	Molecular function	12	3	3.33E-03	1	GANAB; GBA3; KL
17	Protein-lipid complex subunit organization	GO:0071825	Biological process	49	6	3.40E-03	1	ABCA5; BIN1; APOC1; LIPC; APOM; PCDHGA3
18	Cell cycle checkpoint	GO:0000075	Biological process	213	15	4.78E-03	1	FOXN3; RAD9B; DLG1; EIF2AK4; MSH2; ORC1L; NSUN2; CHFR; USP28; KIAA1967; SETMAR; XRCC3; CDT1; C12orf32; MAD1L1
19	Protein-lipid complex remodeling	GO:0034368	Biological process	28	4	4.82E-03	1	ABCA5; APOC1; LIPC; APOM
20	Plasma lipoprotein particle remodeling	GO:0034369	Biological process	28	4	4.82E-03	1	ABCA5; APOC1; LIPC; APOM
21	Nucleotide-sugar metabolic process	GO:0009225	Biological process	36	5	5.84E-03	1	GNPDA1; GUK1; PMM2; CSGALNACT1; GFPT2
22	Protein-containing complex remodeling	GO:0034367	Biological process	30	4	6.07E-03	1	ABCA5; APOC1; LIPC; APOM
23	Sterol transport	GO:0015918	Biological process	107	9	6.29E-03	1	ABCA5; STX12; APOC1; LIPC; LRP6; PPARG; APOM; SYT7; OSBPL2
24	Protein autoubiquitination	GO:0051865	Biological process	62	7	6.82E-03	1	DTX3L; RNF133; SASH1; UHRF1; RNF4; UBE3A; MTA1
25	Mature B cell differentiation	GO:0002335	Biological process	21	4	6.87E-03	1	C17orf99; GPR183; NOTCH2; ITFG2
26	Regulation of chromatin binding	GO:0035561	Biological process	23	4	6.90E-03	1	DTX3L; CDT1; MED25; PARP9
27	Negative regulation of lipoprotein particle clearance	GO:0010985	Biological process	21	3	6.92E-03	1	MYLIP; APOC1; LRPAP1
28	Cholesterol transport	GO:0030301	Biological process	97	8	7.00E-03	1	ABCA5; STX12; APOC1; LIPC; LRP6; PPARG; APOM; SYT7
29	Regulation of G0 to G1 transition	GO:0070316	Biological process	38	5	8.95E-03	1	E2F6; CBX5; APAF1; EHMT1; C12orf32
30	Negative regulation of cell cycle	GO:0045786	Biological process	566	29	9.00E-03	1	HEXIM1; FOXN3; RAD9B; DLG1; E2F6; CBX5; GAS2; BIN1; APAF1; EIF2AK4; MSH2; NOTCH2; ORC1L; NSUN2; USP47; CHFR; RPTOR; USP28; KIAA1967; SETMAR; XRCC3; EHMT1; CDT1; MED25; C12orf32; MAD1L1; SLC25A33; RUNX3; KAT2B

Supplementary Table 11. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
31	Radial glial cell differentiation	GO:0060019	Biological process	12	3	9.39E-03	1	CDH2; LEF1; METTL3
32	G0 to G1 transition	GO:0045023	Biological process	40	5	1.02E-02	1	E2F6; CBX5; APAF1; EHMT1; C12orf32
33	Plasma lipoprotein particle	GO:0034358	Cellular component	37	4	1.08E-02	1	HDLBP; APOC1; LIPC; APOM
34	Lipoprotein particle	GO:1990777	Cellular component	37	4	1.08E-02	1	HDLBP; APOC1; LIPC; APOM
35	Aldehyde dehydrogenase (NAD) activity	GO:0004029	Molecular function	14	3	1.15E-02	1	ALDH2; ALDH1A3; ALDH1A2
36	Plasma lipoprotein particle clearance	GO:0034381	Biological process	71	6	1.17E-02	1	MYLIP; HDLBP; APOC1; LIPC; LRPAP1; APOM
37	Intrinsic apoptotic signaling pathway by p53 class mediator	GO:0072332	Biological process	74	7	1.27E-02	1	ANKRD2; HIPK2; MSH2; WWOX; USP28; AEN; BCL2L12
38	Positive regulation of heterotypic cell-cell adhesion	GO:0034116	Biological process	15	3	1.33E-02	1	ALOX15; LCK; SKAP1
39	Intrinsic apoptotic signaling pathway in response to dna damage by p53 class mediator	GO:0042771	Biological process	42	5	1.41E-02	1	HIPK2; MSH2; USP28; AEN; BCL2L12
40	Base conversion or substitution editing	GO:0016553	Biological process	20	3	1.41E-02	1	ADARB2; RBM47; METTL3
41	Regulation of blood pressure	GO:0008217	Biological process	165	11	1.44E-02	1	CORIN; CPA3; CYP11B1; EDNRB; DDAH1; ERAP1; PPARG; UCN; NCALD; NAV2; RASL10B
42	Triglyceride lipase activity	GO:0004806	Molecular function	23	3	1.48E-02	1	CES5A; CES8; LIPC
43	B cell activation involved in immune response	GO:0002312	Biological process	67	6	1.50E-02	1	C17orf99; GPR183; CD180; MSH2; NOTCH2; ITFG2
44	Protein-lipid complex assembly	GO:0065005	Biological process	32	4	1.57E-02	1	BIN1; APOC1; APOM; PCDHGA3
45	Retinoid X receptor binding	GO:0046965	Molecular function	15	3	1.63E-02	1	PPARG; RARB; MED25
46	SNAP receptor activity	GO:0005484	Molecular function	29	4	1.67E-02	1	VTI1A; STX2; STX12; STX8

Supplementary Table 11. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
47	Reverse cholesterol transport	GO:0043691	Biological process	19	3	1.70E-02	1	ABCA5; LIPC; APOM
48	Very-low-density lipoprotein particle clearance	GO:0034447	Biological process	10	2	1.77E-02	1	APOC1; LRPAP1
49	Secondary palate development	GO:0062009	Biological process	24	4	1.92E-02	1	DLG1; LEF1; MMP25; TGFBR3
50	Synaptic vesicle exocytosis	GO:0016079	Biological process	118	10	2.01E-02	1	TRIM9; GRIN3A; RIMS4; STX2; GRIK5; GSK3B; PRKCB; CADPS; SYT7; CADPS2

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value and those that are significantly enriched after multiple testing correction (FDR adjusted $p < 0.05$) are seperated from the other terms by a dashed line.

Supplementary Table 12. Top 50 enriched Gene Ontology (GO) terms for 5-hydroxymethylcytosine (5-hmC) in the locus coeruleus (LC)

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
1	Regulation of calcium ion-dependent exocytosis	GO:0017158	Biological process	104	12	1.83E-04	8.38E-01	CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; GRIK5; GSK3B; NOTCH1; PRKCB; WNT7A; CACNA1I; CACNA1H
2	Neurotrophin TRK receptor signaling pathway	GO:0048011	Biological process	31	6	2.78E-04	8.38E-01	GRB2; NDN; CASP3; WASF1; HAP1; BCAR1
3	Calcium channel activity	GO:0005262	Molecular function	122	14	3.00E-04	8.38E-01	GRIN3A; CACNG4; ITPR1; CATSPER4; CACNA2D3; JPH1; JPH3; RYR3; TRPV1; CACNA1C; CACNB4; ORAI2; CACNA1I; CACNA1H
4	Response to carbohydrate	GO:0009743	Biological process	222	16	4.86E-04	8.38E-01	COL6A2; EIF2B1; PTK2B; GIPR; GRIK5; LEP; PKLR; PRKCB; APOM; MAPK13; PTPRN2; SMARCA4; SPARC; CDT1; CASP3; BRSK2
5	Negative regulation of epithelial cell differentiation	GO:0030857	Biological process	46	6	5.36E-04	8.38E-01	YAP1; CTNNB1; JAG1; GSK3B; NOTCH1; GRHL2
6	Beta-catenin-TCF complex assembly	GO:1904837	Biological process	30	6	5.68E-04	8.38E-01	CTNNB1; LEF1; SMARCA4; TERT; TLE2; TCF7L1
7	Regulation of regulated secretory pathway	GO:1903305	Biological process	147	13	1.13E-03	9.11E-01	CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; ABR; GRIK5; GSK3B; NOTCH1; PRKCB; WNT7A; CACNA1I; CACNA1H
8	Calcium ion transmembrane transporter activity	GO:0015085	Molecular function	138	14	1.26E-03	9.11E-01	GRIN3A; CACNG4; ITPR1; CATSPER4; CACNA2D3; JPH1; JPH3; RYR3; TRPV1; CACNA1C; CACNB4; ORAI2; CACNA1I; CACNA1H

Supplementary Table 12. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
9	Embryonic limb morphogenesis	GO:0030326	Biological process	125	12	1.52E-03	9.11E-01	CTNNB1; NOTCH1; CHST11; LEF1; SKI; C5orf42; TFAP2A; WNT7A; CACNA1C; GRHL2; FRAS1; ALDH1A2
10	Embryonic appendage morphogenesis	GO:0035113	Biological process	125	12	1.52E-03	9.11E-01	CTNNB1; NOTCH1; CHST11; LEF1; SKI; C5orf42; TFAP2A; WNT7A; CACNA1C; GRHL2; FRAS1; ALDH1A2
11	Ionotropic glutamate receptor activity	GO:0004970	Molecular function	19	5	1.52E-03	9.11E-01	GRIN3A; PTK2B; GRID2; GRIK4; GRIK5
12	Neurotrophin signaling pathway	GO:0038179	Biological process	39	6	1.73E-03	9.11E-01	GRB2; NDN; CASP3; WASF1; HAP1; BCAR1
13	Appendage morphogenesis	GO:0035107	Biological process	146	13	1.83E-03	9.11E-01	CTNNB1; FMN1; NOTCH1; CHST11; LEF1; SKI; C5orf42; TFAP2A; WNT7A; CACNA1C; GRHL2; FRAS1; ALDH1A2
14	Limb morphogenesis	GO:0035108	Biological process	146	13	1.83E-03	9.11E-01	CTNNB1; FMN1; NOTCH1; CHST11; LEF1; SKI; C5orf42; TFAP2A; WNT7A; CACNA1C; GRHL2; FRAS1; ALDH1A2
15	Calcium ion regulated exocytosis	GO:0017156	Biological process	152	13	1.94E-03	9.11E-01	CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; ERC2; GRIK5; GSK3B; NOTCH1; PRKCB; WNT7A; CACNA1I; CACNA1H
16	Response to muramyl dipeptide	GO:0032495	Biological process	20	4	2.18E-03	9.11E-01	JAG1; NOTCH1; CARD9; TRIM41
17	Calcium-release channel activity	GO:0015278	Molecular function	21	5	2.20E-03	9.11E-01	ITPR1; JPH1; JPH3; RYR3; TRPV1
18	Voltage-gated calcium channel complex	GO:0005891	Cellular component	43	7	2.40E-03	9.11E-01	CACNG4; CATSPER4; CACNA2D3; CACNA1C; CACNB4; CACNA1I; CACNA1H
19	Centrosome	GO:0005813	Cellular component	510	25	2.41E-03	9.11E-01	NUBP2; TUBGCP2; TTC8; CTNNB1; RAP1GAP2; ZFYVE26; HEPACAM2; DISC1; GNAI1; GSK3B; NDN; NPM1; KIAA1377; SKI; TBCD; TFAP2A; TTC26; DYNLRB1; MAD1L1; NLRC5; FBF1; HAP1; BRSK2; KIAA0101; CDC25B
20	Pulmonary valve morphogenesis	GO:0003184	Biological process	16	4	2.52E-03	9.11E-01	JAG1; HEYL; NOTCH1; ROBO2
21	Forelimb morphogenesis	GO:0035136	Biological process	41	6	2.59E-03	9.11E-01	CTNNB1; FMN1; TFAP2A; WNT7A; CACNA1C; ALDH1A2
22	Nephron tubule development	GO:0072080	Biological process	93	9	2.83E-03	9.11E-01	YAP1; CTNNB1; ADAMTS16; DLG1; JAG1; HEYL; FMN1; NOTCH1; KIF26B
23	Voltage-gated calcium channel activity	GO:0005245	Molecular function	47	7	3.17E-03	9.11E-01	CACNG4; CATSPER4; CACNA2D3; CACNA1C; CACNB4; CACNA1I; CACNA1H
24	Synaptic vesicle cycle	GO:0099504	Biological process	195	15	3.23E-03	9.11E-01	CDH2; CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; CTNNB1; ERC2; GRIK5; GSK3B; PRKCB; DENND1A; WNT7A; BRSK2; DNAJC6
25	Radial glial cell differentiation	GO:0060019	Biological process	12	3	3.31E-03	9.11E-01	CDH2; LEF1; METTL3
26	Renal tubule development	GO:0061326	Biological process	95	9	3.56E-03	9.11E-01	YAP1; CTNNB1; ADAMTS16; DLG1; JAG1; HEYL; FMN1; NOTCH1; KIF26B

Supplementary Table 12. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
27	Z disc	GO:0030018	Cellular component	126	12	3.59E-03	9.11E-01	CTNNB1; BIN1; ANK3; PARVB; MURC; NRAP; JPH1; RYR3; CACNA1C; CSRP3; SORBS2; ACTN1
28	Response to monosaccharide	GO:0034284	Biological process	200	13	3.73E-03	9.11E-01	COL6A2; EIF2B1; PTK2B; GIPR; GRIK5; LEP; PKLR; APOM; PTPRN2; SMARCA4; SPARC; CASP3; BRSK2
29	Synaptic vesicle localization	GO:0097479	Biological process	165	13	3.89E-03	9.11E-01	CDH2; CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; CTNNB1; ERC2; GRIK5; GSK3B; PRKCB; WNT7A; BRSK2
30	Foregut morphogenesis	GO:0007440	Biological process	10	3	3.95E-03	9.11E-01	CTNNB1; SMAD3; NOTCH1
31	Ionotropic glutamate receptor signaling pathway	GO:0035235	Biological process	25	5	4.00E-03	9.11E-01	GRIN3A; PTK2B; GRID2; GRIK4; GRIK5
32	SH3/SH2 adaptor activity	GO:0005070	Molecular function	54	7	4.17E-03	9.11E-01	CLNK; GRB2; BLNK; SHB; TP53BP2; SKAP1; GRAP2
33	Ligand-gated ion channel activity	GO:0015276	Molecular function	144	12	4.35E-03	9.11E-01	GRIN3A; DLG1; PTK2B; GABRP; GRID2; GRIK4; GRIK5; ITPR1; JPH1; JPH3; RYR3; TRPV1
34	Ligand-gated channel activity	GO:0022834	Molecular function	144	12	4.35E-03	9.11E-01	GRIN3A; DLG1; PTK2B; GABRP; GRID2; GRIK4; GRIK5; ITPR1; JPH1; JPH3; RYR3; TRPV1
35	Regulation of translational initiation in response to stress	GO:0043558	Biological process	14	3	4.39E-03	9.11E-01	EIF2AK4; NPM1; PPP1CA
36	Cell cycle	GO:0007049	Biological process	1768	62	4.47E-03	9.11E-01	ERN2; CGREF1; SEPT9; TUBGCP2; POLS; CTNNB1; DLG1; E2F3; WAPAL; UNC84A; ZFYVE26; HEPACAM2; SGSM3; BIN1; GNAI1; LIN9; ANK3; NR3C1; UHRF1; HELLS; KIF22; LEP; SMAD3; MAX; EIF2AK4; TRIM37; MYH9; NOTCH1; NPM1; MOV10L1; DDX4; PPP1CA; USP47; PPP2R2A; PPP6C; CHFR; PRKCB; MAPK13; METTL3; PCBP4; STARD9; MICAL3; KIAA1377; USP29; MRPL41; BOLL; TBCD; TERT; TP53BP2; XRCC3; CCNJL; WDR76; CDT1; CASP3; C12orf32; MAD1L1; AFAP1L2; RAE1; BRSK2; BRE; KIAA0101; CDC25B
37	Calcium channel complex	GO:0034704	Cellular component	65	8	4.48E-03	9.11E-01	CACNG4; CATSPER4; CACNA2D3; RYR3; CACNA1C; CACNB4; CACNA1I; CACNA1H
38	Embryonic forelimb morphogenesis	GO:0035115	Biological process	33	5	4.48E-03	9.11E-01	CTNNB1; TFAP2A; WNT7A; CACNA1C; ALDH1A2
39	Positive regulation of neuroblast proliferation	GO:0002052	Biological process	22	4	4.58E-03	9.11E-01	CTNNB1; DCT; DISC1; NOTCH1
40	Labyrinthine layer morphogenesis	GO:0060713	Biological process	21	4	4.60E-03	9.11E-01	GRB2; LEF1; GRHL2; NCOA1
41	Nucleobase-containing compound catabolic process	GO:0034655	Biological process	546	23	4.61E-03	9.11E-01	LRPPRC; ERN2; POLS; ACOT7; DPYS; XRN2; NT5C2; TNRC6B; SMG6; KPNA1; NPM1; CECR1; PKLR; PPP2R2A; METTL3; PCBP4; RNASEL; RPL4; ERI3; ADPGK; CASP3; KHSRP; ENTPD1

Supplementary Table 12. *(Continued)*

Rank	GO term	ID	Ontology	Number of genes	Differentially modified	*p* value	FDR adjusted *p*	Genes
42	ncRNA metabolic process	GO:0034660	Biological process	537	20	4.63E-03	9.11E-01	NPM3; ERN2; TBL3; POLS; XRN2; RPRD2; IARS; SMAD3; TYW1B; RRP15; POLR2A; MOV10L1; DDX4; METTL3; RNASEL; SMARCA4; SP1; ERI3; METTL8; KHSRP
43	Establishment of blood-brain barrier	GO:0060856	Biological process	10	3	4.93E-03	9.11E-01	CTNNB1; TRPV1; WNT7A
44	Ionotropic glutamate receptor complex	GO:0008328	Cellular component	51	7	4.96E-03	9.11E-01	GRIN3A; DLG1; PTK2B; CACNG4; GRID2; GRIK4; GRIK5
45	Response to UV	GO:0009411	Biological process	135	9	5.05E-03	9.11E-01	ERCC5; EIF2AK4; NPM1; USP47; MAPK13; METTL3; CASP3; C12orf32; KIAA0101
46	Glutamatergic synapse	GO:0098978	Cellular component	350	23	5.13E-03	9.11E-01	WASF3; CPLX1; CHRM2; GRIN3A; DLG1; PTK2B; ABLIM3; ERC2; CACNG4; GRID2; ABR; GRIK5; GSK3B; HIP1; KPNA1; PPP1CA; PPP2R2A; ARHGAP22; CLSTN2; SPARC; WNT7A; GPR123; ACTN1
47	DNA metabolic process	GO:0006259	Biological process	987	36	5.15E-03	9.11E-01	MCRS1; POLS; CTNNB1; RNF168; ERCC5; PTK2B; XRN2; WAPAL; SMG6; ZFYVE26; SHPRH; LIN9; UHRF1; GTF2H3; HELLS; KIF22; KPNA1; NPM1; LEF1; POLR2A; MOV10L1; DDX4; USP47; NEIL3; FAM111A; COPS7B; TERT; XRCC3; ERI3; GRHL2; CDT1; CASP3; C12orf32; TOP3B; BRE; KIAA0101
48	Regulation of exocytosis	GO:0017157	Biological process	203	14	5.26E-03	9.11E-01	CPLX1; CHRM2; TRIM9; GRIN3A; RIMS4; ABR; GRIK5; GSK3B; NOTCH1; PRKCB; WNT7A; CACNA1I; CACNA1H; HAP1
49	Semi-lunar valve development	GO:1905314	Biological process	32	5	5.36E-03	9.11E-01	JAG1; ELN; HEYL; NOTCH1; ROBO2
50	Protein-DNA complex	GO:0032993	Cellular component	189	9	0.00551393	0.910798446	CTNNB1; ERCC5; SHPRH; GTF2H3; MAX; NPM1; LEF1; SP1; TERT

Displayed for each ranked GO term is the identifier (ID), the related ontology, the total number of genes belonging to the GO term, the number of genes that are differentially modified, the *p* value, the false discovery rate (FDR) adjusted *p* value and the names of the differentially modified genes. The GO terms are ranked based on their *p* value.

Supplementary Table 13. Primer overview Tenascin XB (*TNXB*)

PCR primers

Gene	Forward primer (5'-3')	Reverze primer (5'-3')	Target region (GRCh37)	Product size (bp)	Analysis
TNXB	(Bio-)GTGAAGTAGGTTATT TAGGGGAAGAT	CTCACTCACCTAACCT TCCCAACAAAT	6:32063788-32063996	209	Pyroseq. validation
TNXB (Multiplex)	TGTGAAGTAGG TTATTTAGGGGA AGA	CCTCCAAAACAAACCC TAACTAAAAACTCT	6:32063558-32063997	440	Cell-specific DRN validation
TNXB (Singleplex)	TGTGAAGTAGG TTATTTAGGGGA AGA	(Bio-)ACAAACCCTAACTAAA AACTCTTCCTACC	6:32063566-32063997	432	Cell-specific DRN validation
DNAJB13 (Multiplex)	GGTATTTTGGGA GGATGAGTTA	CCCTTCAAAAACAAAC CAACTAAT	11:73668561-73668825	265	Cell-specific DRN validation
DNAJB13 (Singleplex)	GTATTTTGGGAG GATGAGTTATAA TTG	(Bio-)CTTCAAAAACAAACC AACTAATAC	11:73668563-73668824	262	Cell-specific DRN validation

Pyrosequencing primers

Gene	Sequencing primer (5'-3')	Number of CpGs	Target region (GRCh37)	PyroMark Orientation	Analysis
TNXB	CAATCATTCCAACACT ACCT	8	6:32063869-32063940	Lower strand (5'-3')	Pyroseq. validation
TNXB (Primer 1)	GGTTAGTGTTTAGATG GG	5	6:32063874-32063913	Lower strand (5'-3')	Cell-specific DRN validation
TNXB (Primer 2)	AGGATTGAGGTGTGA	4	6:32063823-32063867	Lower strand (5'-3')	Cell-specific DRN validation
TNXB (Primer 3)	GGTGTGTGTATTTGTT G	2	6:32063774-32063806	Lower strand (5'-3')	Cell-specific DRN validation
DNAJB13 (Primer 1)	GGATGAGTTATAATTG GG	5	11:73668745-73668795	Lower strand (5'-3')	Cell-specific DRN validation
DNAJB13 (Primer 2)	TTTTGTTATTTGTTTGA AAA	4	11:73668618-73668673	Lower strand (5'-3')	Cell-specific DRN validation

Displayed is an overview of the polymerase chain reaction (PCR) and pyrosequencing primers for *TNXB* and *DNAJB13* that were used in this study. Provided for each of the target genes are the primer sequences (Bio, biotinylated), as well as corresponding genomic coordinates (Ensembl GRCh37 assembly), amplicon sizes in base pairs (bp) and the analysis in which the primers have been applied (Pyrosequencing validation or cell-specific dorsal raphe nuclei (DRN) validation). For the pyrosequencing primers, the total number of CpG sites, the targeted region for sequencing, the orientation of the assay in the PyroMark assay design software 2.0 and the analysis in which the primers have been applied are listed per gene.

Supplementary Table 14. Overview of the limiting dilution bisulfite pyrosequencing (LDBSP) parameters for Tenaxcin XB (*TNXB*)			
Singleplex PCR	**Control**	**AD**	**Total**
Number (+ percentage) of reactions with PCR product	104 (13.13)	133 (16.79)	237 (14.96)
Allele estimation			
Number (+ percentage) of reactions with 1 allele	97 (93.27)	120 (90.23)	217 (91.56)
Number (+ percentage) of reactions with 2 alleles	5 (4.81)	8 (6.02)	13 (5.49)
Number (+ percentage) of reactions with 3 alleles	2 (1.92)	5 (3.76)	7 (2.95)
Number (+ percentage) of multi-allele reactions	7 (6.73)	13 (9.77)	20 (8.44)
Number (+ percentage) of recovered alleles	113 (3.14)	151 (4.19)	264 (3.67)
A summary of all the LDBSP parameters for pools of 50 serotonergci neurons isolated from unfixed post-mortem brain tissue using laser capture microdissection (LCM). Displayed are the total number of obtained polymerase chain reaction (PCR) products per experimental groups and the total amount of estimated alleles. Please refer to the written text for further specifications on the allele estimation method and criteria.			

Chapter 6

Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Twan van den Beucken[3]

[1]Department of Psychiatry & Neuropsychology, Graduate School MHeNS (School for Mental Health and Neuroscience), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilian University, Wuerzburg, Germany.
[3]Department of Toxicogenomics, Graduate School GROW (Research School for Oncology and Developmental Biology), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.

Neurobiology of Learning and Memory. 2020 Sep 22;107318.
Doi: 10.1016/j.nlm.2020.107318.

Abstract
The discovery of induced pluripotent stem cell (iPSC) technology has the potential to accelerate scientific research for Alzheimer's disease (AD). iPSCs are therefore increasingly considered for AD modeling and drug development. Nevertheless, most of the work conducted so far has mainly focused on iPSC models from patients with familial AD (fAD), while actually sporadic AD (sAD) is more prevalent and represents over 90% of the AD cases in the population. The development of more sAD models is therefore key for studying this multifactorial disorder. In fact, probing the unique genomes of sAD patients and their interaction with AD-associated environmental factors could contribute to a better understanding of this disease. However, initial iPSC-based models for sAD have shown a high degree of variability and inconsistencies in terms of AD hallmarks. In this review, we provide an overview of the studies that have been conducted for sAD so far. In addition, we critically assess important sources of variability related to the model in addition to those that might be explained by the heterogeneous nature of sAD. These considerations might aid in developing more consistent iPSC models of sAD, which could help in developing a better understanding of the molecular mechanisms underlying the disease.

Keywords
Alzheimer's Disease, induced pluripotent stem cells, somatic mosaicism.

Introduction

With the increasing life span of the general population, Alzheimer's disease (AD), which is an age-related neurodegenerative disorder, is becoming a major disease burden and socio-economic challenge worldwide. It was estimated in 2015 that 46.8 million people in the world are living with AD or a related form of dementia and it is expected that this number will double by 2040 [1]. The vast majority, *i.e.* more than 95%, of the AD cases are characterized by a late-onset form and sporadic development of the disease (that is sporadic AD (sAD)). Its origin is multi-factorial, having both genetic and environmental factors, as well as their complex interactions, contributing to the onset and course of the disease. On the other hand, familial AD (fAD), which represent the early-onset form, is known to arise from mutations in genes such as amyloid precursor protein (APP), presenilin 1 (PSEN1) and presenilin 2 (PSEN2) [2]. Although distinguishable based on their time of onset and mode of transmission, both types of AD appear identical in light of clinical features and neuropathological hallmarks. Lesions of accumulated amyloid beta (Aβ) protein and neurofibrillary tangles consisting of hyperphosphorylated tau protein are typically observed in AD brains, and are known to contribute to the neuronal cell death and immune activation, leading to region-specific atrophy. The loss of neural integrity in regional neurotransmitter systems is thought to underlie the clinical symptoms, such as cognitive decline and learning deficits, which are typically observed in AD patients.

In the past decades, our knowledge on the pathophysiology that underlies AD has advanced tremendously. In particular, the neurotoxicity of aberrant Aβ and neurofibrillary tangles are now much better understood. However, efforts to translate these insights into clinical success have floundered. Although researchers have conducted a large number of clinical trials for potential treatments for AD, almost none of the drugs have been brought to the market. Despite the blame being placed on a variety of factors, one of the main concerns has been the animal and cellular models that were used in the initial stages of drug development [3]. Most of these have relied on transgenic models harboring genetic mutations associated with fAD. Despite the fact that these have proven to be instrumental for basic research, they only partially reflect AD phenotypes and have thus far have been unable to explain the complex and heterogeneous nature of sAD. It has therefore been debated that more accurate models for sAD harboring the enormous amounts of individual genetic backgrounds within the patient population are required to further improve our insights into sAD development and progression. Moreover, it has even been hypothesized that such models would potentially be able to aid in the identification of novel biomarkers for early detection of sAD, as well as in the discovery of new and much needed therapeutics.

With the foundation of human induced pluripotent stem cell (iPSCs) over a decade ago, new doors for AD-research have been opened [4]. Since their discovery, human iPSCs have been offering a promising avenue to fill the gap between animal and human research, providing a new platform for disease modelling and drug testing that has the potential to broaden our knowledge on the underlying pathophysiology. Pluripotent stem cell technology is currently providing an unlimited source of patient-derived cells that can be differentiated into disease-relevant cells, including neurons and glia (reviewed in [5, 6]). The majority of the conducted work so far in AD has, however, used iPSC-cells derived from patients with fAD where, similar to previous animal studies, specific genetic mutations were studied that drive the disease onset. The establishment of more iPSC-based models from sAD patients is therefore necessary and could potentially benefit research into disease etiology and development of therapeutic strategies. However, developing iPSC models of sAD remains challenging due to the multifactorial nature of the disease, the live-long disease progression and the high degree of inter-individual heterogeneity that might be reflected in cells derived from sAD patients. Furthermore, aside from the variability associated with the disease, several technological challenges currently remain that could further impact on the variability that is observed in these sAD-iPSC models. In this review, we therefore provide an overview of the studies that have been conducted thus far and we critically assess important sources of variability related to the model in addition to those that might be explained by the heterogeneous nature of the disease.

iPSC-derived models for sAD

To date, a limited number of studies has been performed with iPSC-derived neurons from sAD patients (Table 1). Modeling AD using patient-derived iPSCs was initiated from fAD cases with known mutations in disease-causing genes, such as *APP*, *PSEN1* and *PSEN2*. iPSC-derived neurons from sAD patients in these studies were often studied in parallel, with the main goal of seeking to find AD-associated cellular phenotypes for the validation of their potential for sAD modeling. For this reason, many of these reports specifically look into the presence of AD-associated hallmarks such as amyloid beta, hyperphosphorylated tau and elevated levels of GSK-3β. The very first study in 2012, which is a comparative study using iPSC-derived neurons from fAD and sAD patients [7], demonstrated increased levels of Aβ(1–40) that were found to be secreted from sAD neurons in comparison to non-AD controls. These findings in the sAD neurons were in line with the observations in neurons from fAD patients. However, this was only observed using iPSC-derived neurons from one out of two sAD patients. Secreted levels of Aβ(1–40) using iPSC-derived neurons from a second patients were similar to levels obtained using iPSC-derived

neurons from two non-demented control subjects. Also other hallmarks of AD, *i.e.* phosphorylated tau and active GSK-3β, were elevated in the iPSC-derived neurons that exhibited elevated Aβ(1–40) secretion, while the levels of these markers for the other sAD patient-derived neurons were again close to those observed in non-demented controls. Overall, this study suggests that iPSC-derived neurons from sAD patients can display similar hallmarks to fAD lines, but that these are not evident for every sAD patient. Interestingly, this heterogeneity corresponds with the complexity of sAD and our general understanding on disease causation and progression. In a later study, Kondo and colleagues reported that iPSC-derived neurons from 2 distinct sAD patients did not show elevated secretion of Aβ(1–40) and Aβ(1–42) compared to iPSC-derived neurons obtained from 3 control subjects [8]. Instead, neurons derived from one sAD patient displayed elevated levels of intracellular Aβ fragments, while this was not observed for the second sAD patient line. Finally, Foveau *et al.* reported the generation of 3 iPSC-derived neural models from sAD patients and 3 control lines from non-cognitively impaired individuals [9]. Secreted levels of Aβ(1–40) and Aβ(1–42) tended to be lower in sAD derived neural models although this was not statistically significant. No changes in the ratio Aβ(1–42)/Aβ(1–40) were observed for sAD derived neural models compared to controls

In addition, several studies have also directly focused on patient-derived iPSC lines that harbor known sAD genetic risk factors. In the study by Duan and coworkers, iPSC-derived neurons were generated from 3 sAD patients all carrying the *APOE3/E4* genotype [10]. In this study, increased extracellular levels of Aβ(1–42) were reported for two out of three patient-derived neural models, while Aβ(1–40) levels remained unchanged compared to a panel of iPSC-derived neurons from control subjects. Ochalek and coworkers found increased extracellular levels of both Aβ(1–40) and Aβ(1–42) in neurons derived from 4 sAD patients compared to control neurons, while the ratio Aβ(1–42)/ Aβ(1–40) did not change relative to controls [11]. In addition, hyperphosphorylation of tau, increased GSK3β activity, APP synthesis and APP C-terminal cleavage was detected in neurons from sAD compared to neurons obtained from iPSC control lines. The study performed by Young *et al.* focused on *SORL1*, which is an endocytic trafficking factor whose levels modulate the processing of APP to Aβ and other proteolytic products implicated in sAD [12]. Loss of *SORL1* expression has been documented in sAD cases and has been associated with sAD in both candidate gene and GWAS analyses. By studying patient iPSC-derived neurons, this latter study confirmed the importance of the SORL1/APP pathway in sAD, and their findings corroborated previous studies in cell and animal models [12]. In another study by Hossini *et al.*, sAD iPSC-derived neurons were analyzed to assess their reflection of disease phenotype in gene expression patterns and the

expression of typical AD proteins [13]. The differentiated neurons reflected sAD phenotypes due to the presence of phosphorylated tau proteins and the upregulation of GSK-3β. Further analysis of the neurons revealed significant changes in the expression of other genes associated with AD, including subunits of the proteasome complex. Moreover, a disease-specific protein association network that models AD pathology on the transcriptome level could be generated from the AD-iPSC lines.

Taken together, these studies have demonstrated that sAD patients' iPSC-derived neurons are able to recapitulate neuropathological processes of the disease, which represent critical first steps in assessing the potential of using iPSCs in sAD research. Simultaneously, these studies have also proven that a high degree of variability in terms of disease hallmarks in iPSC-derived neurons from sAD patients is common. Variability between cell lines in terms of the Aβ type, *i.e.* Aβ(1–42) or Aβ(1–40), which is altered in sAD derived neurons, as well as variability between iPSC-derived neurons from different sAD patients within a study (Table 1). While some of this might be explained due to heterogeneous nature of sAD, it is likely that other source of variability related to the model play a role in this as well.

Until now, most of the described work on sAD models has been conducted using iPSC-derived neurons, mainly because protocols to generate human iPSC-derived microglia and astrocytes were absent or laborious. However, over recent years, several protocols have been published for differentiating iPSC into microglia [14-17]. This has been a breakthrough for studying the function of microglia in human disease as human microglia appear very distinct from its murine counterpart especially in regards to immune activation [18, 19]. Human microglia were found to have higher expression of genes involved in pathways related to longevity and anti-inflammatory responses compared to rodent microglia [19]. This suggests that the function of human microglia may not be well recapitulated in murine AD models.

Applying iPSC-derived microglia, Lin *et al.* investigated the role of the most significate AD risk gene, *APOE4* on human microglia function and gene expression [20]. For this an *APOE3/3* iPSC line was edited using CRISPR/Cas9 technology to create an isogenic *APOE4/4* line. This modification was found to result in reduced phagocytic activity of the microglia and altered gene expression levels of over 1,000 genes. In a reciprocal approach, gene editing of a patient-derived *APOE4/4* bearing iPSC line into an *APOE3/3* genotype restored the impaired phagocytic phenotype.

Table 1. Overview of published studies using induced pluripotent stem cell (iPSC)-derived neurons from sporadic Alzheimer's disease (sAD) patients.

Study	ID	Aβ(1–40)	Aβ(1–42)	Aβ42/40	int.Aβ	p-TAU	aGSK3β	Sex	#iPCS lines per patient
Israel *et al.* (2012)	sAD1	-	N.R.	N.R.	N.R.	-	-	F	3
	sAD2	↑	N.R.	N.R.	N.R.	↑	↑	M	5
Kondo *et al.* (2013)	AD3E211	-	-	-	-	N.R.	N.R.	M	1
	AD8K213	-	-	-	↑	N.R.	N.R.	M	1
Duan *et al.* (2014)	AG04402	N.R.	↑	↑	N.R.	N.R.	N.R.	M	1
	AG11414	N.R.	↑	↑	N.R.	N.R.	N.R.	M	1
	AG05810	N.R.	-	-	N.R.	N.R.	N.R.	F	1
Ochalek *et al.* (2017)	BIOT-0904-LOAD	↑	↑	-	N.R.	↑	↑	F	2
	BIOT-0630-LOAD	↑	↑	-	N.R.	↑	↑	M	1
	BIOT-4828-LOAD	↑	↑	-	N.R.	↑	↑	F	2
	BIOT-0726-LOAD	↑	↑	-	N.R.	↑	↑	F	1
Foveau *et al.* (2019)	CQ2	-	-	-	N.D.	N.R.	N.R.	F	2
	CQ3	-	-	-	N.R.	N.R.	N.R.	M	2
	CQ6	-	-	-	N.R.	N.R.	N.R.	M	2

ID is identifier, in.Aβ is intracellular Aβ, p-TAU is phosphorylated TAU, aGSK3β is GSK3β activity, N.R. is not reported, N.D. is not detected, F is female, M is male, - is no change, ↑ is increased.

A series of studies illustrates the importance of studying another AD risk gene, triggering receptor expressed on myeloid cells-2 (*TREM2*) in a human microglia model. In humans a heterozygous *TREM2 R47H* mutation confers increased AD risk, almost to a similar degree as a single APOE4 e4 allele [21, 22]. Mice bearing a heterozygous *TREM R47H* mutation exhibited aberrant *TREM2* mRNA splicing and reduced expression of *TREM2* mRNA and protein [23]. Strikingly, *TREM2* mRNA levels and splicing were normal in human iPSC-derived microglia carrying the *TREM2 R47H* mutation. In line with this, Claes and colleagues found that heterozygous *TREM2 H47H* iPSC-derived human microglia displayed normal phagocytosis capacity in contrast to *TREM2* heterozygous and homozygous knock-out microglia [24]. This indicates that *TREM2 R47H* phenotypes found in murine models cannot be translated to humans, and highlights the need for adequate human microglia models to study sAD.

Differentiation protocols to generate human iPSC-derived astrocytes from iPSC lines have been published nearly a decade ago [8, 25]. However, these protocols were laborious and yielded only a low percentage of differentiated astrocytes. Nevertheless, using this method it was found that astrocytes differentiated from

sAD and fAD patient-derived iPSC lines displayed Aβ accumulation [8]. Using a slightly modified protocol, Oksanen *et al.* generated astrocytes using iPSC lines from three fAD patients with *PSEN1* exon 9 deletion [26]. The astrocyte cultures from fAD patients showed elevated secretion of Aβ(1–42) , while Aβ(1–40) secretion was unaltered in comparison to astrocyte cultures obtained from healthy control lines or gene-corrected isogenic controls. In addition, the iPSC-derived astrocytes from fAD patients showed increased oxidative stress and altered metabolism. Finally, Jones *et al.* described a protocol that can generate enriched populations of mature cortical astrocytes (>95%), within 30 days of induction starting from NPCs [27]. Aβ levels were not investigated in this study, however, iPSC-derived astrocytes from both sAD and fAD patients displayed altered morphological appearance, aberrant localization of astrocyte markers and an overall atrophic profiles.

Differentiation of neurons, astrocytes and microglia from patient-derived iPSC offers a powerful toolbox for studying sAD. Nevertheless, *in vitro* cultures of these cell types often lack essential aspects from their natural environment including cell-cell contact between multiple cell types, which negatively impacts on the maturity and functionality of these cells. To overcome these limitations researchers have been largely using two strategies; growing iPSC-derived cells in 3 dimensional (3D) structures, including spheres and organoid models or by transplanting iPSC-derived cells into the brain of mice. For instance, astrocyte maturation can be greatly improved by cultivating iPSC-derived astrocytes in 3D human cortical spheroids (hCSs) [28] in direct contact with neurons for prolonged periods of time [29]. Astrocytes in long-term cultures (over 20 months) display gene expression patterns and functional properties of matured astrocytes [29]. Also microglia function can be enhanced by allowing interactions from a 3D brain environment (reviewed in [30]). For example, Abud and coworkers showed that culturing human iPSC-derived microglia (iMGs) on hiPSC 3D brain organoids (BORGs) results in migration and engraftment of the iMGs [14]. Integrated iMGs display a more mature and ramified morphology and respond to injury similar as microglia in brain tissue. Trujillo *et al.* describe the generation of functional cortical organoids from NPCs [31]. By applying single-cell transcriptional profiling and functional validation over a 10-month period of time, the authors could demonstrate that long-term development of cortical organoids is a highly dynamic process involving a range of different cell types. In addition to cellular maturation, the cortical organoids displayed increased electrophysiological activity. Eventually the organoids displayed electrophysiological network activity that was reminiscent of the spontaneous and synchronized activity in developing human neocortex [31].

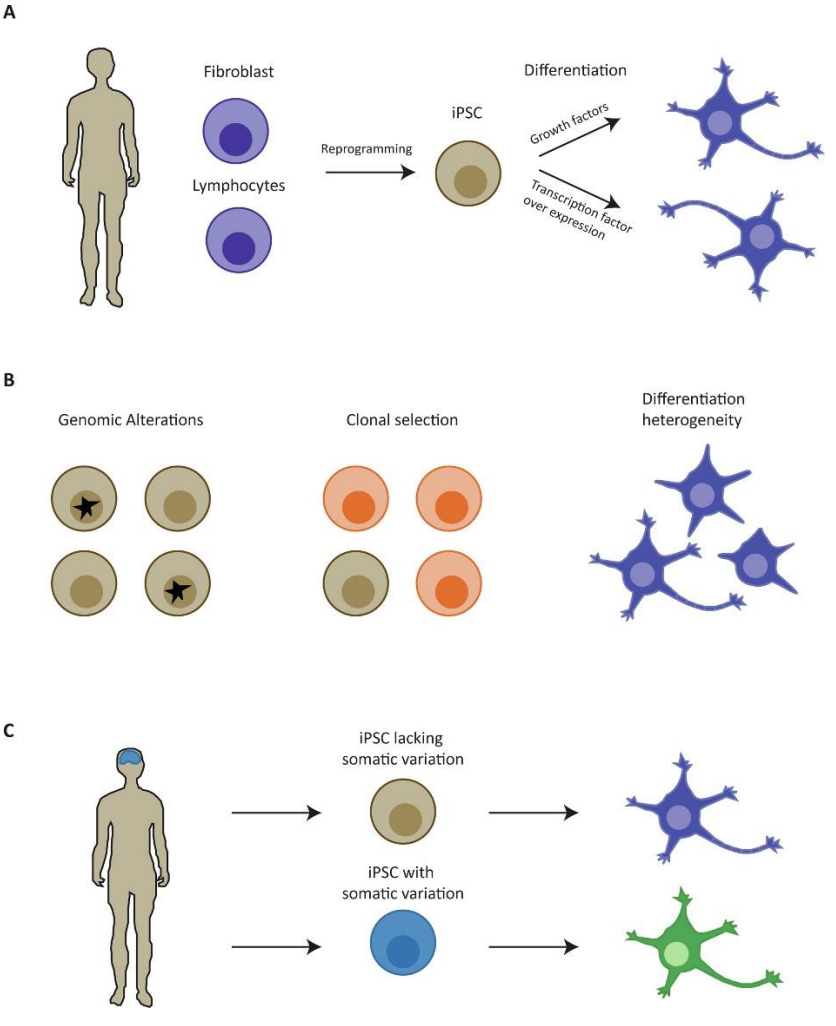


Figure 1. Generation of human induced pluripotent stem cell (iPSC)-derived cell models for sporadic Alzheimer's disease (sAD). **A.** Somatic cells can be reprogrammed to iPSCs. Skin fibroblasts or peripheral blood lymphocytes (PBLs) are frequently used as source of somatic cells. Final neuronal cell types of interest can be generated through directed differentiation of iPSCs using signaling molecules or by exogenous expression of lineage-specific transcription factors. **B.** Potential sources of variability in iPSC-derived cell models. Reprogramming can induce genomic alterations that might stay undetected in case of small alterations (Left panel). Human somatic cells might not be as equipotent as previously assumed. Selection of superior iPSC clones can bias the ultimate disease model (middle panel). Long and complex differentiation schemes can influence robustness of the ultimate disease model. Variability can occur in terms of the maturity of the desired cellular phenotype, the percentage of cells that obtained the correct phenotype and the percentage of unwanted phenotypes created during differentiation (left panel). **C.** Somatic mosaicism is frequently detected in brains of patients with neurological diseases like AD (indicated in blue). Generated iPSC models from fibroblasts or PBLs might lack these unique genotypes that could be essential for an accurate sAD iPSC-derived cell model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

These examples illustrate how the 3D microenvironment in organoid structures can affect maturity and functionality of cells of the CNS. For a more comprehensive review on brain organoids, please refer to [32].

A powerful alternative way to grow iPSC-derived cells within the context of a CNS microenvironment relies on transplanting cells into the brain of mice. For example, Espuny-Camacho *et al.* transplanted human iPSC-derived cortical neuronal precursors into the brain of a murine AD model and could show that these precursor cells differentiate into mature human neurons that are integrated into the murine brain [33]. These neurons express 3R/4R tau splice forms, display aberrant changes in phosphorylation and conformational of tau, and undergo neurodegeneration. Hasselmann and colleagues, transplanted hiPSC-derived hematopoietic progenitor cells into the postnatal brain of humanized, immune-compromised mice [34]. This resulted in context-dependent differentiation of the HPCs into microglia and other CNS macrophages. Transcriptome analysis revealed a high degree of clustering between human microglia obtained through transplantation in mice and ex vivo human microglia. This suggest that the murine brain environment promotes differentiation of HPCs into human microglia. Human microglia surrounding Aβ plaques in transplanted mice display robust transcriptional response that are partly distinct from murine microglia. Linaro *et al.* found that cortical pyramidal neurons derived from human ESC line H9 can integrate as single cells in neonatal mouse cortex upon transplantation [35]. In particular, the transplanted neurons displayed robust maturation at the single-cell level and strong interactions with the host neurons. Moreover, sensory stimulation resulted in physiological responses with good resemblance to that of native cortical neurons.

All together, these studies show great potential for chimeric xenograft models for studying diseases like sAD. Transplation of patient-derived and/ or genetically modified cell types and assessing their *in vivo* function in a natural microenvironment may lead to a more complete understanding of sAD.

Sources of variability in iPSC-derived neural models for sAD

The ability to generate iPSC-derived cell models of sAD patients harbors great potential for studying the etiology of sAD as well as for the development of biomarkers or therapeutic targets for sAD (Figure 1A). More specifically, this platform provides a valuable tool in exploring the complex heterogeneous nature in the etiology of sAD through the interrogation of functional effects of genetic, epigenetic and transcriptional variants linked to risk, as well as protective environmental factors. However, there are several issues associated with iPSC models that require further investigation before the actual potential of this

technology in the context of sAD becomes more evident. Here we will discuss for several of these issues, and how these could impact the interpretation of the findings in relation to sAD.

iPSC generation and differentiation

Generation of iPSCs from sAD patients and differentiation into disease relevant neural lineages with high reproducibility is a key factor for developing robust sAD models. However, iPSC models and derived neurons have their limitations and using them to study sAD is particularly challenging. First of all, it is well known that reprogramming of somatic cells into iPSC can give rise to random genomic alterations including copy number variants, indels and karyotypic aberations that were not present in the parental somatic cells (Figure 1B) (reviewed in [36]). While it is relatively easy to select iPSC lines with normal karyotypes it is less straightforward to avoid smaller rearrangements or indels. Evidently, any alteration that is not present in somatic cells is undesirable and could potentially affect study outcome.

Second, evidence supports that not all iPSC lines have equal potency to differentiate into the desired mature phenotype (Figure 1B) [37, 38]. Hu *et al.* report reduced and more variable neural differentiation efficiencies using a panel of human iPSC lines [37]. This was not related to technical or cell culture related issues since various human embryonic stem cells used in parallel revealed robust and efficient differentiation. All tested iPSC lines in this study were generated by reprogramming of fibroblasts. In addition, the tissue of origin from which iPSC lines are derived can also affect differentiation potential. Profound differences in differentiation efficiency were observed for iPSC lines that originated from umbilical cord blood cells (CB-iPSCs) and foreskin keratinocytes (K-iPSCs) [38]. While CB-iPSCs were more efficient in generating hematopoietic cell types than K-iPSCs, K-iPSCs differentiated more efficiently into keratinocytes compared to CB-iPSCs. This study suggests that iPSCs can retain epigenetic marks from the original tissue the iPSCs were derived from. It will therefore be necessary to assess how variability between different neuronal lines derived from donor cells of distinct developmental lineages could affect the interpretation of disease phenotypes that might be observed in sAD lines.

Third, reprogramming somatic cells into iPSCs has been associated with loss of aging traits such as certain epigenetic chromatin marks, telomerase shortening and altered mitochondrial metabolism [39, 40]. These aging hallmarks are likely of high relevance for the development of an age-related disease like sAD. Certain phenotypes associated with sAD might not be discovered in the absence of aging. Brain cells derived from iPSCs might therefore not be the most optimal model for

studying sAD, but this requires further investigation. Efforts to create neurons that have preserved the aging signature have led to direct reprogramming of fibroblasts into neurons. However, it is unclear how other neural traits are affected by direct reprogramming or what the impact of direct reprogramming is on genomic integrity. In addition, limited neurons can be produced since there's no cell with self-renewing capacity [40, 41]. Whether neurons with a preserved aging signature presents a better *in vitro* cell model for studying sAD compared to iPSC-derived neurons remains to be investigated. For this reason, a wealth of research in currently focused on developing physiologically complete models of accelerating aging and defining methods for the concept of age-preservation during reprogramming. In this regard, one could also think of age accelerating compounds, prolonged cultures times or the production of cellular intermediates that allow expansion before differentiation, while maintaining their aging hallmarks [42]. As such, these efforts might provide a suitable solution to this issue.

Fourth, aside from the fact that the current differentiation protocols remain limited in the generation of all neural cell types found in the brain (reviewed in [6]), great variability in the final neuronal phenotype is introduced by the extents and complexity of the differentiation protocols themselves (Figure 1B). A variety of protocols have been developed to generate neuronal cells from iPSCs (reviewed in [6, 43]) and can be obtained by directed differentiation using signaling molecules or by artificial overexpression of lineage-specific transcription factors. The lack of consensus in differentiation methodology complicates study comparisons. However, this does not account for all variability that is introduced by culturing practices of iPSC models. A multicenter study investigated the transcriptome and proteome of two human iPSC lines at two time points during differentiation into cortical projection neurons using the same standard operating procedure in three independent inductions [44]. The authors report poor reproducibility in differential gene expression signatures between these two lines across the five test sites. These results illustrate that even when intra-laboratory variability is acceptable, huge inter-laboratory variability can exist. Despite the use of a standard operating procedure and exactly the same iPSC lines, very subtle differences in culturing conditions between laboratories can lead to large differences in phenotypic outcome.

Control cells

A completely distinct problem that complicates interpretation of studies using iPSC sAD lines besides the issues already described above lies in the identification of adequate control cell lines. It is not trivial to define a healthy or non-diseased control. Most studies utilizing iPSC-derived disease models apply

iPSC lines from unrelated healthy donors as controls, since isogenic non-diseased controls are not available or impossible to produce. A genome-wide association studies performed by the 1000 genomes project consortium revealed that a set of control cell lines obtained from multiple individuals contained 50 to 100 disease-associated genetic variants [45]. This illustrates the problem we face in finding good reference models. Other initiatives have shown that the most prominent source of variability in a reprogrammed iPSC is related to its own genetic background [46-50]. Studying a sufficiently large reference panel may help to reduce the noise that results from the heterogeneity in the genetic background from individuals. Ultimately, this will be beneficial for the discovery of subtle, but genuine disease causing factors.

Somatic mosaicism

In the recent years, multiple studies have found proof for the presence of *de novo* somatic mutations in neurological disease (reviewed in [51]). Somatic mutations can arise postzygotically during embryonic development or during postnatal life. The impact of this so-called somatic mosaicism is depending on the time in life and the cell type in which these somatic mutations occur. Somatic mutations that take place early in life will be present in the majority of cells in the human body. Whereas somatic mutations that occur late will only be present in a small subset of cells and may be restricted to a single tissue. It has been found that somatic mutations with alternate allele frequencies as little as 1% can cause disease [52]. If indeed somatic mutations in the human brain play an important role in onset and progression of sAD than it is important that the iPSC also carry these somatic mutations. In a study by Bushman and coworkers it was reported that neurons isolated from sAD patients displayed increases in DNA contents and elevated *APP* gene copy numbers (up to 12 copies per neuron) that were not detectable in lymphocytes from the same patients [53]. Later the same group found that *APP* mosaicism in the brains of sAD patients was caused by integration of *APP* genomic complementary DNAs into the genome [54]. Both wild-type sequences as well as smaller, exon lacking variants were found to be incorporated in the genome of sAD neurons. Some of these *APP* mRNA variants encoded proteins can cause neurotoxicity. It remains to be investigated whether *APP* mosaicism in the brains of sAD patients is a cause or an effect of AD.

However, in order to generate adequate disease models for sAD it is important to determine whether somatic mosaicism in the brains of sAD patients plays a causal role or whether this occurs as consequence of AD. In case that somatic mosaicism is driving sAD development then the iPSC-derived disease model should also carry this genetic variation. If somatic mosaicism occurs as a result of AD then the genetic variation is not essential in the iPSC-derived disease

model (Figure 1C). Whether it is possible to generate an iPSC line bearing such a somatic mosaicism is highly questionable for several reasons. First, the somatic mosaicism needs to be present in the tissue from which iPSC are generated. This can be problematic, as the study by Bushman and coworkers showed that somatic variants in the brain were absent in lymphocytes [53]. Second, such relatively large genetic alteration may be tolerable in post-mitotic neurons, but is likely not compatible with cell proliferation.

Generating iPSCs from patients with somatic mosaicism can in certain situation also be advantageous. In cases where the disease causing somatic mutations are known, iPSC lines can be generated for mutant and wild-type from the same patient [55]. This offers the unique opportunity to generate an isogenic pair of iPSC clones, thereby reducing some of the noise associated with generic iPSC control lines.

Reprogramming capacity

For a long time it was thought that all individual somatic cells had the same reprogramming capacity or so called clonal equipotency [56]. However, a recent study revealed that a small subset of “elite” clones dominates reprogramming of somatic mouse cells intro pluripotent stem cells [57] (Figure 1B). Expression of Wnt1 was identified as a key hallmark of elite cells. If this phenomenon also applies for humans, than this will have important consequences for the iPSC-derived cell models generated with these selected “elite” clones. This raises the question to what degree the selected “elite” clones can capture the disease-associated genotype and thus how representative these models are for the disease. Potentially, somatic mosaicism could affect reprogramming capacity, and lead to counter selection against the desired iPSC genotype. Worst-case scenario is that the disease-associated genotype cannot be captured by reprogramming somatic cells into iPSCs. This is crucially important for accurate disease modelling and drug screening studies.

Conclusions and future perspectives

The development of iPSC technology has great potential for studying complex multifactorial diseases like sAD. Although hallmarks of AD can be detected in iPSC-derived cell models, there is a high degree of variability in terms of their presence and severity between different sAD lines. This variability can mask important, but subtle genetic differences between individuals. As the number of studies using iPSC-derived sAD models increases, it will be important to ensure highly reproducible and informative data. The compilation of a large reference panel representing non-demented control subjects will be instrumental for adequate interpretation of genomic variants found in sAD patients. It will be

important to better understand clonal selection during reprogramming, and the relevance of somatic mosaicism and aging in sAD, to ensure representative iPSCs-derived cell models of sAD.

In order to increase translatability of pre-clinical findings it will be important to continue to develop and improve sAD models that reflect the brains natural environment. Sophisticated iPSC-derived cell models should consist of multiple cell types, in 3D conformations allowing cell-cell interactions. Further development of more accurate models for studying sAD like long-term organoid cultures and chimeric xenograft models will be essential in the coming years.

While it is well established that genetic alterations contribute to the pathophysiology of sAD, detailed multi-omics characterization of iPSC-based models and the role of environmental factors associated with sAD pathophysiology remain currently underexplored. It is recognized that in the aging brain, environmental risk factors together with complex gene-environment interactions, are playing crucial roles in reinforcing sAD pathogenesis. Recent studies have shown that exposure to these risk factors can bring about epigenomic, transcriptomic, proteomic and metabolomic changes related to AD, which can bring about sustained alterations in molecular processes leading into the manifestation of the full-blown disease [58]. Furthermore, many AD researchers tend to suggest that these environmental risk factors operate during the pre-clinical phase of AD, even decades before the appearance of the first clinical symptoms. In either way, there is overwhelming support for environmental and extra-genetic risk factors as inducers of sAD pathogenesis. Nevertheless, the exact molecular nature of these interactions, as well as their temporal relationship with the development of the disease, remain largely unknown. Further research from experimental and epidemiological studies that focus on the interaction of these factors with the aforementioned multi-omics modalities therefore remains necessary to develop a better understanding about their contribution to the course and development of the disease. In this regard, iPSC-derived systems might provide exciting opportunities for modelling these interactions, by allowing one to expose cells derived from healthy individuals and sAD patients with different genetic backgrounds to environmental insults and study their cellular and molecular responses in an in vitro setting.

Acknowledgments
This research was made possible by BReIN (Brightlands e-infrastructure for Neurohealth), an initiative which is co-funded by the Province of Limburg, Maastricht University and Maastricht University Medical Centre + in the Netherlands.

References

1. Clay E, Zhou J, Yi Z-M, Zhai S, Toumi M. Economic burden for Alzheimer's disease in China from 2010 to 2050: a modelling study. Journal of Market Access & Health Policy. 2019;7(1):1667195.
2. Chakrabarti S, Khemka VK, Banerjee A, Chatterjee G, Ganguly A, Biswas A. Metabolic risk factors of sporadic Alzheimer's disease: implications in the pathology, pathogenesis and treatment. Aging and disease. 2015;6(4):282.
3. Laurijssens B, Aujard F, Rahman A. Animal models of Alzheimer's disease and drug development. Drug Discovery Today: Technologies. 2013;10(3):e319-e27.
4. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell. 2006;126(4):663-76.
5. Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nature reviews Drug discovery. 2017;16(2):115-30.
6. Riemens RJ, van den Hove DL, Esteller M, Delgado-Morales R. Directing neuronal cell fate in vitro: achievements and challenges. Progress in neurobiology. 2018;168:42-68.
7. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012;482(7384):216-20.
8. Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, et al. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. Cell Stem Cell. 2013;12(4):487-96.
9. Foveau B, Correia AS, Hébert SS, Rainone S, Potvin O, Kergoat M-J, et al. Stem cell-derived neurons as cellular models of sporadic Alzheimer's disease. Journal of Alzheimer's Disease. 2019;67(3):893-910.
10. Duan L, Bhattacharyya BJ, Belmadani A, Pan L, Miller RJ, Kessler JA. Stem cell derived basal forebrain cholinergic neurons from Alzheimer's disease patients are more susceptible to cell death. Mol Neurodegener. 2014;9:3.
11. Ochalek A, Mihalik B, Avci HX, Chandrasekaran A, Téglási A, Bock I, et al. Neurons derived from sporadic Alzheimer's disease iPSCs reveal elevated TAU hyperphosphorylation, increased amyloid levels, and GSK3B activation. Alzheimer's research & therapy. 2017;9(1):1-19.
12. Young JE, Boulanger-Weill J, Williams DA, Woodruff G, Buen F, Revilla AC, et al. Elucidating molecular phenotypes caused by the SORL1 Alzheimer's disease genetic risk factor using human induced pluripotent stem cells. Cell Stem Cell. 2015;16(4):373-85.
13. Hossini AM, Megges M, Prigione A, Lichtner B, Toliat MR, Wruck W, et al. Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks. BMC Genomics. 2015;16:84.
14. Abud EM, Ramirez RN, Martinez ES, Healy LM, Nguyen CH, Newman SA, et al. iPSC-derived human microglia-like cells to study neurological diseases. Neuron. 2017;94(2):278-93. e9.
15. Haenseler W, Sansom SN, Buchrieser J, Newey SE, Moore CS, Nicholls FJ, et al. A highly efficient human pluripotent stem cell microglia model displays a neuronal-co-culture-specific expression profile and inflammatory response. Stem cell reports. 2017;8(6):1727-42.
16. Muffat J, Li Y, Yuan B, Mitalipova M, Omer A, Corcoran S, et al. Efficient derivation of microglia-like cells from human pluripotent stem cells. Nature medicine. 2016;22(11):1358-67.
17. Pandya H, Shen MJ, Ichikawa DM, Sedlock AB, Choi Y, Johnson KR, et al. Differentiation of human and murine induced pluripotent stem cells to microglia-like cells. Nature neuroscience. 2017;20(5):753.

18. Friedman BA, Srinivasan K, Ayalon G, Meilandt WJ, Lin H, Huntley MA, et al. Diverse brain myeloid expression profiles reveal distinct microglial activation states and aspects of Alzheimer's disease not evident in mouse models. Cell reports. 2018;22(3):832-47.
19. Geirsdottir L, David E, Keren-Shaul H, Weiner A, Bohlen SC, Neuber J, et al. Cross-species single-cell analysis reveals divergence of the primate microglia program. Cell. 2019;179(7):1609-22. e16.
20. Lin Y-T, Seo J, Gao F, Feldman HM, Wen H-L, Penney J, et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. Neuron. 2018;98(6):1141-54. e7.
21. Guerreiro R, Bilgic B, Guven G, Brás J, Rohrer J, Lohmann E, et al. A novel compound heterozygous mutation in TREM2 found in a Turkish frontotemporal dementia-like family. Neurobiology of aging. 2013;34(12):2890. e1-. e5.
22. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. New England Journal of Medicine. 2013;368(2):107-16.
23. Xiang X, Piers TM, Wefers B, Zhu K, Mallach A, Brunner B, et al. The Trem2 R47H Alzheimer's risk variant impairs splicing and reduces Trem2 mRNA and protein in mice but not in humans. Molecular neurodegeneration. 2018;13(1):49.
24. Claes C, Van Den Daele J, Boon R, Schouteden S, Colombo A, Monasor LS, et al. Human stem cell–derived monocytes and microglia-like cells reveal impaired amyloid plaque clearance upon heterozygous or homozygous loss of TREM2. Alzheimer's & Dementia. 2019;15(3):453-64.
25. Krencik R, Zhang S-C. Directed differentiation of functional astroglial subtypes from human pluripotent stem cells. Nature protocols. 2011;6(11):1710-7.
26. Oksanen M, Petersen AJ, Naumenko N, Puttonen K, Lehtonen Š, Olivé MG, et al. PSEN1 mutant iPSC-derived model reveals severe astrocyte pathology in Alzheimer's disease. Stem cell reports. 2017;9(6):1885-97.
27. Jones VC, Atkinson-Dell R, Verkhratsky A, Mohamet L. Aberrant iPSC-derived human astrocytes in Alzheimer's disease. Cell death & disease. 2017;8(3):e2696-e.
28. Paşca AM, Sloan SA, Clarke LE, Tian Y, Makinson CD, Huber N, et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. Nature methods. 2015;12(7):671-8.
29. Sloan SA, Darmanis S, Huber N, Khan TA, Birey F, Caneda C, et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. Neuron. 2017;95(4):779-90. e6.
30. Claes C, Van den Daele J, Verfaillie CM. Generating tissue-resident macrophages from pluripotent stem cells: Lessons learned from microglia. Cellular Immunology. 2018;330:60-7.
31. Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, et al. Complex oscillatory waves emerging from cortical organoids model early human brain network development. Cell stem cell. 2019;25(4):558-69. e7.
32. Benito-Kwiecinski S, Lancaster MA. Brain organoids: Human neurodevelopment in a dish. Cold Spring Harbor perspectives in biology. 2020;12(8):a035709.
33. Espuny-Camacho I, Arranz AM, Fiers M, Snellinx A, Ando K, Munck S, et al. Hallmarks of Alzheimer's disease in stem-cell-derived human neurons transplanted into mouse brain. Neuron. 2017;93(5):1066-81. e8.
34. Hasselmann J, Coburn MA, England W, Velez DXF, Shabestari SK, Tu CH, et al. Development of a chimeric model to study and manipulate human microglia in vivo. Neuron. 2019;103(6):1016-33. e10.
35. Linaro D, Vermaercke B, Iwata R, Ramaswamy A, Libé-Philippot B, Boubakar L, et al. Xenotransplanted human cortical neurons reveal species-specific development and functional integration into mouse visual circuits. Neuron. 2019;104(5):972-86. e6.

36. Assou S, Bouckenheimer J, De Vos J. Concise review: assessing the genome integrity of human induced pluripotent stem cells: what quality control metrics? Stem Cells. 2018;36(6):814-21.
37. Hu B-Y, Weick JP, Yu J, Ma L-X, Zhang X-Q, Thomson JA, et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proceedings of the National Academy of Sciences. 2010;107(9):4335-40.
38. Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, et al. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nature biotechnology. 2011;29(12):1117-9.
39. Lapasset L, Milhavet O, Prieur A, Besnard E, Babled A, Aït-Hamou N, et al. Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. Genes & development. 2011;25(21):2248-53.
40. Mertens J, Paquola ACM, Ku M, Hatch E, Bohnke L, Ladjevardi S, et al. Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell. 2015;17(6):705-18.
41. Hu W, Qiu B, Guan W, Wang Q, Wang M, Li W, et al. Direct conversion of normal and Alzheimer's disease human fibroblasts into neuronal cells by small molecules. Cell stem cell. 2015;17(2):204-12.
42. Mertens J, Reid D, Lau S, Kim Y, Gage FH. Aging in a dish: iPSC-derived and directly induced neurons for studying brain aging and age-related neurodegenerative diseases. Annual review of genetics. 2018;52:271-93.
43. McCaughey-Chapman A, Connor B. Human cortical neuron generation using cell reprogramming: a review of recent advances. Stem cells and development. 2018;27(24):1674-92.
44. Volpato V, Smith J, Sandor C, Ried JS, Baud A, Handel A, et al. Reproducibility of molecular phenotypes after long-term differentiation to human iPSC-derived neurons: a multi-site omics study. Stem cell reports. 2018;11(4):897-911.
45. Consortium GP. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061.
46. Carcamo-Orive I, Hoffman GE, Cundiff P, Beckmann ND, D'Souza SL, Knowles JW, et al. Analysis of transcriptional variability in a large human iPSC library reveals genetic and non-genetic determinants of heterogeneity. Cell stem cell. 2017;20(4):518-32. e9.
47. Warren CR, O'Sullivan JF, Friesen M, Becker CE, Zhang X, Liu P, et al. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. Cell Stem Cell. 2017;20(4):547-57. e7.
48. Pashos EE, Park Y, Wang X, Raghavan A, Yang W, Abbey D, et al. Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. Cell stem cell. 2017;20(4):558-70. e10.
49. DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, Olson KM, et al. Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells. Cell stem cell. 2017;20(4):533-46. e7.
50. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature. 2017;546(7658):370-5.
51. D'Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. Nature neuroscience. 2018;21(11):1504-14.
52. D'Gama AM, Woodworth MB, Hossain AA, Bizzotto S, Hatem NE, LaCoursiere CM, et al. Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. Cell reports. 2017;21(13):3754-66.
53. Bushman DM, Kaeser GE, Siddoway B, Westra JW, Rivera RR, Rehen SK, et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. Elife. 2015;4:e05116.

54. Lee M-H, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. Nature. 2018;563(7733):639-45.
55. van der Wal E, Den Hamer B, van der Vliet PJ, Tok M, Brands T, Eussen B, et al. Generation of genetically matched hiPSC lines from two mosaic facioscapulohumeral dystrophy type 1 patients. Stem cell research. 2019;40:101560.
56. Hanna J, Saha K, Pando B, Van Zon J, Lengner CJ, Creyghton MP, et al. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. 2009;462(7273):595-601.
57. Shakiba N, Fahmy A, Jayakumaran G, McGibbon S, David L, Trcka D, et al. Cell competition during reprogramming gives rise to dominant clones. Science. 2019;364(6438).
58. Lahiri DK, Maloney B, Zawia NH. The LEARn model: an epigenetic explanation for idiopathic neurobiological diseases. Molecular psychiatry. 2009;14(11):992-1003.

Chapter 7

Directing neuronal cell fate in vitro: Achievements and challenges

Renzo J.M. Riemens[1,2,3], Daniël L.A. van den Hove[3,4] Manel Esteller[1,5,6], Raul Delgado-Morales[1,3]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute – IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain.
[6]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

Progress in Neurobiology. 2018 Sep;168:42-68.
Doi: 10.1016/j.pneurobio.2018.04.003.

Abstract
Human pluripotent stem cell (PSC) technology and direct somatic cell reprogramming have opened up a promising new avenue in the field of neuroscience. These recent advances allow researchers to obtain virtually any cell type found in the human brain, making it possible to produce and study functional neurons in laboratory conditions for both scientific and medical purposes. Although distinct approaches have shown to be successful in directing neuronal cell fate in vitro, their refinement and optimization, as well as the search for alternative approaches, remains necessary to help realize the full potential of the eventually derived neuronal populations. Furthermore, we are currently limited in the number of neuronal subtypes whose induction is fully established, and different cultivation protocols for each subtype exist, making it challenging to increase the reproducibility and decrease the variances that are observed between different protocols. In this review, we summarize the progress that has been made in generating various neuronal subtypes from PSCs and somatic cells, with special emphasis on chemically defined systems, transcription factor-mediated reprogramming and epigenetic-based approaches. We also discuss the efforts that are being made to increase the efficiency of current protocols and address the potential for the use of these cells in disease modelling, drug discovery and regenerative medicine.

Keywords
Brain disorders, cellular reprogramming, directed differentiation, disease modelling, epigenetics, neuronal differentiation, pluripotent stem cells, somatic cells, transdifferentiation.

Highlights

- Directed differentiation and reprogramming of stem cells and somatic cells allow induction of human neurons in vitro.
- Chemical stimulation for in vitro differentiation recapitulates the neurodevelopmental patterning principles in the embryo.
- Overexpression of transcription factors induces dramatic changes in transcriptional networks that drive cellular conversions.
- Epigenetic manipulation allows induction of profound epigenome remodelling processes that underlie cell fate switches.
- The availability of human neurons in vitro has critical implications for future brain-related studies and drug development.

Introduction

Over the last decades, our general knowledge on human brain functions has grown exceedingly thanks to the availability of animal models and human brain tissue. Although their utility is undeniable, both are challenged with limitations that have been impeding progress in gaining complete mechanistic insights, as well as in the development of therapeutic interventions for many brain disorders [1]. The numerous transgenic animal models that have been established mimic pathological mechanisms of the human brain to some extent, but they do not yet satisfactorily capture human disease phenotypes completely [1]. An animal is obviously not a human being and interspecies differences might, therefore, be critical factors underlying the failure of translating a wealth of preclinical findings into clinical implementations [2, 3]. Aside from the distinct (epi)genetic backgrounds, there are major physiological differences that could affect the development of disease phenotypes or differentially affect drug mechanisms, leading to misinterpretation of experimental findings. Human brain tissue samples, on the other hand, can generally only be obtained post-mortem, which complicates the study of disease aetiology and progression, since they simply do not allow to discriminate between cause and consequence of the disorder [4]. Cellular *in vitro* model systems, however, have the potential to overcome this latter challenge, owing to the possibility of manipulating the (epi)genetic architecture, as well as the environmental exposome, in culture conditions, which allows to study cause-effect relationships of pathological hallmarks in a controlled setting. Unfortunately, many early *in vitro* model systems have heavily relied on a combination of non-neuronal human cell lines, primary rodent cultures and transgenic rodent cell lines, which, similar to animal models, also all exclude the human neuronal (epi)genomic and phenotypic context [5].

In view of the aforementioned limitations and the unmet necessity of finding therapeutic interventions for brain disorders, the recent availability of human pluripotent stem cell (PSC) technology and somatic cell reprogramming has offered new opportunities for human brain-related studies [6, 7]. PSC technology is an umbrella term that encompasses both embryonic stem cells (ESCs) and induced PSCs (iPSCs) [8]. Aside from being derived from humans, the distinct advantages that these cells have are their unlimited proliferative capacity and the ability to differentiate towards virtually any cell type, including specific neuronal subtypes [9]. These characteristics equip PSCs with the unique feature to provide a theoretically inexhaustible and replenishable source of cells *in vitro*. Somatic cells on the other hand, albeit restricted by their limited proliferative potential, can be obtained relatively easy from healthy individuals and patients, and can then be directly reprogrammed towards desirable neuronal subtypes in a comparative, but faster framework [10]. Consequently, PSCs, somatic cells and their

differentiated progeny can nowadays be used to model disease mechanisms in a humanized-setting, where they allow for the investigation of unique human cellular and molecular features in a cell-specific and personalised matter [11]. In fact, we are currently in the middle of an exciting era where human PSC and somatic cell reprogramming studies are contributing to the understanding of underlying neurobiological processes, as well as the consequences of personal molecular variations on the development and course of brain disorders [10, 12]. In addition, iPSC- and somatic cell-derived neuronal populations provide a platform for high-throughput drug screening and toxicity testing in an upcoming epoch of personalised medicine, which assists the production of therapeutic interventions and might at the same time provide cues for therapeutic resistance [11]. Furthermore, such neuronal populations also harbour therapeutic potential in the field of regenerative medicine, since they might be used for transplant therapies [11].

Given the heterogeneity of neuronal cells found in the human brain and the complex interactions between them, one of the main opportunities PSC technology, directed differentiation and somatic cell reprogramming are offering, is the study of multicellular neural cultures organized in a manner reminiscent to what is seen in distinct anatomical structures of the brain. Such a method offers a promising approach to study higher-order neuronal networks during development and allows studying single neuron connectivity, as well as various other cellular interactions found in the neuronal niche, such as those between glial cells and endothelial cells [13, 14]. Another major focus in the scientific community that emerges from the unique contributions each of the neuronal subtypes have in specific brain functions and disease, has been to obtain and study pure populations of neurons free of other subtypes. Derivation of such homogenous populations is especially important for the study of disease-associated neurons, where confounding effects of other subtypes should be avoided, such as in transcriptomic- and epigenetic profiling for example. Furthermore, the acquisition of protocols that can produce highly pure populations of neurons can in turn be utilized to customize the features of a multicellular culture. Successfully directing differentiation and cellular reprogramming *in vitro* has, therefore, been an area of intense research during the past two decades [15]. Many strategies have been explored utilizing mouse and human PSCs, as well as somatic cells, to generate anatomically specified neural precursor cells (NPCs) and differentiated neuronal subtypes (See Fig. 1). As a definition used here, NPCs refer to the mixed population of cells consisting of all undifferentiated progeny from neural stem cells, therefore including both neural stem cells and neural progenitor cells. Altogether, these differentiation methods try to recapitulate the multistep processes of neural development that

occur in the early embryo [15]. Thus, an improved understanding of developmental signalling pathways and gene regulatory networks has guided the design of neuronal differentiation and cellular reprogramming strategies [8]. Although significant advances have been made, *in vitro* neuronal differentiation is actually not a process that is fully disciplined yet, as very often other cell types are produced in parallel to what was first intended by the researcher. Furthermore, differentiation protocols for multiple neuronal subtypes remain either unestablished or unstable, or they do not allow to obtain the terminally differentiated neuron with its entire functional maturated characteristics [16, 17].

To date, the majority of directed neuronal differentiation and direct somatic cell reprogramming protocols involve chemical stimulation through patterning cues or the use of ectopic overexpression of lineagespecific transcription factors that are known for their genuine involvement during neural development [18-20]. Although it is currently accepted that these networks of patterning molecules and transcription factors orchestrate neuronal induction and differentiation, it is becoming more and more evident that cell-intrinsic mechanisms are also key players within these circuits [21-23]. Neuronal induction and differentiation generally rely on the interplay of activation and inhibition on multiple developmental signalling pathways tightly controlled by the epigenetic machinery [8, 24]. Within these networks, the epigenetic machinery is essential for fine-tuning genetic programs that coordinate distinct developmental processes, as well as shaping the neuronal identities at a phenotypic resolution [25, 26]. With the current advances in epigenetic editing [27-29] and RNA interference (RNAi) [30, 31], an alternative strategy for directed neuronal differentiation and direct somatic cell reprogramming has become available. Such approaches allow the design and construction of novel specific artificial epigenetic pathways or the redesign of existing endogenous molecular systems, in order to intentionally change epigenetic information at desired loci [32]. Those systems can now be used as an additional tool to guide neuronal cell fate *in vitro* and also allow to address fundamental questions concerning the role of epigenetics in assigning neuronal cell fates.

In the present review, we will provide an overview on the progress made in generating various neuronal subtypes from PSCs and somatic cells *in vitro*. We will dissect the current used chemically defined systems, transcription factor-mediated reprogramming methods and epigenetic-based approaches. Furthermore, we will discuss the efforts that are being made to increase the efficiency of current protocols and highlight the potential for the use of this platform in disease modelling, drug discovery and regenerative medicine.

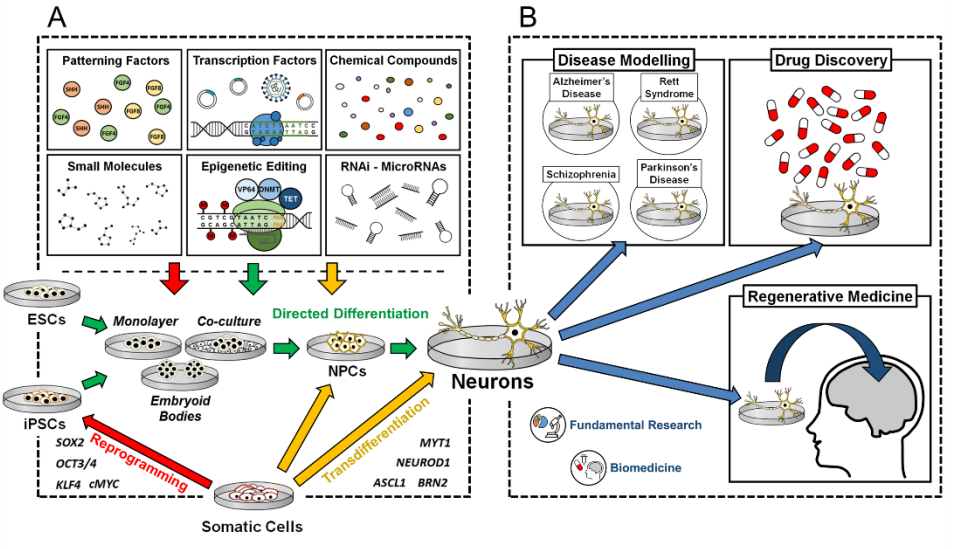


Figure 1. Overview of methods and applications for neurons obtained by directed differentiation and (direct) cellular reprogramming in embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs) and somatic cells. **A.** The figure shows the complete directed differentiation and (direct) cellular reprogramming toolkit currently available, including the use of patterning factors, chemical compounds, small molecules, transcription factors, epigenetic editing and RNA interference (RNAi) with the use of microRNAs (miRNAs) as example. The neurodevelopmental transitions starting from PSCs towards neural precursor cells (NPCs) and eventually mature neurons are depicted. Directed differentiation protocols can generally be divided into three methods: monolayer methods, co-culture methods and embryoid body (EB) methods. Reprogramming of somatic cells towards iPSCs can be achieved by increased expression of transcription factors, such as Oct3/4, Sox2, Klf4 and Myc (OSKM, Yamanaka factors). Direct reprogramming of somatic cells towards specific neuronal subtypes can be achieved by co-expression of the pan-neuronal factors Brn2, Ascl1, Myt1 (BAM) and Neurod1 among others, and neuronal subtype-specific transcription factors. **B.** The direct applications of the obtained neurons in fundamental research and biomedicine, including the use for disease modelling, drug discovery and regenerative medicine, are shown.

Chemically defined systems

Neural induction and differentiation

Research to develop protocols for the differentiation of PSCs into clinically relevant cell subtypes has progressed at a rapid pace. Hundreds of protocols have become available that allow to obtain early NPCs and eventually the derivation of desired neuronal populations [33-38]. In the classical embryoid body (EB) method, scientists let PSC form self-assembling aggregates from dissociated suspension cultures, known as EBs [39, 40]. This method initially aimed to spontaneously differentiate the PSCs into mixed cellular populations encompassing the three germ-layers [41]. As PSCs readily differentiate into many different cell types, the efficiency of neuronal conversion is limited and culture media that enhances neuronal production, as well as further selection

procedures, are usually necessary to increase the homogeneity of a specific neuronal subpopulation [42]. The second method to reconstitute neural commitment *in vitro* and to achieve efficient neural induction from PSCs has relied upon adherent monolayer culture differentiation, which eliminates the use of multicellular aggregations [33, 37, 38]. In this method, PSCs are dissociated into single cells and further cultured with conditioned media that also enhances neuronal conversion. Various aspects of the monolayer differentiation protocol have been extensively studied, adapted and optimized recently by different research groups, currently allowing neural induction within a few days [43, 44]. The third method to promote neural induction from PSCs can be achieved by co-cultivation of PSCs with stromal cell feeder layers [37, 45]. Co-culturing PSCs with other cell types is based on the idea that these surrounding cells provide cues that assign cell fates along the neural lineage. The use of stromal feeder layers is, however, only an efficient PSC differentiation strategy for certain neuronal subtypes, such as dopaminergic neurons [46]. Additional efforts on optimising previous methods have recently also lead to the establishment of a combinatory approach for PSC differentiation that is characterised by a chemically transitional EB-like state [47]. A low density monolayer culture on a feeder is differentiated under appropriate culture conditions and induces intermediate progenitor cells with the capability of differentiating into the three germ layers. The main advantage of this approach is that it harbours a reduced innate differentiation propensity of PSCs, even if the PSCs are known for their unfavourable differentiation [47].

Although different strategies for neuronal conversion in PSCs have been explored, the general trend moves towards stromal-free methods combined with chemically defined culture systems. These chemically defined systems utilize culture media that are supplemented with patterning molecules, such as morphogens and growth factors, which force enrichment of the desired neuronal cells. Additionally, an increasing number of studies have also illustrated the use of small molecules and emphasized their significant benefits for neural induction and differentiation in PSCs, as well as for somatic cell reprogramming [48]. Aside from providing positive signals to induce neuronal fates, many of these factors also inhibit signalling pathways that control the differentiation into other germ layers then the ectoderm. Of note, just as in vivo embryonic development can be broken down into distinct stages where distinct patterning molecules are required to induce neuronal cell fates, so too can the *in vitro* specification of differentiating neuronal population from PSCs [49]. The compositions of the chemically defined media, as well as the compound concentrations at any given time point, can, therefore, significantly redirect the anatomical and functional identity of the differentiating cells.

Table 1. Chemically defined differentiation systems for neural induction and differentiation *in vitro* per neuronal phenotype.

Chemically defined systems

Phenotypes	Species	Starting cell types	Culture methods	Chemical driving factors	Phenotypic markers (% cells)	References
Glutamatergic neurons	Mouse	ESCs	EB	RA	±85% TUJ1+ 93 ± 4.7% VGLUT1+/TUJ1+	Bibel *et al.* (2008)
		ESCs (*TAU::GFP*)	Monolayer	Cyclopamine	±70% TUJ1+ ±70% VGLUT1+/TUJ1+ ±13% VGLUT2+/TUJ1+ <50% TBR1+/TUJ1+ <35% CTIP2+/TUJ1+	Gaspard *et al.* 2008
	Mouse and human	ESCs (*Sox1::GFP* and *Bf1::Venus*)	EB	DKK1, Lefty-1 and SB431542	*Mouse* 83% VGLUT1+/TUJ1+	Eiraku *et al.* (2008)
	Human	ESCs	EB	None	±79% TBR1+/TUJ1+ ±57% CTIP2+	Li *et al.* (2009)
		ESCs (*ACTB::GFP*) and iPSCs	Monolayer	NOG	*ESCs* <65% TUJ1+ ±60% VGLUT1+/TUJ1+ <75% TBR1+/TUJ1+ <72% CTIP2+/TUJ1+ <18% CTIP2+/TBR1+/TUJ1+	Espuny-Camacho *et al.* (2013)
		ESCs and iPSCs	(Spin) EB	DOR, DKK1, EGF, FGF2 and NOG	*iPSCs (8011)* 94.7 ± 2.5% VGLUT1+/MAP2+ 9.9 ± 4.7% GABA+/MAP2+ *iPSCs (BJiPS#1)* 32.8 ± 8.1% VGLUT1+/MAP2+ 71.1 ± 5.5% GABA+/MAP2	Kim *et al.* (2011c)
			Monolayer (Dual SMAD inhibition)	DOR, FGF2, NOG, SB431542 and Vitamin A	*ESCs* ±27% TBR1+ ±28% CTIP2+ ±34% BRN2+ *iPSCs* 22-29% TBR1+ 25-30% CTIP2+ 28-36% BRN2+	Shi *et al.* (2012a, 2012b)
GABAergic neurons	Mouse	ESCs	Co-culture (MS-5)	FGF2, FGF8 and SHH	±68% GABA+/TUJ1+	Barberi *et al.* (2003)
		ESCs (*Lhx6::GFP*)	EB	FGF2, NOG and SHH-C25II	91.6 ± 4.4% BF1+/Lxh6-GFP+ 70.5 ± 7.7% DLX2+/Lxh6-GFP+ 97.7 ± 3.1% LHX6+/Lxh6-GFP+ 94.6 ± 3.7% GABA+/Lxh6-GFP+	Maroof *et al.* (2010)

Table 1. *(Continued)*

Chemically defined systems

Phenotypes	Species	Starting cell types	Culture methods	Chemical driving factors	Phenotypic markers (% cells)	References
GABAergic neurons	Mouse	ESCs (*Sox1::GFP*)	EB	DKK1, Lefty-2 and SHH-C25II	60−70% NKX2-1+/BF1+ 20-30% GAD65/67+/TUJ1+ 20-30% GABA+/TUJ1+	Watanabe *et al.* (2005)
	Human	ESCs	EB	DKK1 and SHH	±84% NKX2-1+	Li *et al.* (2009)
		ESCs and iPSCs	EB	SHH-C24II, SHH-C25II and PUR	*ESC and iPSCs* ±90% TUJ1+ ±90% NKX2-1+ >90% BF1+ >90% GABA+ ±25% CALB1+ ±15% SST+ ±13% PV+	Liu *et al.* (2013)
			Monolayer	DKK1, DOR, NOG, SHH-C25II and SB431542	*ESC and iPSCs* ±58% BF1+ ±51% MAP2+ ±80% TUJ1+ ±78% GABA+/MAP2+ ±60.3% CTIP2+/MAP2+ ±86% GABA+/CTIP2+/MAP2+ ±53% CALB1+/MAP2+ ±70.6% CTIP2+/CALB1+/MAP2+	Carri *et al.* (2013)
		ESCs (*NKX2-1::GFP*) and iPSCs	EB	BMPRIA-Fc, DKK1, PUR and SB431542	*ESCs* 74.9% ± 2.1% NKX2-1-GFP+ 81.5 ± 3.6% BF1+/NKX2-1-GFP+ 75.8 ± 2.3% GABA+/NKX2-1-GFP+ 31.1 ± 5.4% CALB1+	Nicholas *et al.* (2013)
		ESCs (*NKX2-1::GFP*) and iPSCs	Monolayer (Dual SMAD inhibition)	DKK1, FGF2, LDN193189, NOG, PUR, SB431542, SHH-C25II and XAV939	*ESCs* ±90% BF1+ ±80% NKX2-1+ ±90% BF1+/NKX2-1-GFP+ <88% GABA+ <16% CALB1+	Maroof *et al.* (2013)
		Neural precursor cell line (ReNcell VM)	Monolayer	EGF, FGF2 and VPA[1]	68 ± 4% MAP2+ 90% GABA+/MAP2+ 54% CALB1+/MAP2+	Lin *et al.* (2015)
				DKK1, EGF, FGF2 and SHH	63 ± 4% MAP2+ 96% GABA+/MAP2+ 84% CALB1+/MAP2+	Lin *et al.* (2015)
Dopaminergic neurons	Mouse	ESCs	Co-culture (MS-5)	FGF2, FGF8 and SHH	50 ± 10% TH+/TUJ1+	Barberi *et al.* (2003)

[1] Epigenetic factor

Table 1. *(Continued)*

Chemically defined systems

Phenotypes	Species	Starting cell types	Culture methods	Chemical driving factors	Phenotypic markers (% cells)	References
Dopaminergic neurons	Mouse	ESCs	Co-culture (PA6)	None	52 ± 9% TUJ1+ 30 ± 4% TH+/TUJ1+	Kawasaki *et al.* (2000)
			EB	FGF2, FGF8 and SHH	71.9 ± 6.9% TUJ1+ 33.9 ± 5.5% TH+/TUJ1+	Lee *et al.* (2000)
	Human	ESCs	Co-culture (MEF-NOG, MS-5-NOG and MS-5-SHH)	FGF2	52.5 ± 2.56% TUJ1+ 38.2 ± 2.15% TH+ 75.0 ± 3.02% TUJ1+/TH+	Lim *et al.* (2015)
			Dual SMAD inhibition with EB	CHIR99021, NOG, SB431542 and SHH-C24II	*After in vivo transplantations* 54.2 ± 2.5% TH+ 81% LMX1A+/FOXA2+	Kirkeby *et al.* (2012)
			EB	FGF2, FGF8 and SHH	50-60% TH+/TUJ1+	Yan *et al.* (2005)
		iPSCs	EB	FGF2, FGF8 and SHH	30 ± 5% TH+ ±100% GIRK2+/TH+	Swistowski *et al.* (2010)
				FGF2	6.5 ± 1.4% TH+	Cai *et al.* (2010)
			Monolayer (Dual SMAD-inhibition)	A83-01, CHIR99021, FGF8, LDN193189 and PUR	42 ± 4.4% TH+ 19.9% ± 6.9% NURR1+ 70%–75% FOXA2+	Doi *et al.* (2014)
		ESCs and iPSCs	Monolayer (Dual SMAD-inhibition) and Co-culture (MS-5)	CHIR99021, FGF8, LDN193189, NOG, PUR, SB431542 and SHH-C25II	*ESCs* ±75% TH+ ±50% NURR1+ ±80% FOXA2+ ±60% LMX1A+	Kriks *et al.* (2011)
	Human and primate	ESCs and iPSCs	Monolayer (Dual SMAD inhibition) with subsequent cellular aggregations	CHIR99021, FGF8b, LDN193189, SB431542 and SHH-C25II	*Human ESCs and iPSCs* 43.6 ± 6.2% TH+ 95.3 ± 2.4% NURR1+/TH+ 96.7 ± 1.8% FOXA2+/TH+ 96.5 ± 2.3% LMX1A+/TH+ 56.3 ± 6.7% GIRK2+/TH+	Xi *et al.* (2012)
Serotonergic neurons	Mouse	ESCs	Co-culture (MS-5)	FGF2, FGF4 and SHH	±57% 5-HT+/TUJ1+	Barberi *et al.* (2003)
			EB	FGF2, FGF8 and SHH	11 ± 0.5% 5-HT+/TUJ1+	Lee *et al.* (2000)
		ESCs (*ePet::EGFP*) and iPSCs	Monolayer	NOG	*ESCs* ±6% 5-HT+/TUJ1+	Shimada *et al.* (2012)
	Human	ESCs	EB	5-HT, FGF2, FGF10, Forskolin and RA	±20% TUJ1+ <69 ± 4% 5-HT+/TUJ1+ 40% TPH+/MAP2+ 40 ± 4% MAP2+/5-HT+	Kumar *et al.* (2009)

Table 1. *(Continued)*

Chemically defined systems

Phenotypes	Species	Starting cell types	Culture methods	Chemical driving factors	Phenotypic markers (% cells)	References
Serotonergic neurons	Human	ESCs (*TPH2::GFP, TPH2::TdT and SYN1*::dsRed)	Dual SMAD inhibition with EB	FGF8, LDN193189, NOG, SB431542 and SHH	±8% TPH+/MAP2ab+ ±27% 5-HT+/TPH+ ±64% TPH+/TPH-GFP+ ±60% TPH-GFP+/TPH+ ±5% 5-HT+/MAP2ab+	Vadodaria *et al.* (2015)
		ESCs and iPSCs	Monolayer	CHIR99021, DMH-1, FGF4, SB431542 and SHH-C25II	*ESCs and iPSCs* ±52% TPH2+ >60% 5-HT+/TUJ1+	Lu *et al.* (2016)
Cholinergic motor neurons	Mouse	ESCs	Co-culture (MS-5)	FGF2, RA and SHH	±60% HB9+/TUJ1+	Barberi *et al.* (2003)
		ESCs (*Hb9::GFP*)	EB	Hh-Ag1.3, RA and SHH	25% ± 5% HB9+ 25% ± 5% HB9-GFP+ >70% ISL1+/ HB9+	Wichterle *et al.* (2002)
	Human	ESCs	EB	FGF2, RA and SHH	>50% ISL1+/TUJ1+/MAP2+ ±50% HB9+/ISL1/2+ ±21% HB9+	Li *et al.* (2005)
		ESCs (*Hb9::EGFP*)	EB	RA and SHH	35.3 ± 24.9 TUJ1+ 56.1 ± 9.9% ISL1+/CHAT+/TUJ1+ 37.4 ± 3.3% HB9+/TUJ1+ 96.3 ± 12.5% HB9+/Hb9-GFP+ 88.7 ± 7.4% TUJ1+/Hb9-GFP+	Singh Roy *et al.* (2005)
		ESCs and iPSCs	EB	RA and SHH agonist	*iPSCs* 20% HB9+ >90% ISL1/2+/HB9+ >50% CHAT+/ISL1/2+/HB9+	Dimos *et al.* (2008)
			Monolayer	DOR, RA and SHH	*ESCs* 69.5 ± 11.2% HB9+/CHAT+/TUJ1+ 78.3 ± 3.5% ISL1+ *iPSCs* 51.6 ± 5.7% HB9+/TUJ1+ 65.4 ± 5.1% ISL1+	Qu *et al.* (2014)
		ESCs and iPSCs (*Hb9::(E)GFP*)	EB	PUR and RA	*ESCs* 59.1 ± 7.07% OLIG2+/SOX3+ 28.2 ± 5.7% ISL1+/TUJ1+ *iPSCs* 57.6% ± 9.88% OLIG2+/SOX3+ 33.6% ± 12% ISL1+/TUJ1+	Karumbayaram *et al.* (2009)

Table 1. *(Continued)*

Chemically defined systems

Phenotypes	Species	Starting cell types	Culture methods	Chemical driving factors	Phenotypic markers (% cells)	References
Cholinergic motor neurons	Human	ESCs and iPSCs (*Hb9::(E)GFP*)	Dual SMAD inhibition with EB	LDN193189, PUR, RA, SAG, SB435142 and SHH-C25II	*ESCs* 83 ± 1% TUJ1+ <29 ± 4% HB9-GFP+ 98 ± 0% HB9-GFP+/TUJ1+ 30 ± 6% ISL1+/HB9-GFP+ 16 ± 5% HB9+/HB9-GFP+ 37 ± 2% ISL1+/HB9+/HB9-GFP+	Amoroso *et al.* (2013)
		ESCs and iPSCs (*HB9::Venus*)	Dual SMAD inhibition with EB	BIO, CHIR99021, DOR, LDN193189, PUR, RA and SB431542	*ESCs* 50-58% HB9+ 45-50% ISL1+ 43-68% CHAT+ 84.1 ± 2.4% HB9+/HB9-Venus+ 39.5 ± 13.2% ISL1+/HB9-Venus+ 83.4 ± 1.7% CHAT+/HB9-Venus+ *iPSCs* 42-48% HB9+ 37-48% ISL1+	Shimojo *et al.* (2015)
	Human and primate	ESCs (*RUES1-EGFP)*	Co-culture (MS-5)	NOG, RA and SHH	*Human* 20% HB9+ ±26% CHAT+ *Primate* 43% HB9+ 65% HB9+/TUJ1+	Lee *et al.* (2007)

Neuronal phenotypes, species, starting cell types, culture methods, chemical driving factors and representative phenotypic markers that have been used to assess the differentiation efficiency and culture homogeneity are broadly summarized. + indicates the percentage of cells in the population that stained positive for a certain marker. *Abbreviations*: 5-HT, serotonin (5-hydroxytryptamine); A83-01, TGF-β kinase/activin receptor-like kinase inhibitor; ACTB, actin beta/beta-actin; BF1, brain factor 1/forkhead box protein G1 (FOXG1); BIO, GSK3β inhibitor 6-bromoindirubin-3'-oxime; BMPRIA-Fc, bone morphogenetic protein receptor 1a-fragment crystallizable; BRN2, brain-specific homeobox/POU domain protein 2 (POU3F2);; CALB1, calbindin 1; CHAT, choline o-acetyltransferase; CHIR99021, GSK3β inhibitor; CTIP2, b-cell CLL/lymphoma 11b (BCL11B)/ COUP-TF-interacting protein 2 (COUP-TFII); DKK1, dickkopf-1; DLX2, distal-less homeobox 2; DMH-1, dorsomorphin homolog 1; DOR, Dorsomorphin; dsRed, discosoma sp. red fluorescent protein; EB, embryoid body; EGF, epidermal growth factor; EGFP, enhanced green fluorescent protein; ePet, enhancer of the mouse Pet-1 (human FEV) gene; ESCs, embryonic stem cells; FGF10, fibroblast growth factor 10; FGF2, fibroblast growth factor 2/basic fibroblast growth factor (bFGF); FGF4, fibroblast growth factor 4; FGF8, fibroblast growth factor 8; FGF8b, fibroblast growth factor 8 isoform b; FOXA2, forkhead box protein A2; GABA, γ-aminobutyric acid; GAD65/67, glutamic acid decarboxylase isoform 65/67 (GAD2/1); GFP, green fluorescent protein; GIRK2, G protein-activated inward rectifier potassium channel 2 (KCNJ6); HB9, homeobox HB9/motor neuron and pancrease homeobox 1 (MNX1); Hh-Ag1.3, small molecule agonist of SHH signalling; iPSCs, induced pluripotent stem cells; ISL1, ISL LIM homeobox 1; ISL1/2, ISL LIM homeobox 1/2; LDN193189, selective BMP signalling inhibitor; Lefty-1, left-right determination factor 1; Lefty-2, left-right determination factor 2; LHX6, LIM homeobox 6; LMX1A, LIM homeobox transcription factor 1 alpha; MAP2, microtubule-associated protein 2; MAP2ab, microtubule-associated protein 2ab; MEF, mouse embryonic fibroblast; MS-5, stromal cell line derived from irradiated murine bone marrow cultures; NKX2-1, NK2 homeobox 1; NOG, Noggin; NURR1, nuclear receptor

related 1 protein;OLIG2, oligodendrocyte transcription factor 2; PA6, stromal cell line derived from newborn calvaria tissue of the C57BL/6 mice; PUR, purmorphamine; PV, parvalbumin; RA, retinoic acid; RUES1-EGFP, human ESC line expressing EGFP; SAG, smoothened agonist; SB431542, transforming growth factor beta inhibitor; SHH, sonic hedgehog; SHH-C24II, recombinant human SHH; SHH-C25II, recombinant mouse SHH; SMAD, transcription factor and member of the BMP and TGF-β signalling pathways; Sox1, SRY box 1; SOX3, SRY box 3; SST, somatostatin; SYN1, synapsin 1; TAU, microtubule-associated protein tau (MAPT); TBR1, T-box brain 1; TdT, tandem dimer tomato red fluorescent protein; TH, tyrosine hydroxylase; TPH, tryptophan hydroxylase; TPH2, tryptophan hydroxylase 2; TUJ1, neuron-specific class III beta-tubulin (TUBB3); VGLUT1, vesicular glutamate transporter 1; VGLUT2, vesicular glutamate transporter 2; VPA, valproic acid; XAV939, WNT/β-catenin inhibitor.

Neuronal conversion follows the neural induction principle that is first initiated by removing medium components that promote self-renewal, which on itself is sufficient to trigger differentiation towards all the three embryonic germ layers [50]. Inhibition of extraembryonic and meso-endoderm differentiation can be further enhanced by culturing the cells in serum-free medium and by the actions of early patterning molecules, through which the PSCs progressively start restricting their differentiation potential towards the neural lineage to form early NPCs [50, 51]. Many neural induction protocols include the simultaneous inhibition of transforming growth factor beta (TGF-β)/Activin/Nodal and bone morphogenic protein (BMP) signalling pathways, *i.e.* dual SMAD inhibition, which is similar to what is observed in vivo [33, 52]. Inhibition of the TGF-β and BMP pathways is thought to promote differentiation of PSCs along the neural lineage primarily through inhibition of PSC self-renewal, as well as by blocking differentiation towards alternative cellular lineages [49]. Consequently, signalling molecules such as Noggin (NOG), left-right determination factor (Lefty)-1, Lefty-2, Dickkopf-1 (DKK1), TGF-β inhibitor SB431542, glycogen synthase kinase 3 beta (GSK3β) inhibitor CHIR99021 and BMP inhibitor dorsomorphin homolog 1 (DMH-1) have all been used to promote neural induction from PSCs [53, 54]. Other pathways, including epidermal growth factor (EGF), fibroblast growth factor (FGF) and wingless-type MMTV integration site family (WNT) signalling, have been described to regulate neuronal differentiation by promoting induction and survival of the NPCs [49, 55]. In particular, FGF2 has been shown to enhance the neural induction phase and to increase the number of NPCs, whereas omitting it during subsequent stages promotes their differentiation into mature neurons [56].

After neural induction, second series of lineage-specific patterning molecule cocktails have been used to direct the further differentiation of the NPCs towards mature neuronal subtypes. Although the availability of protocols for subtypes within specific neurotransmitter classes is limited and different cultivation conditions for each of these classes exist, protocols for glutamatergic, dopaminergic, γ-aminobutyric acid (GABA)-ergic, serotonergic, and cholinergic motor neurons have become available over the years (See Table 1) [53, 57-60]. Specification of the NPCs takes place both along the rostral-caudal and the

dorsal-ventral axes of the brain, coordinated by the synergistic actions of temporally and spatially available patterning molecules [61]. The presence and concentrations of these molecules define the transcriptional code and, hence, the identity of the NPCs in a particular domain along both axes. NPCs are generally specified first in the head region and extend caudally, meaning that they become committed to an anterior forebrain fate by a default programme [50]. Correspondingly, NPCs differentiated from PSCs, independent of the differentiation method, carry a rostral identity that is free of caudal markers [50]. Indeed, this anterior phenotype is transient and NPCs will take on a definitive regional identity depending on further cues. Treatment with increasing concentrations of sonic hedgehog (SHH) has shown to promote ventralization of the NPCs, while addition of retinoic acid (RA) promotes caudal fates [49] and activation of WNT signalling exerts a dose-dependent effect where increasing concentrations are patterning the NPCs to forebrain, midbrain, hindbrain and anterior spinal cord identities, respectively [49, 50, 62]. Although it remains largely unknown how all the patterned NPCs acquire functional and anatomical specificity, it is this regional patterning principle recapitulating in vivo morphogenesis that guides PSC neuronal differentiation in vitro (See Fig. 2) [63]. Moreover, a specific neurotransmitter subtype can often be generated in different parts of the human brain and at different stages, demonstrating that different spatiotemporal cues can likely converge on the same terminal selectors and thereby induce a similar terminal fate [64].

The anatomically directed differentiation processes seen in these neuronal differentiation protocols are characterized by the same expression and temporal regulation of lineage-specific transcription factors as observed in vivo [49]. For instance, NPCs differentiated towards forebrain neurons carry an anterior identity by expressing PAX6 and OTX2, but not more caudal markers like EN1, GBX2 or HOX [33, 50, 65]. Thus, expression levels of these transcription factors represent useful markers that are widely used to assess the differentiation status of the cellular population (See Table 1). Furthermore, extensive gene expression analysis, electrophysiological characterization, biochemical assessments, and in vivo transplantation into rodent brains have been applied to examine the population characteristics, as well as quality and efficacy of the differentiation protocols (See Table S1). However, with many of these protocols, in vitro-directed differentiation in PSCs still results in highly variable neuronal populations with fluctuating yields of neuronal cells and remarkable differences in efficiency [66]. This variability emerges from clear differences between the protocols, such as the neuronal induction method used, the chemical compositions of the media, the compound concentrations and the chemical exposure times that were used, as

well as more undefined differences, including the culture densities and the passage number for example [49].

Glutamatergic neurons

Despite the differences in developmental principles that underlie the specification of their subpopulations, excitatory glutamatergic neurons can be found throughout the whole central nervous system, such as in the cerebral cortex [67], as well as in subcortical regions like the thalamus [68], and even in the spinal cord among others [69]. To date, strategies that have been established to successfully generate glutamatergic neurons from PSCs (See Table 1) are mainly based on the derivation of cortical glutamatergic neurons of which the vast majority originates from dorsal telencephalic regions [70]. As mentioned earlier, forebrain identity is a default programme for neuronal differentiation of PSCs, and existing protocols yield neurons with a glutamatergic identity without the need of an extra second series of patterning molecules. However, attributed to endogenous SHH signalling, mouse ESCs have been shown to differentiate into anterior NPCs with a ventral phenotype under serum- and morphogen-free culture conditions, resulting in a neuronal population of which only a minority was considered to be glutamatergic [71]. Consequently, inhibition of intrinsic SHH signalling with the small molecule cyclopamine has been shown to prevent ventralization of the mouse NPCs and significantly increases the derivation of dorsal glutamatergic neurons [71]. In a separate study on the other hand, it has also been demonstrated that mouse ESCs cultured in the presence of retinoic acid (RA) induced highly homogenous neuronal populations, which, similar to previous report, where consistent with an identity of cortical pyramidal neurons [72]. Pyramidal neurons constitute more than 80% of the cerebral cortex neurons and are further diversified in distinct cortical layers that establish specific patterns of axonal output and dendritic input, providing the essential substrate of cortical circuitry [73]. In a study by Eiraku et al. [57] for example, it was also demonstrated that human ESC-derived cortical neurons stained positive for transcription factors corresponding to the development of the cortical layers in a temporal manner, including RELN, TBR1, CTIP2 and CUX1. However, opposite to the ventral phenotype observed in mouse ESCs, human PSCs have shown to differentiate by a default programme into synchronised populations of NPCs that predominantly express anterior dorsal markers [70, 71]. This dorsal phenotype has been attributed to expression of endogenous WNT ligands, and, as a consequence, inhibition of WNT signalling or activation of SHH signalling has shown to almost completely convert primitive dorsal telencephalic NPCs to ventral progenitors at the expense of glutamatergic neuron identity [70].

Various studies have also addressed the potential of using the derived glutamatergic cultures for fundamental research. Kim et al. [74] suggested an approach for efficient differentiation of human glutamatergic neurons based on a spin EB protocol in ESCs and iPSCs. Interestingly, when co-cultured with human embryonic kidney cells (HEK293T) expressing NLGN3 and NLGN4, but not those containing autism-associated mutations, the iPSC-derived neurons were able to form functional synapses, demonstrating that these neuronal populations are a potential model for the study of synaptic differentiation and function under normal and disorder-associated conditions [74]. Also in their differentiation approach it was demonstrated that the human spin EB-derived NPCs acquired an anterior dorsal forebrain character by a default pathway. While addition of the SHH agonist purmorphamine (PUR) during the EB stages ventralized the NPCs, inhibition of SHH, however, did not enhance expression of dorsal markers as seen in mouse ESCs [74]. Consistent with this finding, a more recent study also showed that cyclopamine treatment was not required for induction of the dorsal telencephalic fate in a human PSC monolayer system [73]. By determining the time of onset for the expression of layer-specific markers during the course of differentiation, it was demonstrated that most of the pyramidal neurons generated here displayed an identity corresponding to deep cortical layers, while upper layer neurons were underrepresented [73]. This contrasts with a more recent report describing a robust culture system in human ESCs for the generation of both electrophysiological active deep- and upper-layer pyramidal neurons in equivalent proportions that were cultured in the presence of retinoids [75, 76]. In the context of dual SMAD inhibition, they found that vitamin A is crucial for the efficient induction of cortical NPC differentiation and subsequent cortical neurogenesis. These findings agree with previous report [72], which demonstrated that derivatives of retinoids have important roles in the acquisition of NPCs and telencephalic glutamatergic neurons from mouse ESCs. By using this approach, the efficiency of cortical neural induction from PSCs even approaches 100% [76]. Although various organisations are still missing, such as interactions with glial cells that are essential partners in synaptic functioning, these systems provide the first steps towards functional studies of human cerebral cortex development and the generation of patient-specific cortical networks in vitro. These future applications will be particularly interesting for modelling disorders that are known for their cortical synaptic dysfunctions, including epilepsy, schizophrenia and dementia, and allow high-throughput testing for therapeutic interventions [75, 76].

GABAergic neurons

Various subtypes of inhibitory GABAergic neurons exist in the brain and spinal cord, which can be categorised based on their developmental origins,

localization, synaptic connections, co-expression of molecular and neurotransmitter markers and electrophysiological properties [77]. During development, GABA interneurons are synaptically integrated into neuronal networks in the forebrain that originate mostly from the medial ganglionic eminence and, to a lesser extent, from the ventral lateral ganglionic eminence and the anterior dorsal ganglionic eminence [77-79]. These GABAergic NPCs migrate by following radial or tangential pathways, they differentiate into post-mitotic neurons and make connections with local glutamatergic neurons [77, 80, 81]. The vast majority of forebrain GABAergic interneuron progenitors express Nkx2- 1 [82] in addition to the telencephalic transcription factor Bf1, also known as Foxg1 [53], and they can be distinguished from other types of GABAergic neurons, including the striatal GABAergic projection neurons, which originate from the lateral ganglionic eminence [77, 83]. Forebrain GABA interneurons can be divided into many subgroups on the basis of molecular markers and their expression of neuropeptides or calciumbinding proteins, including somatostatin (SST), parvalbumin (PV), calretinin, calbindin and neuropeptide Y, although the medial ganglionic eminence progenitors mostly give rise to SST and PV interneurons [77].

The majority of studies on the differentiation of GABAergic neurons in vitro have focused on guiding PSCs toward ventral telencephalic NPCs, primarily defined by the co-expression of Bf1 and Nkx2-1 (See Table 1) [53, 70, 84]. As stated in the previous section of glutamatergic neurons, ventral telencephalic pecursors have been generated from mice ESCs without the need of additional patterning factors, leading to a neuronal population that is enriched in cortical GABAergic interneurons [71].Ventralization of human dorsal telencephalic NPCs on the other hand, has been achieved by addition of concentrated SHH or inhibition of WNT by DKK1 together with low concentrations of SHH, leading to the generation of enriched populations of human cortical GABAergic interneurons [70]. Interestingly, various groups have tried to purify the generated GABAergic NPCs using cellular selection systems. For instance, Maroof et al. [85] described a protocol for the generation of cortical GABAergic interneurons from mouse ESCs based on EB formation and SHH signalling. In this study, a Lhx6-GFP bacterial artificial chromosome reporter construct was used, which allowed for the isolation and enrichment of the newly generated NPCs. Using a similar approach with a previously established NKX2-1::GFP human ESC reporter line, two other groups [16, 86] have developed a protocol based on the combination of small molecules with the timed activation of SHH signalling. In both studies they showed that the human PSCs develop into GABAergic interneurons with mature physiological properties, both in vitro, as well as after transplantation into rodent brains. Liu et al. [77] on the other hand, have described a protocol without transgenic modification or cell sorting that involves treatment with SHH or its antagonist PUR

for directed GABAergic differentiation of human PSCs with high efficiency. After 2 weeks of differentiation, more than 90% of the neurons were estimated to be GABAergic interneurons, which was confirmed by immunostaining and electrophysiological analysis [77].

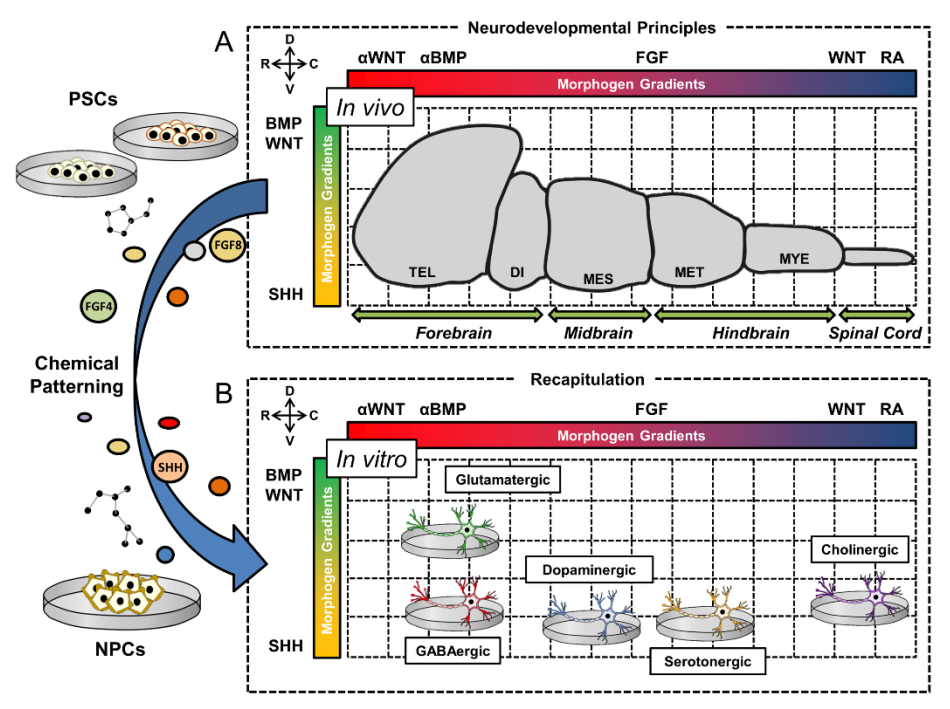


Figure 2. Schematic overview of the working mechanisms of chemically defined neuronal differentiation systems and the underlying neurodevelopmental principles that they recapitulate *in vitro*. **A.** The figure shows how various morphogen signalling gradients, including bone morphogenic protein (BMP), wingless-type MMTV integration site family (WNT), fibroblast growth factor (FGF), sonic hedgehog (SHH) and retinoic acid (RA), as well as inhibitors/antagonists (α) of these pathways, pattern the various brain regions during early embryonic development both along the rostral-caudal and dorsal-ventral axes. The depicted brain regions include the telencephalon (TEL), diencephalon (DI), mesencephalon (MES), metencephalon (MET), myelencephalon (MYE) and spinal cord. **B.** By using the same chemical patterning principles as seen *in vivo*, pluripotent stem cells (PSCs) and neural precursor cells (NPCs) can be patterned towards neuronal subtypes *in vitro*, corresponding to the brain regions where they typically originate from.

Aside from the cortical GABAergic interneurons, several studies have also managed to acquire striatal GABAergic projection neurons, also known as medium spiny neurons [87]. In one of the procedures, neural induction via BMP/TGF-β inhibition was coupled with exposure to SHH and DKK1 to drive ventral telencephalic specification in human ESC and iPSCs, followed by the terminal differentiation towards authentic medium spiny neurons [87]. Authenticity of the resulting neuronal population was monitored by the appearance of

BF1/GSX2-positive progenitor cells typical for the lateral ganglionic eminence, followed by appearance of CTIP2-, FOXP1- and FOXP2-positive cells. These precursor cells then matured into MAP2/ GABA-positive neurons with 20% of them co-expressing DARPP-32 and CTIP2, and also carried electrophysiological properties expected for fully functional medium spiny neurons [87]. Most recently, a reliable and simplified two- and three-step protocol to derive striatal GABAergic neurons from immortalized NPCs has also been established, using valproic acid (VPA) or SHH and DKK1, respectively [88]. The differentiated cells expressed appropriate GABAergic markers and responded to ionotropic glutamate receptor stimulation. In accordance, the cells also expressed various glutamate receptor subunits and released GABA upon stimulation [88]. In relation to disease modelling, the derivation of these GABAergic neurons represents a possible critical resource for the study of Huntington's disease and Rett syndrome for example [88, 89].

Dopaminergic neurons

Dopaminergic neurons are localized in the diencephalon, mesencephalon and the olfactory bulb [90], although the most prominent group resides in the mesencephalon, containing approximately 90% of the total number of brain dopaminergic neurons [90]. Differentiation protocols for dopaminergic neurons, in particular for the ones originating from the midbrain, have received a lot of attention due to their applicability into regenerative medicine for Parkinson's disease, with numerous differentiation protocols published over the last years (See Table 1) [51, 91, 92]. Studies in mice have shown that dopaminergic midbrain NPCs are specified from the floor plate in the mesencephalon, which is located at the ventral midline of the neural tube [91, 92]. These NPCs are transcriptionally characterised by the expression of Lmx1a, Foxa2, En1 and Otx2, and are controlled by two regulatory feedback loops also involving WNT and SHH signalling [93, 94]. In more detail, WNT1 induces expression of Otx2, which represses Gbx2 to coordinate the mid-hindbrain organizer and represses Nkx2-2, which defines the midbrain dopaminergic NPC domain from the lateral located progenitors of serotonergic neurons [95]. Thereby, it induces the expression of Lmx1a which either induces the pro-neural gene Ngn2 through Msx1 or inhibits the NPCs from acquiring alternative cell fates by repressing Nkx6-1 [93, 96].

It is this developmental principle in mice described above that currently also forms the guideline for differentiating midbrain dopaminergic neurons from human PSC [94]. However, initial reports for the differentiation of dopaminergic neurons have also heavily relied on the use of PA6 and MS-5 feeder cells [45, 46, 97]. Furthermore, studies have shown the successful differentiation of TH (tyrosine hydroxylase)- expressing neurons from mouse ESCs and iPSCs based on

chemically defined systems that relied on the generation of EBs [98] and the activation of key signalling pathways by SHH and FGF8, a morphogen important for the formation of the isthmus [51, 97-99]. Gradually, studies have tailored and applied these initial studies in primate [94] and human iPSCs [60, 100, 101], and demonstrated efficient induction of neurons with a dopaminergic phenotype. Importantly, in most of the previous mentioned reports it was actually not conclusively determined whether the obtained TH-positive neurons were really representing midbrain dopaminergic neurons, at least they did not always carry abundant midbrain markers, suggesting that the combination of FGF8 and SHH had the potency to induce the dopaminergic identity but was possibly not sufficient to restrict the neurons to the midbrain fate [94].

A better understanding of essential signalling pathways and transcriptional networks important for dopaminergic neuron midbrain differentiation, as well as their more precise temporal implementation, has improved the protocols over time. Incorporation of WNT/β-catenin signalling due to the availability of the GSK3β inhibitor CHIR99021 eventually led to an improved midbrain specification in a reliable and efficient manner [62, 94, 102-105]. This was first demonstrated by Kriks et al. [105], who produced cultures containing around 75% of floor plate-derived dopaminergic neurons, assessed by immunostaining of markers such as FOXA2 and TH. Further expression analysis in this study also demonstrated abundant co-expression of lineage-specific genes necessary for appropriate dopaminergic neuron specification [105]. Furthermore, the cells were able to be efficiently engraft in rodent brains and survived in vivo without overgrowing, a phenomenon that was previously only observed with very poor performance using human PSC-derived dopaminergic neurons [105]. Their findings were later also confirmed in different human ESC lines, iPSCs, and rhesus monkey iPSCs, showing that a narrow range of CHIR99021 at a particular developmental stage restricts the cells to form midbrain floor plate progenitors which, in the presence of FGF8, acquire a dopaminergic neuron identity [94]. Based on the latter approaches, differentiated dopaminergic populations from Parkinson's disease patient-derived iPSCs have even demonstrated to mimic several pathological mechanisms of the neurological disorder in vitro [106, 107]. Interestingly, disease phenotypes were only observed in the PSC-derived dopaminergic neurons and not in patients' fibroblasts, which emphasizes the significance of directed dopaminergic differentiation protocols for disease modelling in vitro [107].

Serotonergic neurons

Serotonergic neurons are found in the raphe nuclei that arise from progenitors in the rhombencephalon during development and can be divided into two main clusters: A rostral division located just caudal to the isthmus and a more caudal

division situated in the myelencephalon [108]. While the rostral division has widespread innervating projections throughout the brain, the caudal division mainly projects down to the spinal cord [108]. These primary anterior and posterior clusters are further segmented along the rostro-caudal axis according to 9 rhombomeric divisions, where specific transcriptional codes confer positional identities. For instance, the progenitors located at rhombomeric segments 2–3 are distinguished from segment 1 by expression of Hoxa2 but not En1, while progenitors in segment 4 express Phox2b, leading into an intervening gap between cluster 3 and 5, where serotonergic identity is repressed and visceromotoneurons are formed instead [59, 109]. Specification of the distinct segments is thought to be induced by different combinations of morphogen gradients of which the most important ones include SHH, FGF8 and FGF4 [110]. In addition, WNT- and TGF-beta signalling have also been shown to be important for determining boundaries and specifying a hindbrain fate [62, 111, 112]. Developing serotonergic NPCs gradually start expressing Nkx2-2, Ascl1 and Foxa2, which constitutes a primary gene regulatory network for serotonergic specification [108]. Although it is evident that their postmitotic neurons in different regions are transcriptional and functional heterogeneous, this primary genetic cascade activates a secondary network that consists of a set of core transcription factors, including GATA2, GATA3, INSM1, LMX1B and PET1 in mice or FEV in humans, which is key in their terminal specification [108].

Thus far, there have been only few attempts to differentiate serotonergic neurons in chemically defined systems from PSCs (See Table 1). Mice studies have given primary knowledge on the combinations of developmental signals that allow the generation of serotonergic neurons in vitro [110]. One of the first approaches to induce serotonergic neurons from ESCs was based on the formation of EBs in combination with the activation of SHH and FGF8 signalling [98]. The protocol was primarily intended to enrich for dopaminergic neurons, resulting in considerably low yields of serotonergic neurons (± 11%) compared to the total neuronal population (± 72%). Mouse ESCs have then also been co-cultured with stromal cells in the presence of SHH and FGF4, yielding a substantially higher proportion (± 57%) of serotonergic neurons [97]. More recently, a simplified method to generate serotonergic neurons from mouse PSCs in monolayer cultures has also been developed [113]. The cells were cultured on a layer of matrigel in the presence of NOG, a BMP inhibitor, and reached 80% pure serotonergic cultures after cell sorting [113]. Although their approach increased the homogeneity of serotonergic neurons and appears to be an appealing simpler alternative to culturing mouse ESCs on feeder cells, effectively only ± 6% of the initial neuronal cells were considered serotonergic, which remains a relatively low proportion of cells that can be generated from mouse PSCs in vitro.

Studies have consequently demonstrated that the key developmental signalling molecules involved in serotonergic differentiation could also be applied for directing their differentiation utilizing human PSCs. One of the first protocols established in order to obtain serotonergic neurons from human ESCs was based on the induction and enrichment of NPCs that differentiated toward serotonergic neurons under empirically determined culture conditions [114]. In this study, neural differentiation was estimated to be around 20% with up to 70% of all neurons staining positive for serotonin (5-HT), generating 14% of serotonergic neurons in vitro. Interestingly, acidic FGF, which is localized in raphe neurons in rats [115], and 5-HT, which is known to contribute to the development of serotonergic neurons *in vivo* [116], were added to the growth factor cocktail for differentiation and maturation of the NPCs. Using a differentiation protocol based on FGF8 and SHH, another study [117] found that approximately 8% of the differentiating human ESCs committed to a serotonergic fate, which was determined by a co-staining of 5-HT and tryptophan hydroxylase (TPH).

Thus, although these first protocols in mice and human PSCs were very promising, they only met limited success considering their low yields, which limits the ability to use these cells for further applications. However, most recently, Lu et al. [59] were the first to develop a robust chemically defined system to induce human iPSC to enriched populations of serotonergic neurons with very high efficiency. Induction of ventral hindbrain NPCs was achieved by maintaining PSCs in medium containing SB431542, DMH-1 and CHIR99021. Following neural induction, the NPCs were exposed to FGF4 together with SHH to promote the acquisition of a serotonergic cell fate, eventually reaching more than 60% serotonergic neurons [59]. The key aspect here, is the activation of the WNT pathway by CHIR99021 that, within a narrow window of concentrations, gives rise to ventral hindbrain NPCs that further differentiate into serotonergic with a typical identity for rhombomeric divisions 2–3. Interestingly, treatment with the FDA-approved antidepressants tramadol and escitalopram oxalate in this latter study resulted in release or uptake of 5-HT in a dose- and timedependent manner, which emphasized their utility for the evaluation of drug candidates in depression [59]. Dysregulation of the serotonergic system is typical in depression and a common target for antidepressants [118, 119].

Cholinergic motor neurons

Cholinergic motor neurons can be broadly divided into two main groups according to the location of their cell body: (I) Upper motor neurons located in the motor regions of the cerebral cortex, and (II) lower motor neurons, which are located in the brainstem and spinal cord [49]. Upper motor neurons have ascending

pathways to lower motor neurons, which project to the musculature, where they control muscle contractions via neuromuscular junctions [49]. Spinal motor NPCs originate from a highly restricted foci in the ventral neural tube in response to RA, FGFs and SHH [52]. These cells express the basic helixloop-helix transcription factor Olig2, which, together with Ngn2, direct the expression of motor neuron fate consolidating genes such as Hb9 and Isl1. For a more detailed description of the underlying developmental cascade that results in the acquisition of these cells, we would like the reader to refer to other excellent reviews [49, 52]. Motor neurons can be further developmentally allocated to discrete motor columns, which extend along the rostral-caudal neural tube and contain motor neuron pools that are responsible for innervating a single skeletal muscle, each of which is also arranged by an anatomical logic corresponding to their targets [52]. Retinoid signalling plays key roles in the diversification of motor neuron subtypes from the common NPC pool and additionally contributes to spinal cord columnar organisations, which are then again characterized by unique transcriptional codes that define the regional identity of the neuronal subtypes [52].

By recapitulating the developmentally rationalized programme of morphogenic cues, considerable advances using chemically defined systems have primarily been made in differentiating PSCs into lower spinal motor neurons (See Table 1). The desire to regenerate in vitro motor circuitry in the contexts of motor neuron disease and spinal cord injury has been motivating the attempts to produce motor neurons for translational research [49]. Initial studies have outlined methods to derive functional cholinergic motor neurons from mice [97] and human ESCs [58, 120, 121], while more recent studies have applied similar methods to human iPSC lines [36, 122-124]. The numerous protocols that have been developed utilize various directed differentiation methods, including co-culture with stromal feeders [58, 97], adherent monolayer cultures [123], or the use of EB induction followed by neuralization, cholinergic neuron differentiation and neuronal maturation [36, 124]. In most cases, studies have reported the use of RA treatment with addition of recombinant SHH or small molecule agonists of the SHH signalling pathway to induce differentiation of PSCs into cholinergic motor neurons. Patterning NPCs by RA and SHH confers caudal and ventral anatomical identities, respectively, and gives rise to OLIG2 expressing neurons, which in turn begin to express CHAT, HB9 and ISL1 [124]. Cholinergic motor neurons obtained through these methods have been shown to possess numerous characteristics of their in vivo equivalents, including electrophysiological properties, the possibility to engraft into the developing spinal cord and the presence of correctly labelled neuromuscular junction complexes, demonstrating the potential ability to form functional muscular junctions in vitro [49]. However, in contrast to the generic developmental principles that allow cholinergic motor neuron

specification, the process determining how individual motor neuron subtypes can be generated is relatively less well understood [49]. Nevertheless, in the context of mouse ESC differentiation for example, protocols based on a treatment with RA typically result in spinal motor neurons with a rostral cervical character, as judged by expression of Hoxc5 and Hoxc6, but not Hoxc8 [125]. In the context of human motor neuron differentiation on the other hand, a recent report demonstrated a shift in the proportion of motor neurons expressing the median motor column marker LHX3 or the lateral motor column marker FOXP1, when SHH signalling was activated via a combination of smoothened agonist (SAG) and PUR instead of recombinant SHH [126]. This sensitivity of the differentiating NPCs emphasizes the need for thorough evaluation of the differentiation protocols and also presents the opportunity for the optimization of motor neuron subtype specification [49].

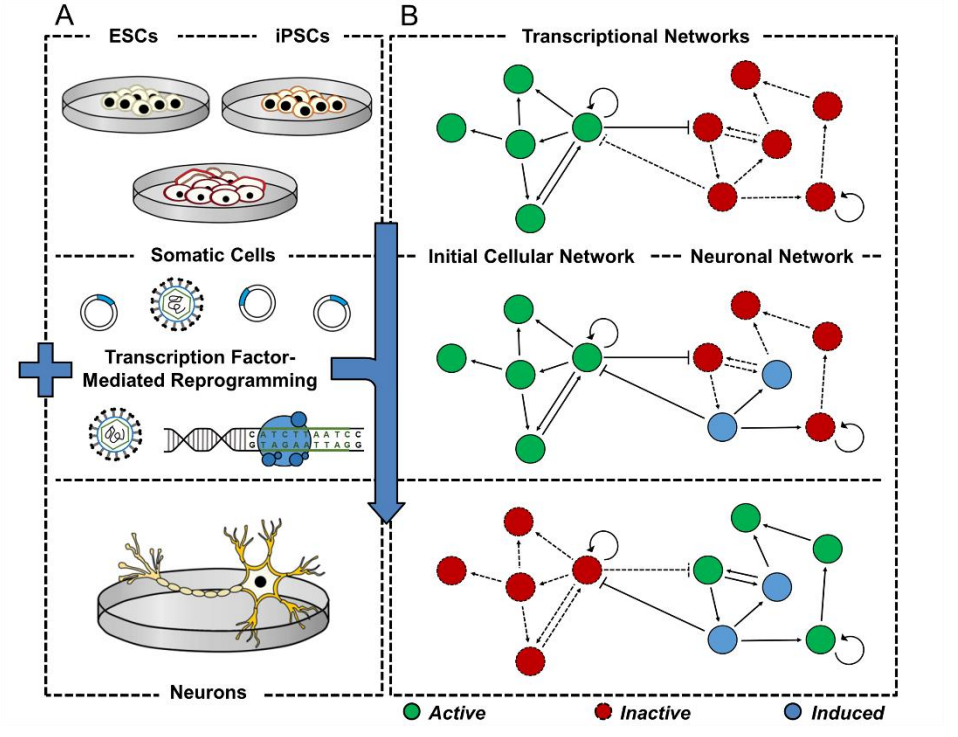


Figure 3. Schematic overview of the working mechanisms of transcription factor-mediated reprogramming along the neural lineage *in vitro*. **A.** The figure shows how embryonic stem cells (ESCs), induced pluripotent stem cells (PSCs) and somatic cells can be reprogrammed towards neurons by transcription factor-mediated fate instructions. **B.** The initial active (in green) transcriptional network that defines and reinforces cellular identity can be perturbed by induction of exogenous key-lineage determinant factors (in blue), leading to the activation (in green) of the neuronal transcription factor network and deactivation (in red) of the initial cellular network, which underlies the cellular fate switch.

Table 2. Transcription factor-mediated neuronal reprogramming protocols in vitro per neuronal phenotype.

Transcription factor-mediated reprogramming							
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Transcription factors**	**Chemical driving factors**	**Phenotypic markers (% cells)**	**References**
Glutamatergic neurons	Mouse	Fibroblasts (*Tau::EGFP)*	Monolayer	*Ascl1, Brn2* and *Myt1l*	DOX	>20% TUJ1+ 53% TBR1+/TUJ1+	Vierbuchen *et al.* (2010)
		Astrocytes	Monolayer	*Ngn2*	None	70.2 ± 6.3% TUJ1+ 85.4 ± 5.0% VGLUT1+ ±48.2% TBR1+	Heinrich *et al.* (2010)
		ESCs	Monolayer	*Ngn2*	None	<40% TUJ1+	Thoma *et al.* (2012)
	Mouse and human	Fibroblasts (*Tau::EGFP*) and ESCs	Monolayer	*Ascl1*	DOX	*Mouse embryonic fibroblasts* ±45% VGLUT1+/TAU-EGFP+	Chanda *et al.* (2014)
	Human	Fibroblasts (*Tau::EGFP*), ESCs and iPSCs	Monolayer	*Ascl1, Brn2, Myt1l* and *Neurod1*	DOX	*Fetal fibroblasts* ±60% TUJ1+ >50% VGLUT1+/TUJ1+ >50% VGLUT2+/TUJ1+ 17 ± 8% TBR1+/TUJ1+ *Postnatal fibroblasts* 81 ± 17% TBR1+/TUJ1+	Pang *et al.* (2011)
		ESCs and iPSCs	Monolayer	*Ngn2 and Neurod1*	DOX	*Ngn2- ESCs* ±80% MAP2+ *Ngn2 - iPSCs* ±90% MAP2+	Zhang *et al.* (2013)
Dopaminergic neurons	Mouse	Astrocytes (*MAP2::CD4* and *MAP2::GCaMP3*) and fibroblasts	Monolayer	*ASCL1, LMX1B* and *NURR1*	DOX	*Astrocytes* 35.1 ± 1.5% TUJ1+ 50.9 ± 3.3% TH+/TUJ1+ 18.2 ± 1.5% TH+	Addis *et al.* (2011)

Table 2. *(Continued)*

Transcription factor-mediated reprogramming							
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Transcription factors**	**Chemical driving factors**	**Phenotypic markers (% cells)**	**References**
Dopaminergic neurons	Mouse	Astrocytes (*MAP2::CD4* and *MAP2::GCaMP3*) and fibroblasts	Monolayer	*ASCL1, LMX1B* and *NURR1*	DOX	*Fibroblasts* 14.9 ± 2.3% TUJ1+ 9.1 ± 0.9% TH+	Addis *et al.* (2011)
		Fibroblasts (*Pitx3::EGFP*)	Monolayer	*Ascl1, En1, Foxa2, Lmx1a, Nurr1* and *Pitx3*	DOX, FGF8 and SHH	9.1% Pitx3-EGFP+	Kim *et al.* (2011b)
	Mouse and human	Mouse fibroblasts (*TH::GFP*) and human fibroblasts	Monolayer	*Ascl1, Lmx1a* and *Nurr1*	DOX	*Mouse* ±22% TUJ1+ ±17% TH+ *Human* 10 ± 4% TUJ1+ 6 ± 2% TH+	Caiazzo *et al.* (2011)
	Human	Fibroblasts	Monolayer	*Ascl1, Brn2, Myt1l, Foxa2* and *Lmx1a*	DOX	±15% TUJ1+/MAP2+ ±10% TH+/TUJ1+/MAP2+	Pfisterer *et al.* (2011)
		iPSCs	Monolayer	*Ascl1, Lmx1a* and *Nurr1*	DOX	*Fetal fibroblast-derived iPSCs* 51 ± 4% TUJ1+ 65 ± 5% TH+/TUJ1+ ±30% CALB1+/TH+ ±40% GIRK2+/TH+ *Parkinson's disease patient-derived iPSCs* 48 ± 4% TUJ1+ 26 ± 3% TH+/TUJ1+	Theka *et al.* (2013)
GABAergic neurons	Mouse	Fibroblasts	Monolayer	*Ascl1, Brn2, Dlx1, Dlx2, Lbx1, Lhx1, Lhx2, Myt1l, Pax2, Pitx2* and *Pft1a*	None	<35 ± 4% GABA+/TUJ1+	Wasko (2013)

Table 2. *(Continued)*

Transcription factor-mediated reprogramming							
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Transcription factors**	**Chemical driving factors**	**Phenotypic markers (% cells)**	**References**
GABAergic neurons	Mouse	Astrocytes	Monolayer and neurosphere	*Dlx2*	None	*Monolayer* 35.9 ± 13.0% TUJ1+ 33.7 ± 3.6% VGAT+	Heinrich *et al.* (2010)
	Mouse and human	Mouse fibroblasts (*GAD67::GFP*), human fibroblasts and human iPSCs	Monolayer	*Ascl1, Dlx5, Bf1, Lhx6* and *Sox2*	DOX	*Mouse fibroblasts* ±94% GAD65/67+/GAD67-GFP+ ±97% GABA+/GAD67-GFP+ ±93% PV+/GAD67-GFP+ ±3% SST+/GAD67-GFP+ *Human fibroblasts* ±70% GABA+/TUJ1+ ±90% PV+/TUJ1+ *Human iPSCs* ±50% GABA+/MAP2+ ±90% PV+/MAP2+ ±2% SST+/GABA+	Colasante *et al.* (2015)
Serotonergic neurons	Human	Fibroblasts	Monolayer	*ASCL1, FOXA2, FEV, LMX1B* and *hp53shRNA*	DOR, DOX, PD0332991 and SB431542	<49% TUJ1+ <23% 5-HT+	Xu *et al.* (2015)
		Fibroblasts (*TPH2::GFP* and *SYN1::dsRed*)	Monolayer	*ASCL1, FEV, GATA2, LMX1B, NGN2* and *NKX2-2*	A83-01, CHIR99021, DOX, Forskolin, LDN193189, NOG and SB431542	58.4 ± 4.2% TUJ1+ ±60% MAP2ab+ 61 ± 15% TPH+/MAP2ab+ 38 ± 2% 5-HT+	Vadodaria *et al.* (2015)
Cholinergic motor neurons	Mouse	ESCs	EB	*Isl1, Lhx3, Ngn2* and *Phox2a*	DOX	*Ngn2, Isl1 and Lhx3* 99.82 ± 0.17% HB9+/ISL1+ 0.24 ± 0.28% PHOX2B+/ISL1+	Mazzoni *et al.* (2013)

Table 2. *(Continued)*

Transcription factor-mediated reprogramming							
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Transcription factors**	**Chemical driving factors**	**Phenotypic markers (% cells)**	**References**
Cholinergic motor neurons	Mouse	ESCs	EB	*Isl1, Lhx3, Ngn2* and *Phox2a*	DOX	*Ngn2, Isl1, Phox2a* 0.11 ± 0.11% HB9+/ISL1+ 99.03 ± 0.08% PHOX2B+/ISL1+	Mazzoni *et al.* (2013)
	Mouse and human	ESCs	Monolayer (mouse) and EB (human)	*Phox2a* and *Phox2b*	FGF2, FGF8b and Hh-Ag1.3	*Mouse* 61% PHOX2B+/ISL1+/TUJ1+	Mong *et al.* (2014)
		Mouse fibroblasts (*Hb9::GFP*) and human fibroblasts	Monolayer	*Ascl1, Brn2, Myt1l, Lhx3, Isl1, Hb9, Ngn2* and *NEUROD1*	None	*Mouse* 5-10% Hb9-GFP+ 97.6% VACHT+/Hb9-GFP+	Son *et al.* (2011)
	Human	ESCs (*Hb9::GFP*) and iPSCs	EB	*Isl1, Lhx3* and *Ngn2*	Forskolin, RA and SHH	*ESCs* 55% Hb9-GFP+ ±50-62% HB9+/CHAT+ *iPSCs* 49-72% HB9+/CHAT+	Hester *et al.* (2011)
Noradrenergic neurons	Mouse and human	ESCs	Monolayer (mouse) and EB (human)	*Phox2b*	BMP5, BMP7, Cyclopamine, FGF2 and FGF8b	*Mouse* 17.9% TH+/PHOX2A+/TUJ1+	Mong *et al.* (2014)

Neuronal phenotypes, species, starting cell types, culture methods, transcription factors, chemical driving factors and representative phenotypic markers that have been used to assess the differentiation efficiency and culture homogeneity are broadly summarized. + indicates the percentage of cells in the population that stained positive for a certain marker. *Abbreviations*: 5-HT, serotonin (5-hydroxytryptamine); A83-01, TGF-β kinase/activin receptor-like kinase inhibitor; ASCL1, achaete-scute homolog 1 (MASH1/HASH1); Bf1, brain factor 1/forkhead box protein G1 (FOXG1); BMP5, bone morphogenic protein 5; BMP7, bone morphogenic protein 7; Brn2, brain-specific homeobox/POU domain protein 2 (POU3F2); CALB1, calbindin 1; CD4, cluster of differentiation 4; CHAT, choline o-acetyltransferase; CHIR99021, GSK3β inhibitor; Dlx1, distal-less homeobox 1; Dlx2, distal-less homeobox 2; Dlx5, distal-less homeobox 5; DOX, doxycycline; dsRed, discosoma sp. red fluorescent protein; EGFP, enhanced green fluorescent protein; En1, homeobox protein engrailed 1; ESCs, embryonic stem cells; FEV, ETS transcription factor (PET1); FGF2, fibroblast growth factor 2/basic fibroblast growth factor (bFGF); FGF8, fibroblast growth factor 8; FGF8b, fibroblast growth factor 8 isoform b; FOXA2, forkhead box protein A2; GABA, γ-aminobutyric acid; GAD65/67, glutamic acid decarboxylase isoform 65/67 (GAD2/1); GAD67, glutamic acid decarboxylase isoform 67 (GAD1); GATA2, GATA binding protein 2; GCaMP3,

a GFP-based calcium sensor for imaging calcium dynamics; GFP, green fluorescent protein; GIRK2, G protein-activated inward rectifier potassium channel 2 (KCNJ6); HB9, homeobox HB9/motor neuron and pancrease homeobox 1 (MNX1); Hh-Ag1.3, small molecule agonist of SHH signalling; hp53shRNA, human p53 small hairpin RNA; iPSCs, induced pluripotent stem cells; ISL1, ISL LIM homeobox 1; Lbx1, ladybird homeobox 1; LDN193189, selective BMP signalling inhibitor; Lhx1, LIM homeobox 1; Lhx2, LIM homeobox 2; Lhx3, LIM homeobox 3; Lhx6, LIM homeobox 6; Lmx1a, LIM homeobox transcription factor 1 alpha; LMX1B, LIM Homeobox transcription factor 1 beta; MAP2, microtubule-associated protein 2; MAP2ab, microtubule-associated protein 2ab; Myt1l, myelin transcription factor 1 like; NEUROD1, neurogenic differentiation 1; NGN2, neurogenin 2; NKX2-2, NK2 homeobox 2; NOG, Noggin; NURR1, nuclear receptor related 1 protein; Pax2, paired box 2; PD0332991, cyclin-dependent kinase 4/6 inhibitor; Pft1a, pancreas specific transcription factor 1a; PHOX2A, paired-like homeobox 2a; PHOX2B, paired-like homeobox 2b; Pitx2, paired-like homeodomain 2; Pitx3, paired-like homeodomain 3; PV, parvalbumin; RA, retinoic acid; SB431542, transforming growth factor beta inhibitor; SHH, sonic hedgehog; Sox2, SRY box 2; SST, somatostatin; SYN1, synapsin 1; Tau, microtubule-associated protein tau (MAPT); TBR1, T-box brain 1; TH, tyrosine hydroxylase; TPH, tryptophan hydroxylase; TPH2, tryptophan hydroxylase 2; TUJ1, neuron-specific class III beta-tubulin (TUBB3); VACHT, vesicular acetylcholine transporter; VGAT, vesicular GABA transporter; VGLUT1, vesicular glutamate transporter 1; VGLUT2, vesicular glutamate transporter 2.

Transcription factor-mediated reprogramming

Cellular reprogramming and transdifferentiation

For long it was thought that cellular differentiation and lineage commitment were irreversible processes established during embryonic development [127]. However, the cloning of animals by nuclear transfer demonstrated that maturated molecular mechanisms are reversible and that a nucleus from the most differentiated cell bears the potential to generate an organism [127]. These cell fusion experiments have proved that transcriptional reprogramming can occur by exposing a distinct nucleus to cytoplasmatic components of cells from distinct lineages, although the exact mechanisms underlying these processes remain challenging to address. Substantial interest in transcriptional reprogramming has been rejuvenated upon the discoveries by Dr. Takahashi and Prof. Dr. Yamanaka, who provided the foundation that somatic cells can be reprogrammed to iPSCs. The cells were initially generated by reprogramming fibroblasts via recombinant overexpression of four transcription factors, including Oct3/4, Sox2, Klf4, and Myc (OSKM, Yamanaka factors) [128]. The use of only four transcription factors was sufficient to induce dramatic cell fate changes and to reprogram fully differentiated cells into a more embryonic cell state. The derivation of iPSCs has been substantially adapted and improved by using other sets of transcription factors, including LIN28 and NANOG [129], by introducing non-integrative transgene expression and by using different types of somatic cells [130, 131]. Earlier work on the other hand, has demonstrated that increased activity of a single transcription factor, namely Myod1, is sufficient to directly convert fibroblasts into myocytes by a process known as transdifferentiation [132]. This has supported the notion that cell fate conversions can be direct without the need of precedent dedifferentiation. Consequently, these studies have raised the question whether transcription factor-mediated reprogramming could also directly induce neuronal fates in somatic cells or even PSCs (See Fig. 3). Notably, and especially important in the context of disease modelling, these direct conversion modalities may prove to be invaluable in the study of late-onset neurodegenerative disorders because the age of somatic cells is maintained in the converted neurons, thus allowing to model the aging process in vitro [133]. In addition, while chemically defined differentiation protocols in PSCs are known for their long multistep protocols, PSCs exposed to specific sets of transcription factors have shown to differentiate much faster without additional culturing steps, therefore providing an appealing, simpler and possibly more effective differentiation strategy for PSCs in vitro [134]. One of the major disadvantages of direct somatic cell reprogramming compared to directed differentiation of PSCs however, is that the former skips the pluripotent state and does not allow expansion of the cells before further applications, thereby limiting the accessibility of cells readily available in vitro. Nevertheless, somatic cell reprogramming into

an intermediate state that allows substantial proliferation, such as NPCs, has already been considered [135, 136]. For instance, mouse and human fibroblasts have been partially reprogrammed by introducing the four Yamanaka factors, which gave rise to NPCs that were capable of differentiating into both neuronal and glial cells in the presence of leukemia inducible factor (LIF) and FGF2 [137]. The direct differentiation of partially reprogrammed cells may be useful for rapidly preparing high numbers of NPCs that could be expanded before terminal differentiation into target neurons.

Induced glutamatergic neurons

Starting from a pool of nineteen candidates, Vierbuchen et al. [138] were the first to identify a combination of only three transcription factors, including Brn2, Ascl1, and Myt1l (BAM), which could rapidly and efficiently convert mouse embryonic fibroblasts (MEF) into induced neuronal (iN) cells (See Table 2). These iN cells expressed multiple neuron-specific markers, generated action potentials and were able to form functional synapses [138]. Electrophysiological recordings demonstrated that mainly excitatory postsynaptic potentials could be recorded, providing functional proof that a large majority of the iN cells exhibited a glutamatergic phenotype. Some cells also expressed GABAergic markers at earlier time points, including Gaba and Gad67, suggesting that both neuronal subtypes could be obtained but culture conditions probably favoured the glutamatergic phenotype [127, 138]. These findings in murine somatic cells led to follow-up experiments using human fibroblasts, which eventually led to the successful generation of human iNs with the addition of Neurod1 to the BAM pool [139]. Just like the generation of iPSCs, during that time it was thought that a combination of factors was necessary to fully reprogram iN cells from fibroblasts and the use of a single transcription factor was considered insufficient [140]. However, later it was shown that Ascl1 alone is satisfactory to generate populations of pre-dominantly glutamatergic iN cells from mouse and human fibroblasts, as well as ESCs [140, 141]. A clear hierarchical role of the reprogramming factors has, therefore, been suggested, demonstrating that ASCL1 acts as a key factor to activate the neuronal program, whereas access of BRN2 to the chromatin is apparently more cell-context-dependent and facilitates reprogramming later on [140]. Moreover, although Ascl1 alone is sufficient to generate iNs, endogenous Mytl is subsequently induced during reprogramming, and exogenous Myt1l has, therefore, demonstrated to greatly improve the efficiency of reprogramming and the functional maturity of the resulting iN cells [141]. In fact, it has been shown that MYT1L exerts its proneuronal function by direct repression of many different somatic lineage programs except the neuronal program. This repressive function of MYT1L is mediated via recruitment of a complex containing SIN3B by binding to a previously uncharacterized N-terminal

domain [141]. In addition, knockdown of Myt1l in primary postmitotic neurons removed the repression of non-neuronal programs and impaired neuronal gene expression and function, indicating that many somatic lineage programs are actively and persistently inhibited by MYT1L to maintain neuronal identity [141]. Aside from fibroblast, for murine astrocytes [142], mouse ESCs [140, 143] and human PSCs [17, 140], it has also been shown that single neurogenic factors, such as Neurod1 and Ngn2, alone are sufficient to rapidly induce the neuronal fate. Although Ascl1-induced iN cells displayed slower maturation kinetics at early developmental stages, their functional properties and neuronal gene-expression profile at later time points are surprisingly similar to that of Ngn2 or BAM iN cells [140].

After the establishment of the BAM pool and the other neurogenic transcription factors, researchers have been attempting to derive other neuronal subtypes by transcription factor-mediated reprogramming and, thanks to that, additional factors have been identified with the ability to induce dopaminergic neurons, GABAergic neurons, serotonergic neurons, cholinergic neurons and adrenergic neurons (See Table 2) [144-148]. Identical to the chemically defined systems, all of these obtained cultures have been subjected to various bioassays and assessments in order to examine their population characteristics (See Table S2). Furthermore, studies have demonstrated that supplementation of chemically defined systems with transcription factor-mediated reprogramming can significantly increase the efficiency of obtaining differentiated neuronal cells, as well as vice versa [145, 149, 150]. Notably, blockade of TGFb/SMAD signalling using Noggin and molecules such as SB431542 and LDN193189, as well as pharmalogical promotion of calcium signalling with cAMP and Forskolin have not only shown to increase iN yield, but have also been used to successfully generate transgene-free iNs [110, 151, 152]. The fast progress in the field of chemically-mediated reprogramming and transdifferentiation provides us new ways to manipulate neuronal fates both in vitro and in vivo. These methods on their own and/or in combination with other approaches may accelerate the eventual applications of patient-specific human neurons generated in vitro, by facilitating the potency and timelines of the protocols, and by aiding in the specification of regional subtypes within neurotransmitter classes.

Induced dopaminergic neurons

By combining the BAM factors with Lmx1a and Foxa2, which are typically expressed in midbrain dopaminergic NPCs, human fibroblasts have been converted into induced dopaminergic (iDA) neurons [146]. This provided proof-of-principle that other subtypes of iN cells can be produced by transcription factor-mediated fate instructions. Since then, iDA neurons have been obtained by

ectopic overexpression of various combinations of transgenes encoding midbrain-specific transcription factors (See Table 2). For instance, a minimal set of three transcription factors, i.e. Ascl1, Nurr1 and Lmx1a, was sufficient to generate functional iDA neurons from murine and human fibroblasts [153]. The three factors were able to elicit dopaminergic neuronal conversion, resulting in iDA neurons that were highly enriched in genes of the dopaminergic phenotype, released dopamine, exhibited proper electrophysiological profiles, and, in case of the murine iDA neurons, were able to integrate into neonatal mouse brains. Interestingly, the endogenous Th and Vmat2 promoter regions were highly demethylated in the iDA neurons, whereas they were fully methylated in the fibroblasts, indicating their epigenetic reactivation during dopaminergic conversion [153]. In addition to fibroblasts, astrocytes have also been efficiently converted into iDA by using a single polycistronic vector containing ASCL1, LMX1B and NURR1 [154]. The fact that different somatic cells can be reprogrammed towards iDA using similar sets of transcription factors highlights their importance in reprogramming processes and reassigning cell fate. Reprogramming mouse fibroblast based only on Ascl1 and Pitx3 in another study resulted in immature iDA neurons after 4 weeks of culture [150]. However, inclusion of additional factors, such as En1, Foxa2, Lmx1a, and Nurr1, could fully reprogram fibroblasts into iDA neurons that were more similar at the molecular level to bona fide dopaminergic neurons [150]. Furthermore, the murine iDA neurons were able to alleviate symptoms in a mouse model of Parkinson's disease, demonstrating their therapeutic potential for transplant therapies [150]. Although transdifferentiation of fibroblasts in the latter study could be achieved by ectopic overexpression of only 2 factors, the results suggested that additional factors are required to ensure proper maturation of the iDA neurons [150]. Furthermore, by adding SHH and FGF8 to the culture media, iDA neuron reprogramming could be enhanced up to 2 fold when using only 2 factors, and up to 3 fold when using the combination of 6 factors [150]. These findings demonstrate that patterning molecules can act as critical enhancing components in promoting the generation of iDA neurons from fibroblasts. Notably, Theka *et al.* [155] have established a fast protocol to obtain dopaminergic neurons by overexpressing Ascl1, Nurr1, and Lmx1a in human iPSCs. They were able to generate mature and functional dopaminergic neurons in as few as 21 days, avoiding all the intermediate steps of induction and selection of EBs and NPCs. Strikingly, the resulting neuronal conversion process was very efficient, since approximately 93% of all the co-infected iPSCs were forced to differentiate into postmitotic iDA neurons [155]. The iPSC-derived neurons expressed all the critical molecular markers of midbrain dopaminergic neurons at the molecular level and exhibited sophisticated functional features, including spontaneous electrical activity and dopamine release [155].

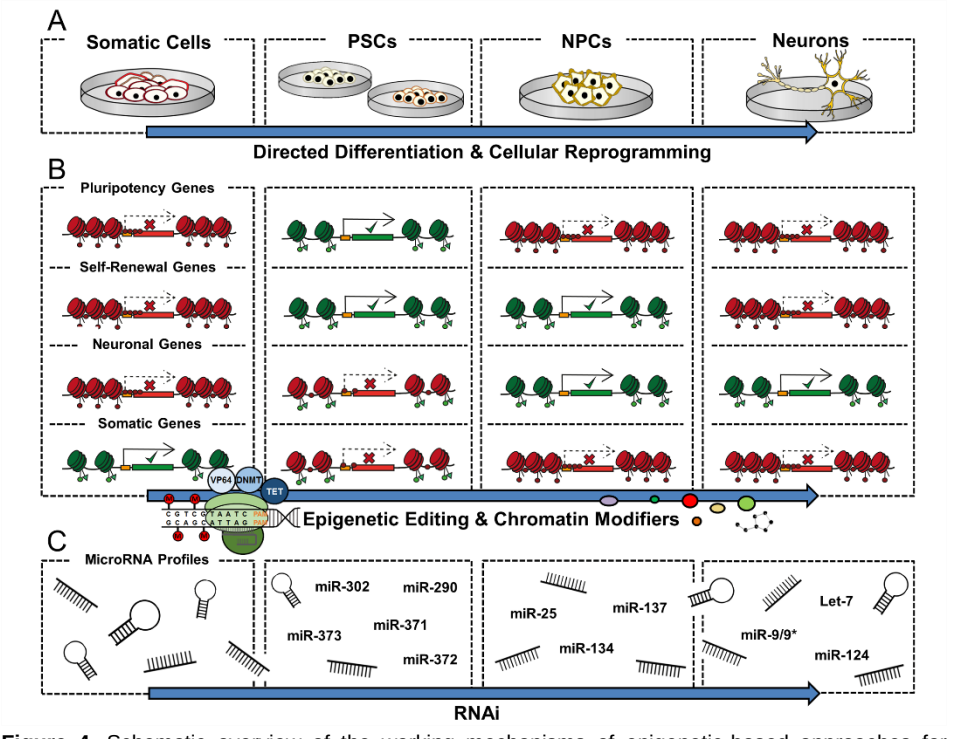


Figure 4. Schematic overview of the working mechanisms of epigenetic-based approaches for neuronal differentiation *in vitro*. **A.** The figure shows how somatic cells, pluripotent stem cells (PSCs) and neural precursor cells (NPCs) can be differentiated or reprogrammed into neurons. **B.** In each of these cells, histone and DNA modifications orchestrate transcriptional activity of pluripotency genes, self-renewal genes, neuronal genes and somatic genes, which define their cellular identity. While somatic genes are only active in somatic cells (in green), genes that regulate pluripotency and self-renewal are activated in PSCs (in green), whereas genes that regulate neuronal differentiation are repressed and poised (in red with green modifications) for activation upon further developmental cues. This activation and repression is inverted upon initiation of neural induction and differentiation, leading to the activation of neuronal genes (in green) and repression of PSC genes (in red). By using epigenetic editing and molecules that allow modifying the chromatin structure, activity of genes that redefine and reinforce cellular identity can be altered, offering control over neuronal fate determination *in vitro* by the natural dynamics of endogenous gene regulation. **C.** Aside from chromatin modifications, it has been proposed that the cellular microRNA (miRNA) milieu is unique in each cellular subtype and required to facilitate developmental transitions during neuronal differentiation. For PSC regulation, miRNAs such as miR-371 and miR-302 are important, while miR-9/9* and miR-124 belong to brain-enriched miRNAs that are activated during neuronal specification. By using RNA interference (RNAi), miRNA profiles that redefine and reinforce cellular identity can be manipulated, leading to changes in transcriptional regulation that could aid neuronal differentiation and specification *in vitro*.

Induced GABAergic neurons

In the first study that achieved to obtain iNs [138], some cells expressed markers of GABAergic differentiation [127]. However, whether enriched populations of induced GABAergic interneurons (iGNs) can be obtained has not been addressed

until recently (See Table 2). Wasko [156] was the first to demonstrate that mouse fibroblasts could be directly reprogrammed to iGN-like cells using different pools of transcription factors, including Dlx1, Dlx2, Lbx1, Lhx1, Lhx2, Pax2, Pitx2 and Pft1a in combination with members of the BAM factors. The author states that different groups containing these transcription factors have demonstrated some capacity for the derivation of iGNs, although the most efficient factor combination remains to be determined. Additional experiments will be required to fully characterize the efficiency of the different transcription factor pools, as well as to assess the functional properties and maturity of the eventual derived iGNs. Aside from fibroblasts, iGNs neurons have also been derived by overexpressing the ventral telencephalic fate determinant Dlx2 in murine astrocyte cultures [142]. Interestingly, they found that the overall efficiency of Dlx2-mediated neuronal reprogramming towards iGNs is much lower compared to Ngn2-mediated reprogramming towards induced glutamatergic neurons, suggesting that cortical astrocytes possess a higher competence to respond to the dorsal telencephalic fate determinant [142]. In addition, five different factors have been identified, including Ascl1, Bf1, Dlx6, Lhx6 and Sox2, which were able to convert mouse and human fibroblast, as well as human iPSCs, into iGNs that possessed characteristics of telencephalic GABAergic interneurons [144]. Molecular profiling showed pronounced activation of forebrain-specific (epi)genetic markers, required for GABAergic fate specification [144]. Furthermore, the iGNs displayed progressively maturing firing patterns comparable to cortical GABAergic interneurons, formed functional synapses, and released GABA upon stimulation [144]. The iGNs also survived and matured upon engraftment into the mouse hippocampus and optogenetic stimulation demonstrated functional integration of the grafted iGNs into the host circuitry, triggering inhibition of host granule neuron activity [144]. This latter study also elegantly demonstrated how human PSCs can be harnessed to generate GABAergic neurons.

Induced serotonergic neurons

There have only been two studies published that were able to directly obtain induced serotonergic (i5HT) neurons by transcription factor-mediated reprogramming of somatic cells (See Table 2). In the first study [148], human fibroblasts could be directly converted to i5HT neurons by the ectopic expression of ASCL1, FEV, FOXA2, and LMX1B. The transdifferentiation was enhanced by p53 knockdown and appropriate culture conditions, including hypoxia [148]. Addition of the small-molecule compounds dorsomorphin (DOR), SB431542, ROCK inhibitor Y-27632 and the CDK4/6 inhibitor PD0332991 improved both the conversion efficiency and morphology of the obtained i5HT neurons, whereas addition of brain derived neurotrophic factor (BDNF) and glial derived neurotrophic factor (GDNF) slightly improved the morphology but not the

conversion efficiency [148]. The i5HT neurons expressed markers for mature serotonergic neurons, had calcium-dependent 5-HT release and selective 5-HT uptake, and exhibited spontaneous action potentials, as well as spontaneous excitatory postsynaptic currents [148]. Moreover, application of 5-HT significantly increased the firing rate of spontaneous action potentials [148]. In the second study, it was demonstrated that overexpression of the transcription factors FEV, GATA2, LMX1B and NKX2-2 in combination with ASCL1 and NGN2 directly and efficiently generated i5HT neurons from human fibroblasts [117]. The i5HT neurons showed increased expression of specific serotonergic genes known to be expressed in the raphe nuclei, displayed spontaneous action potentials, released 5-HT in vitro and functionally responded to selective serotonin reuptake inhibitors (SSRIs) [117].

Induced cholinergic motor neurons

Transcription factor mediated reprogramming has also been used to derive cholinergic motor neurons (See Table 2). Son et al. [147] reported that the forced expression of the BAM factors, in combination with Hb9, Isl1, Lhx3 and Ngn2 was sufficient to convert mouse fibroblasts into induced motor neurons (iMNs). The iMNs were identified based on an HB9::GFP reporter and exhibited a morphology, gene expression signature, electrophysiological profile, synaptic functionality, in vivo engraftment capacity and sensitivity to degenerative stimuli, similar to ESC-derived cholinergic motor neurons [147]. By adding NEUROD1 to the 7 transcription factors they demonstrated that also human fibroblasts could be converted to iMNs [147]. Other studies have also coupled chemically defined systems with transcription factor-mediated reprogramming in PSCs in order to improve the efficiency and timing to obtain cholinergic motor neurons. For instance, adenoviral delivery of Isl1, Lhx3 and Ngn2 combined with exposure to RA and SHH signalling allowed rapid and efficient (> 55%) acquisition of electrophysiological active human iMNs within 11 days [149]. In a separate study, the same set of transcription factors were used to sufficiently differentiate mouse ESCs to a cholinergic motor neuron identity [157]. Interestingly, replacement of Lhx3 by Phox2a led to specification of cranial, rather than spinal motor neurons, emphasizing the possibility to direct subtype specification. Phox2a and Phox2b have also been used to generate both visceral motor neurons and noradrenergic neurons from mouse ESCs in the presence of appropriate patterning molecules [145]. Culturing Phox2b expressing cells under the signalling influence of FGF8 and BMPs promoted the generation of enriched noradrenergic cultures, while culturing Phox2a or Phox2b expressing cells with FGF8 and SHH generated cholinergic motor neurons instead [145]. The authors demonstrated that the obtained neurons were suitable for drug testing in vitro and, therefore, harbour the potential for the discovery of therapeutic interventions. To conclude, these

studies add up to the growing body of protocols, allowing to produce clinically relevant neuronal cells and demonstrate that combinations of distinct methods can work synergistically in obtaining the desired neuronal subtypes.

Epigenetic-based approaches

Epigenetics in directed differentiation and cellular reprogramming

Even though every cell in the human brain shares an identical genotype, it consists of many neuronal subtypes with distinct, yet stable, profiles of gene expression patterns. As addressed above, this diverse repertoire of neurons is produced by extrinsic patterning cues and lineage-specific transcription factors that define and reinforce these neuronal subtype-specific expression patterns. The neuronal phenotypes are further stabilized by the epigenetic machinery that maintains their genetic profile over a lifetime [138, 158]. Epigenetics can be defined as stable and heritable modifications on the chromatin that occur without changes in the underlying DNA sequence [159]. It is generally accepted that the epigenetic machinery includes multiple levels of transcriptional control by (re)organizing the chromatin structure and architecture [160, 161]. Histone modifications and DNA methylation are the most predominant examples of chromatin modifications that have been studied over the last years. Both of these modifications play an important role in neuronal cell fate determination and differentiation [24]. In continuously self-renewing ESCs, genes that regulate pluripotency are activated, whereas genes that regulate neuronal differentiation are repressed in a stable and heritable manner over many cell divisions [162-165]. Moreover, this activation and repression needs to be inverted upon initiation of neural induction and differentiation. In addition to these chromatin modifications, noncoding RNAs are other important modulators that regulate gene expression patterns at the post-transcriptional level [161]. A growing body of evidence is revealing that these also represent strong mediators of neuronal cell fate determination [166, 167]. Examples of noncoding RNAs include small interfering RNAs (siRNAs), small modulatory RNAs, ribosomal RNAs, transfer RNAs, natural antisense transcripts, enhancer RNAs and microRNAs (miRNAs), but many other subclasses have been described [161, 168, 169]. Based on the key role of the epigenetic machinery in assigning neuronal fate and identity, it has currently become evident that the underlying mechanisms of both directed differentiation and direct cellular reprogramming encompass epigenetic phenomena [23, 170]. A significant degree of transcriptional regulation takes place, where epigenetic mechanisms communicate with each other in collaboration with the extrinsic patterning cues and transcription factors to guide neuronal cell fate conversions. The profound epigenome remodelling processes in PSCs and somatic cells eventually lead to the acquisition and stabilization of neuronal subtypespecific gene expression profiles, which reinforce their cellular

phenotypes. Although the epigenetic machinery has been extensively studied in ESC differentiation along the neural lineage, please refer to Hirabayashi and Gotoh [171], little is known about the extent in which these epigenetic mechanisms are identical in iPSC differentiation. The field is still in its infancy and we are just starting to understand similarities and differences in epigenetic and transcriptional states between iPSCs and ESCs [172-174]. In addition, the underlying epigenetic mechanisms of direct reprogramming in somatic cells are also still incompletely understood and the interactions between transcription factors and the chromatin architecture are currently under investigation [23, 175]. Nevertheless, considering the crucial role of chromatin modifications and transcriptional regulations by non-coding RNAs in modulating neuronal cell fate and identity, (re)programming and interfering with the epigenetic machinery offers an alternative approach for directing neuronal fates in vitro (See Fig. 4, Tables 3 and S3).

Chemical compounds targeting chromatin-remodelling proteins

First proof-of-concept on epigenetic-based approaches for neuronal differentiation in vitro came from experiments where beneficial effects on direct cellular reprogramming by chemical compounds that target chromatin-remodelling proteins were found (See Table 3). Genome-wide profiling of PSCs and their differentiated progeny has suggested a global, progressive transition from euchromatin to heterochromatin at various loci during differentiation [176]. The epigenetic state of a differentiated cell is, therefore, considered to be more condensed and has significantly less dynamic exchange with transcriptional regulators [176]. However, by targeting remodelling proteins, heterochromatin enriched regions can be altered and recruitment of transcriptional activators can be orchestrated to DNA sites that were previously inaccessible, allowing them to induce dramatic cell fate changes. As a consequence, epigenetic research in PSC differentiation and somatic cell reprogramming has supplied a plethora of potential drug targets to improve neuronal cell fate conversions. For instance, NPCs were induced from mouse fibroblasts by ectopic overexpression of Pax6 and Bf1 [136]. Treatment with a combination of small molecules that inhibit histone deacetylases (HDACs), H3K27 methyltransferases, and H3K4me2 demethylases accelerated the direct conversion of fibroblasts into NPCs up to ten times [136]. Remarkably, simultaneous inhibition of BMP- and TGF-β-signalling almost doubled the frequency of NPCs, again demonstrating that combinations of different approaches can significantly enhance the derivation of the desired cell types in vitro. HDAC inhibitors and histone demethylase inhibitors coupled with other appropriate chemical patterning cues have also shown to turn mouse fibroblasts and astrocytes, as well as human urinary cells, into NPCs or neurons under physiological hypoxia conditions and without the need of additional

transgene overexpression [177, 178]. Analysis of global gene expression patterns revealed a high degree of similarity between the induced NPCs and control NPCs [178], whereas the neurons induced from cultured astrocytes were electrophysiological active and expressed various subtype specific markers for dopaminergic neurons, GABAergic neurons, glutamatergic neurons and cholinergic motor neurons [177]. The delivery of small molecules or systematic administration of drug cocktails enabling astrocytic-to-neuronal conversions bear the potential for direct induction of desirable cells from resident astrocytes in situ, while bypassing possible adverse effects of genome integrating constructs. Taken together, drugs targeting the chromatin can improve reprogramming efficiency and might function as useful adjuvants in currently used reprogramming protocols, thereby providing a possible alternative strategy to produce patient-specific neuronal cells.

Epigenetic editing

Epigenetic editing systems [27, 29] offer an alternative tool to supplement, or in some cases even replace components of, current widely used directed differentiation and cellular reprogramming protocols [32]. Making locus-specific alterations to the epigenetic code allows to (re) shape the mechanistic relationships among chromatin state, gene regulation, and cellular phenotype by the natural dynamics of gene expression [29]. For this reason, these epigenetic editing systems allow probing of signatures responsible for cellular identity and provide intelligent control to direct neuronal cell fates in PSCs and somatic cells [179-182]. DNA-targeting platforms based on the initially established zinc finger proteins (ZFPs), transcription activator-like effector nucleases (TALENs) and the clustered, regularly interspaced, short palindromic repeats (CRISPR)-Cas9 systems, have allowed the recruitment of transcriptional modulators and epigenome-modifying factors to any genomic locus [27, 29, 183]. Virtually any DNA sequence can be targeted with these customizable synthetic epigenetic tools [29]. The direct fusion of transcriptional effector domains to designed DNA-targeting domains can induce transcriptional activation or repression of endogenous key-lineage-determinant genes [29]. Transcriptional effector domains include epigenetic effectors that directly catalyse covalent modifications to DNA or histones, or that recruit other histone modifying enzymes, as well as interfere with chromatin-binding proteins. For instance, epigenetic effectors that directly catalyse covalent modifications to DNA, such as DNA methyltransferases (DNMTs) or ten-eleven translocation (TET) enzymes, can methylate or demethylate CpGs at non-neuronal and neuronal target promoters, leading to transcriptional repression or induction, respectively [28, 32]. Other effector domains such as VP64 can recruit histone remodelling factors, leading to increased chromatin accessibility and to the deposition of activating histone

modifications at desired neuronal loci [184, 185]. Alternatively, localization of DNA-targeting domains without an epigenetic effector to promoter regions or regions downstream of the transcription start sites can silence nonneuronal gene expression by steric hindrance of lineage-specific transcription factor binding and RNA polymerase elongation [32, 180]. Thus, custom epigenetic and transcriptional regulation by epigenetic editing-based approaches offer refined control over cell fate decisions, providing an invaluable tool for applications such as directed differentiation and cellular reprogramming along the neural lineage (See Table 3).

Several pioneer studies utilizing these epigenetic editing systems have shown successful and precise deposition or removal of different chromatin modifications to induce directed differentiation and cellular reprogramming for multiple cell types, including iPSCs, myocytes and neurons. Gao *et al.* [186] used TALE-based transactivators targeting distal enhancers of Oct4 in concert with Sox2, Klf4 and Myc transgene overexpression to generate mouse iPSCs. More recently, a protocol for the direct conversion of mouse fibroblasts to skeletal myocytes using a dCas9-based transactivator targeting the endogenous Myod1 gene has been developed [184]. Other groups have also applied dCas9-based transcriptional regulation to direct the differentiation of human PSCs [180, 187]. In an elegant study, human iPSCs were derived from human skin fibroblast by replacing OCT4 overexpression with dCas9-mediated activation of the endogenous promoter [187]. The authors demonstrated that directed endodermal differentiation of the iPSCs could be achieved by targeting proximal promoters of endodermal and pancreatic keyregulatory transcription factors, including FOXA2, SOX17, GATA4, PDX1, and NKX6-1 [187]. Directed neuronal reprogramming through multiplex endogenous gene activation using an engineered VP64-dCas9-VP64-based transcriptional activator has also been achieved [179]. Induced activation of the BAM factors successfully converted mouse fibroblasts to iN cells and their expression sustained in high levels during later stages of reprogramming despite the transient delivery of the guide RNAs (gRNAs) [179]. In a separate study, it was demonstrated that rapid and robust neuronal differentiation of human iPSCs could be achieved by targeting NGN2 and NEUROD1 with a VP64-p65-Rta-dCas9-mediated transcriptional activator [180]. Although many of the aforementioned reports include targeted activation of a single gene in combination of concurrent overexpression of multiple transcription factors, these latter examples also demonstrate that multiplex activation with a collection of gRNAs against a set of genes can be used to direct somatic cell reprogramming.

Table 3. Epigenetic-based neuronal conversion approaches *in vitro* per neuronal phenotype.

Chemical compounds targeting chromatin remodelling proteins								
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Epigenetic factors**	**Transcription factors**	**Chemical driving factors**	**Representative phenotypic markers (% cells)**	**References**
NPCs	Mouse	Fibroblasts (*Trp53+/+, Trp53−/−, Sox1+/+, Sox1EGFP/+, Tau+/+* and *TauEGFP/+*)	Monolayer	BIX-01294, t2PCPA and VPA	*Pax6* and *Bf1*	DOX, LDN193189 and SB431542	30% Sox1-EGFP+	Raciti *et al.* (2013)
	Mouse and human	Mouse fibroblasts and human urinary cells	Monolayer and neurospheres	NaB, TSA and VPA	None	CHIR99021, LIF, Li2CO3, LiCl, Repsox, SB431542 and Tranilast	*Mouse* <96% NES+ <96% SOX2+ <96% PAX6+ <93% NES+/SOX2+ <93% NES+/PAX6+	Cheng *et al.* (2015b)
Dopaminergic-, GABAergic-, glutamatergic- and cholinergic motor neurons	Mouse	Astrocytes (*GFAP::GFP* and *Neurod1::GFP*)	Monolayer	VPA	None	CHIR99021, FGF2, FGF8, Repsox, SHH and Tranilast	<13% DCX+ <15% NEUN+ ±5% GAD67+ ±4% CHAT+ ±3% TH+ ±6% VGLUT1+	Cheng *et al.* (2015a)
Epigenetic editing								
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Epigenetic factors**	**Transcription factors**	**Chemical driving factors**	**Representative phenotypic markers (% cells)**	**References**
Neurons	Mouse	Fibroblasts (*SYN1::RFP*)	Monolayer	VP64-dCas9-VP64 gRNAs: *Brn2, Ascl1,* and *Myt1l*	None	CHIR99021, LDN193189 and SB431542	±4% TUJ1+ ±75% MAP2+/TUJ1+	Black *et al.* (2016)
	Human	iPSCs	Monolayer	VP64-p65-Rta-dCas9 gRNAs: *NGN2* and *NEUROD1*	None	DOX	7% TUJ1+	Chavez *et al.* (2015)

Table 3. *(Continued)*

RNAi								
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Epigenetic factors**	**Transcription factors**	**Chemical driving factors**	**Representative phenotypic markers (% cells)**	**References**
Neurons	Mouse	NPCs	Monolayer and neurospheres	siRNA-REST	None	FGF2 and RA	<80% TUJ1+ <80% MAP2+	Low *et al.* (2012)
Dopaminergic neurons	Mouse	ESCs (*TH::GFP*)	EB and co-culture (PA6)	miR-132-ASOs	None	None	>25% TH+/MAP2+ 70% TH+/TH-GFP+	Yang *et al.* (2012)
		ESCs	EB	miR-133b-ASOs	None	FGF2	20% TH+/PITX3+	Kim *et al.* (2007)
	Human	NPCs	Monolayer	miR-124-inhibitor, miR-125b, miR-125b-mimic, miR-181a-mimic, miR-181a*-inhibitor and miR-181a/a*	None	FGF8b and SAG	<30% TUJ1+ <15% TH+	Stappert *et al.* (2013)
GABAergic neurons	Human	Fibroblasts (*SYN1::EGFP*)	Monolayer	miR-9/9*, miR-124 (*Bcl-xL)* and VPA	*CTIP2, DLX1, DLX2,* and *MYT1L*	DOX and RA	*Postnatal fibroblasts* 87% TUJ1+ 90% MAP2+ 72.3% GABA+/MAP2+ 80% BF1+/MAP2+ 60% DLX2+/MAP2+ *Adult fibroblasts* 82% MAP2+ 86% GABA+/MAP2+	Victor *et al.* (2014)
Glutamatergic- and GABAergic neurons	Human	Fibroblasts	Monolayer	miR-9/9*, miR-124 and VPA	*ASCL1, MYT1L* and *NEUROD2*	DOX and FGF2	80% MAP2+ ±24% TBR1+/MAP2+/TUJ1+ ±13% CTIP2+/MAP2+/TUJ1+ ±38% VGLUT1+/MAP2+/TUJ1+ ±30% GAD67+/MAP2+/TUJ1+ ±38% DLX1+/MAP2+/TUJ1+	Yoo *et al.* (2011)

Table 3. *(Continued)*

RNAi								
Phenotypes	**Species**	**Starting cell types**	**Culture methods**	**Epigenetic factors**	**Transcription factors**	**Chemical driving factors**	**Representative phenotypic markers (% cells)**	**References**
Glutamatergic- and GABAergic neurons	Human	Fibroblasts	Monolayer	miR-124 (IRES-RFP)	*BRN2* and *MYT1L*	Cumate, FGF2, NOG and DOX	*Postnatal fibroblasts* 55% MAP2+/RFP+ 46% NEUN+/RFP+ 8% GABA+/RFP+ 12% VGAT+/RFP+ *Adult fibroblasts* 28% RFP+/VGLUT1+	Ambasudhan *et al.* (2011)
Cholinergic motor neurons	Human	Fibroblasts	Monolayer	miR-9/9*, miR-124 (*Bcl-xL)* and VPA	*ISL1* and *LHX3*	DOX and RA	±80% TUJ1+ ±80% MAP2+ ±80% CHAT+/TUJ1+	Abernathy *et al.* (2017)

Neuronal phenotypes, species, starting cell types, culture methods, epigenetic factors, transcription factors, chemical driving factors and representative phenotypic markers that have been used to assess the differentiation efficiency and culture homogeneity are broadly summarized. + indicates the percentage of cells in the population that stained positive for a certain marker. *Abbreviations*: ASCL1, achaete-scute homolog 1 (MASH1/HASH1); ASOs, anti-sense oligonucleotides; Bcl-xL, b-cell lymphoma-extra large; BF1, brain factor 1/forkhead box protein G1 (FOXG1); BIX-01294, histone-lysine methyltransferase inhibitor; BRN2, brain-specific homeobox/POU domain protein 2 (POU3F2); CHAT, choline o-acetyltransferase; CHIR99021, GSK3β inhibitor; CTIP2, b-cell CLL/lymphoma 11b (BCL11B)/ COUP-TF-interacting protein 2 (COUP-TFII); DCX, doublecortin; DLX1, distal-less homeobox 1; DLX2, distal-less homeobox 2; DOX, doxycycline; EB, embryoid body; EGFP, enhanced green fluorescent protein; ESCs, embryonic stem cells; FGF2, fibroblast growth factor 2/basic fibroblast growth factor (bFGF); FGF8, fibroblast growth factor 8; FGF8b, fibroblast growth factor 8 isoform b; GABA, γ-aminobutyric acid; GAD67, glutamic acid decarboxylase isoform 67 (GAD1); GFAP, glial fibrillary acidic protein; GFP, green fluorescent protein; gRNAs, guide RNAs; iPSCs, induced pluripotent stem cells; IRES, internal ribosome entry site; LDN193189, selective BMP signalling inhibitor; Li2CO3, lithium carbonate; LiCl, lithiumchloride; LIF, leukemia inhibitory factor; MAP2, microtubule-associated protein 2; MYT1L, myelin transcription factor 1 like; NaB, sodium butyrate; NES, nestin; NEUN, neuronal nuclei antigen; NEUROD1, neurogenic differentiation 1; NEUROD2, neuronal differentiation 2; NGN2, neurogenin 2; NPCs, neural precursor cells; PA6, stromal cell line derived from newborn calvaria tissue of the C57BL/6 mice; PAX6, paired box 6; PITX3, paired-like homeodomain 3; RA, retinoic acid; REST, RE1-silencing transcription factor; RFP, red fluorescent protein; SAG, smoothened agonist; SB431542, transforming growth factor beta inhibitor; SHH, sonic hedgehog; siRNAs, small interfering RNAs; Sox1, SRY box 1; SOX2, SRY box 2; SYN1, synapsin 1; t2PCPA, trans-2-phenyl-cyclopropylamine hydrochloride; Tau, microtubule-associated protein tau (MAPT); TBR1, T-box brain 1; TH, tyrosine hydroxylase; Trp53, tumor protein p53; TSA, trichostatin A; TUJ1, neuron-specific class III beta-tubulin (TUBB3); VGAT, vesicular GABA transporter; VGLUT1, vesicular glutamate transporter 1; VP64-dCas9-VP64, dCas9 with N-terminal and C-terminal VP64 transactivation domains; VP64-p65-Rta-dCas9, dCas9-based transcriptional activator containing VP64, p65 and Rta; VPA, valproic acid.

Previous studies have demonstrated that targeted epigenetic editing of the regulatory elements controlling expression of lineage-specific transcription factors is sufficient for direct conversion between cell types, emphasizing the feasibility and potential advantages of using these synthetic epigenetic systems to direct neuronal cell fate of PSCs and somatic cells in vitro. Moreover, epigenetic editing in isolation has incredible promise as a platform for disease modelling both in vitro, as well as in vivo, whereas using it for transdifferentiation within the native physiological niche of the human brain might provide an alternative strategy to achieve cell fate conversions for applications in regenerative medicine [28, 188, 189]. Several labs have recently pioneered in vivo reprogramming in the brain and spinal cord by converting endogenous glial cells [190-192] and NPCs [193, 194] into functional neurons. For a more in-depth review on more surpassed studies, please refer to Li and Chen [195] and Srivastava and DeWitt [196]. Epigenetic editing systems applied in a safe and efficient manner that target similar lineage-specific transcription factors, might allow to induce any desirable neuronal subtype in vivo and could eliminate undesired issues in cell transplantations that may arise due to precedent in vitro cultures. To conclude, ground-breaking advances in this field are beginning to yield novel opportunities in the context of inducing neuronal phenotypes and bear excessive potential for many different applications in fundamental research and biomedicine.

RNA interference using miRNAs

On the one hand, the acquisition of distinct histone and DNA modifications at neuronal genes and non-neuronal genes plays a role in determining neuronal identity. On the other hand, neuronal identity is also determined by synergistic actions of extrinsic cues and the combined expression of transcription factors that are modulated by transcriptional regulators, including non-coding RNAs. Among these, miRNAs have been most extensively studied in relation to cellular identity and although most of the other subclasses are also anticipated to play important roles in regulating neuronal cell fate determination, the exact contribution of many remains elusive. Currently, it is accepted that miRNAs promote the transition from ESC self-renewal to differentiation by either directly suppressing the self-renewal state or by stabilizing the differentiated state [197]. In addition, multiple miRNAs target components or modulators of neural developmental signalling pathways, such as BMP and TGF-β signalling, and have been identified to either positively or negatively affect entry along the neural lineage [198]. In mature neurons, it has been proposed that the cellular miRNA milieu might even be unique in each subtype and required to facilitate developmental transitions during differentiation [199]. Some miRNAs even exhibit region-specific expression patterns in the brain, suggesting that neuronal subtypes residing in these regions may express

different miRNA profiles [198]. This is exemplified by a study of He *et al.* [200], in which they demonstrated substantial differences between the miRNA repertoire expressed in glutamatergic neurons and GABAergic interneurons that co-expressed either PV or SST. Specifically, miR-133b and miR-187 were found to be expressed higher in GABAergic interneurons as compared to glutamatergic neurons, where miR-133b was more abundant in PV expressing and miR-187 in SST expressing GABAergic interneurons [200].

Based on the emerging role of miRNAs during neural induction, neuronal differentiation and neuronal subtype specification, techniques such as RNAi using miRNAs holds great promise as an alternative tool to direct neuronal cell fate in vitro (See Table 3) [30, 31]. RNAi is a post-transcriptional gene silencing technique that has therapeutic opportunities for the treatment of various human disorders and has extensively been employed in translational studies to address fundamental biological questions[30]. The potential of RNAi lies in its capacity to virtually target any RNA molecule of interest, which allows fine-tuning of expression of key-determinant factors in neuronal fate determination. Approaches to downregulate endogenous miRNA expression to influence gene expression opposite of RNAi are also available and could for example be achieved by using anti-sense oligonucleotides (ASOs) (See Table 3) [201]. The temporal control of miRNA regulation might facilitate the induction of neuronal subtype-specific transcriptional networks and aids in recapitulating the natural dynamics of transcriptional regulation during neuronal cell fate commitment. Each miRNA has multiple molecular targets that might play essential roles in the derivation of specific neuronal subtypes and modulating a single miRNA could, therefore, guide entire neural developmental processes. Moreover, successful delivery of miRNAs could be achieved in many different ways depending on the needs of the experiments, as each method has different transfection efficiencies and transgene expression duration. Representative methods that allow expression of a miRNA construct include non-viral delivery systems such as lipid-based transfection, electroporation or the use of microvesicles, or viral delivery systems such as lentiviruses and adeno-associated viruses [202]. The potential of miRNAs to complement current practiced directed neuronal differentiation protocols from PSCs was first demonstrated by Kim *et al.* [203]. Interestingly, in this study they reported an unexpected negative impact of miR-133b on the generation of dopaminergic neurons from mouse ESCs. MiR-133b was found to be enriched in the human midbrain, while overexpression in this study impaired the generation of TH-positive cells [203]. Inhibition of miR- 133b on the other hand, resulted in an increased dopaminergic differentiation. The authors speculated that miR-133b regulates the maturation and function of midbrain dopaminergic neurons within a negative feedback circuit that includes the

dopaminergic transcription factor Pitx3. A similar negative impact on the differentiation of dopaminergic neurons from mouse ESCs has been reported in a separate study [204]. Inhibition of miR-132 promoted the differentiation of dopaminergic neurons, while ectopic expression of miR- 132 decreased the derivation of TH-positive cells without affecting the total number of neuronal cells. Through a bioinformatics assay they identified Nurr1 as a potential molecular target of miR-132, which also represents a key transcription factor of dopaminergic neuron specification. Stappert *et al.* [31] showed that miR-125b and miR-181a specifically promote the generation of neurons of dopaminergic fate from NPCs derived from human ESCs, whereas miR-181a* inhibits the development of this neurotransmitter subtype. By using a set of miRNAmimics and –inhibitors, they also demonstrated that inhibition of miRNA-124 enhances the development of dopaminergic neurons [31]. Although other studies on additional neuronal subtypes have not been published, these studies demonstrated that time-controlled modulation of specific miRNA activities can contribute to the derivation of defined neuronal cells in vitro.

Recent developments in direct somatic cell reprogramming also highlighted the potential of miRNAs as mediators for transdifferentiation along the neural lineage. The convergence of transcriptional control by miRNAs that leads into direct cellular transitions is exemplified by miR-9/9* and miR-124, which both belong to a set of brain-enriched miRNAs that are activated upon initiation of neurogenesis [205]. It has been shown that overexpression of miR-124 along with BRN2 and MYT1L is able to reprogram human fibroblasts into functional neurons in the absence of other cell types [206]. These iNs exhibited typical neuronal morphology, appropriate electrophysiological properties and were able to form functional synapses between each other [206]. In a separate study, it was reported that the expression of miR-9/9* and miR-124 in human fibroblasts induced their direct conversion into neurons, a process which was enhanced by the addition of several transcription factors, including ASCL1, MYT1L and NEUROD2 [207]. Importantly, they found that the expression of these transcription factors alone without the miRNAs was inefficient to induce a neuronal phenotype, suggesting that this miRNA-induced neuronal state is indulgent to subtype-specific transcription factors that can initiate and advance differentiation towards mature neuronal identities [205]. Co-expression of miR-9/9* and miR-124 with transcription factors enriched in the developing striatum, including CTIP2, DLX1, DLX2, and MYT1L, guided the conversion of human fibroblasts into enriched populations of GABAergic neurons analogous to striatal medium spiny neurons [182].

Synergism between miR-9/9*, miR-124 and two other neuronal subtype-specific transcription factors, *i.e.* ISL1 and LHX3, has also been shown to be able to induce a highly homogeneous population of spinal cord motor neurons in adult human fibroblasts [205]. Longitudinal analyses of the transcriptome, genome-wide DNA-methylation, and chromatin accessibilities in the latter study revealed that miR-9/9* and miR-124 trigger reconfiguration of the epigenome, including activation of a pan-neuronal program and the reconfiguration of chromatin accessibilities [205]. Neurons solely induced by miR-9/9*- and miR-124 demonstrated to be functionally excitable and uncommitted toward specific subtypes, but possess open chromatin structures at neuronal subtype-specific loci that can be activated upon further instructions [205]. The authors, therefore, also suggested that expression of bot miRNAs in somatic cells initiates gradual but active changes in the activities of multiple chromatin modifiers while simultaneously repressing anti-neuronal genes and activating neuronal genes, resulting in the binary cell fate switch. The fact that pre-existing neuronal loci within the heterochromatic regions opened-up in response to miR-9/9* and miR-124, suggests that miRNA-mediated reprogramming could indeed stem from their ability to induce remodelling of the epigenome. Taken together, these findings demonstrate that miR-9/9* and miR-124 control the neurogenic potential of somatic cells and provide a platform for the foundation of subtype-specific neuronal conversions of human cells. For a further in depth description of all other brain-enriched miRNAs, their target genes and exact functions, please refer to excellent reviews by Åkerblom and Jakobsson [208], Coolen *et al.* [209], Meza-Sosa *et al.* [210] and Stappert *et al.* [198].

To conclude, miRNA-mediated neuronal differentiation and transdifferentiation have enlarged our current toolkit for directing neuronal cell fate in vitro and have the potential to widen our understanding on the transcriptional regulations in cell fate decisions. Further insights could in the end be exploited to develop new protocols in order to obtain enriched populations of the complete repertoire of neurons found in the human brain.

Discussion and future directions

The establishment of efficient stepwise protocols to obtain functional neurons *in vitro* is highly essential for the study of human brain functions, as well as disease modelling, drugs discovery and regenerative medicine. In this review, we have highlighted the advances that have been made over the last two decades in obtaining neuronal cells from PSCs and somatic cells. Insights from basic research and developmental biology have guided the design of current strategies and numerous protocols for glutamatergic, dopaminergic, GABAergic, serotonergic, and cholinergic/motor neurons have become available. The use of

chemically defined systems and ectopic overexpression of key lineage-specific transcription factors have been first-choice to direct neuronal fates in vitro. In the chemically defined systems, PSCs have been exposed to a variety of cocktails containing patterning cues and small molecules that induce differentiation towards early NPCs and eventually specific mature neuronal subtypes. Classical experiments such as nuclear transfer or cell fusion on the other hand, have demonstrated that differentiated cells are not irreversibly committed to their fate. Very recent work has built on these conclusions and discovered that ectopic overexpression of defined transcription factors can directly generate iNs from distinct somatic cell types, as well as from PSCs. Other groups have combined both approaches and have shown successful derivation of neuronal populations and increased protocol efficiencies in both PSCs and somatic cells.

Scientists have also uncovered the existence of intrinsic mechanisms that influence the responsiveness to patterning cues and transcription factors. As we have outlined above, recent work has demonstrated a significant contribution of the epigenetic machinery to assigning neuronal fate and identity. From cancer studies, it is also becoming apparent that shifts in epigenetic signatures underlie phenotypic changes, and can induce stem cell-like properties in cancer cells due to transcriptional reprogramming [211]. Several pioneering studies have now used this knowledge in combination with the current growing availability of epigenetic editing systems and RNAi, enabling the modification of epigenetic marks at key-determinant loci that allow to direct neuronal fates in vitro. Additionally, lists of potential drugs and drug targets such as histone-modifying enzymes have been suggested to improve neuronal conversions. Such approaches on their own or in combination with others could lead to an accelerated application of the obtained neuronal populations. It is therefore, anticipated that epigenetic editing systems and RNAi will be increasingly involved in PSCdifferentiation and cellular reprogramming in the near future. These epigenomic editing tools even have the potential to become a golden standard for probing interactions among specific chromatin modifications, transcriptional programmes and cellular phenotypes. However, care should be taken in terms of advantages and disadvantages that come together with each of the aforementioned methods. For instance, off-target effects with the use of epigenetic-based strategies are common and should be reduced at all times to prevent experimental bias and undesirable outcomes, especially when one would consider the use of these cells in regenerative medicine. Such systems in combination with genome-integrating techniques can affect the genome in a way that it might lead to adverse changes in the biology of the cell, including changes in its differentiation potential. Genome-integrating constructs are randomly incorporated into the host genome and the copy number of the exogenous DNA

per cell may vary to great extent, since there are often no specific genetic elements or no apparent logic for their integration [212]. Furthermore, integration can occur into various chromatin-regulatory elements and interfere with gene-coding sequences, possibly affecting endogenous transcriptional regulation and three-dimensional chromatin structures. Finally, there is a possibility that transgenes maintain their activity or reactivate in the progeny of the initial targeted cell type. Nevertheless, studies have already shown to circumvent the issue with genomic-integrating techniques by the use of removable constructs, non-integrating systems such as adeno-associated vectors, transfections of mRNAs, transduction of reprogramming proteins and the use of small molecule compounds. Furthermore, promoters controlled by chemical compounds that allow to regulate transgene transcription, such as doxycycline (DOX) for example (See Tables 2 and 3), might provide an alternative to prevent continuously expressed transgenes and, thus, allows better temporal control of transgene activation.

Despite the achievements in directing neuronal cell fate in vitro, several other challenges also need to be addressed before their full potential in fundamental research and biomedicine can be utilized. Different cultivation protocols for each neuronal subtype currently exist and the majority of these protocols result in heterogeneous neuronal populations with remarkable differences in efficiency. Furthermore, we are currently also unable to enrich for the full repertoire of neurons found in the human brain, especially when considering specific anatomical subtypes within neurotransmitter classes. For these reasons, the refinement and the search for alternative approaches that will allow us to obtain all known neuronal subtypes found in the human brain remains an ongoing demand. Although multicellular cultures will be pivotal for brain-related studies, optimizing the used parameters and devising strategies to enrich for specific neuronal subtypes will also be essential in order to obtain pure populations of neurons, which will allow mechanistic studies and clinical applications in which confounding effects from other cell types can be kept to a minimum. In this respect, one could think of combining different protocols or adding additional factors to existing protocols that may act synergistically in mimicking the complete molecular processes that (re)assign cell fates along the neural lineage. As has been shown in many available approaches, every single driving factor fulfils a crucial role as a part of a bigger network and absence of a single component can completely redirect the terminal differentiation, as well as induce incomplete differentiation or reprogramming processes, leading to the production of immature cells that may not fully recapitulate bona fide neurons. By combining directed differentiation with transcriptional reprogramming and/or epigenetic-based approaches, one might be able to control these cellular conversions in a

robust way, which could, in turn, also lead to a greater diversity of neuronal subtypes to be specified in vitro. Another option to enhance culture purity could be achieved by finding more effective ways to isolate the desired cell types in different developmental stages. The use of specific cellular markers or expressing constructs combined with cell-sorting techniques, such as fluorescence-activated cell sorting (FACS) and magnetic-activated cell sorting (MACS) might provide suitable solutions (See Tables S1–3).

The choice of the initial tissue and cell type for generating the neuronal populations also needs considerable attention and could significantly affect the efficiency of a neuronal differentiation protocol. For instance, it is currently recognized that there are epigenetic differences between different PSC lines that can induce lineage differentiation bias [213]. This is exemplified in the study by Kim et al. [74], where they observed that the majority of neurons generated from one iPSC line were glutamatergic, whereas populations generated from another iPSC line mainly consisted of GABAergic neurons, when exposed to the same chemical culture conditions. Scientists have committed to the challenge to find markers that will allow them to predict this bias. The presence of such a marker could reveal which cell line has the highest neuronal differentiation capacity, leading to an increased neuronal conversion efficiency later on. It is noted that the origin-dependent epigenetic and transcriptional patterns of the pluripotent state can render iPSC lines with different neuronal differentiation potential. Cell lines that harbour epigenetic signatures, which were maintained and are characteristic of the somatic tissue of origin, have been shown to favour differentiation along lineages related to the donor cell, while restricting alternative cell fates [214]. Furthermore, it has been show that expression levels of other transcriptional regulators in iPSCs, such as miR-371-3, can predict neuronal differentiation propensities [215]. Aside from presenting the concept of epigenetic memory that may influence efforts in directed differentiation in iPSCs, this also emphasizes the role of epigenetic mechanisms in neuronal differentiation and addresses the complexity of the neuronal fate determination that needs the warrant for more comprehensive comparisons between different PSC lines.

Another debate in relation to starting cell choice, but based on direct reprogramming, arises from the initial assumption that developmentally related cells convey a higher conversion efficiency, as compared to cells that originate from distinct germ layers [216]. Indeed, previous studies have shown that cells derived from the same lineage, such as astrocytes and neurons, can be converted with minimal sets of transcription factors, whereas cells of nonectodermal origin require more than one factor or additional chemical stimulation [151]. Even though this is not always the case, as conversion of one

neuronal subtype into another is rather difficult and has been achieved only in immature cells [151, 217], lineage boundaries established during cellular differentiation and specification might be overcome, depending on the potency of the factors that are employed [216]. However, the questions to what extent direct reprogramming recapitulates the natural dynamics of neuronal differentiation and whether the developmental origin and, more specifically, the epigenetic memory of the starting cell type is negligible when using the proper reprogramming factors remain unanswered. The underlying principle of direct reprogramming is based on the expression of key lineage-specific transcription factors that are essential during development, but their action during direct reprogramming can be rather different, since these factors are operating in a completely different context [216]. Further studies on how factors function in various reprogramming environments may bring new insight that can lead into the establishment of more robust neuronal differentiation protocols in various somatic cell subtypes, as well as in PSCs. Additionally, detailed characterization of the cellular and molecular characteristics involved in guiding PSC-differentiation and somatic cell reprogramming along the neural lineage is expected to contribute, not only to enhance our understanding on the developmental aspects, but also to develop more efficient protocols and rational interventions [218, 219]. Continuous characterization of patterning cues, small molecules and other driving factors, as well as a comprehensive understanding of the underlying molecular pathways that they target will be necessary to achieve a higher efficiency, decrease culture heterogeneity and increase neuronal subtype availability. Single cell analysis and direct comparison of differentiated cells versus undifferentiated counterparts will be crucial in order to find signalling mechanisms, as well as to develop a comprehensive understanding of the epigenetic state and transcriptional programmes involved in neuronal fate specification. Such knowledge can in the end be exploited to manipulate single molecules or even complete molecular networks in the developmental processes and, hence, could aid the production of specific neuronal cells of interest. With the current advances of nextgeneration sequencing technology we are now also able to define genome-wide expression patterns and epigenetic modifications in each cell type. Such approaches have already demonstrated that different cellular subtypes display unique epigenetic signatures that persist as ESCs differentiate into the neuronal lineage. However, how these unique signatures are acquired in specific gene promoters or to what extent they are involved in shaping neuronal fates remains to be elucidated.

A final point of consideration with regard to the use of PSC- and somatic cell-derived neuronal populations in fundamental research and biomedicine, is that the conventional cell culture systems do not fully resemble the in vivo cellular microenvironment, where three-dimensional cell-to-cell interactions form the

foundation of the human brain. The simplicity of an *in vitro* culture system is an advantage, as well as a significant disadvantage, when cellular homogeneity becomes a reliability. Studying pure populations of neuronal subtypes that are in principle part of a more complex integrated cellular network, might lead to under- or overrepresentation of experimental findings depending on the research question. Organoids or three-dimensional culture systems in combination with bioprinting might offer a way to circumvent this issue [220-222]. Several studies have already succeeded in establishing organoids that imitate many features of human cortical development in a precise and complex manner [223-225]. Such culture systems derived from human iPSCs hold great potential for the investigation of developmental and evolutionary features of the human brain and provides a useful platform for drug screening and disease modelling [223, 226]. Additionally, the technology of decellularization and recellularization on obtained tissue matrixes to create entire organs *in vitro* is currently also under development [227]. Unfortunately, despite these advances, the challenge of how to generate organs that cherish the highly integrated cellular complexities like in the human brain persists, and will require collaborative multidisciplinary expertise to overcome. Nevertheless, neuronal *in vitro* differentiation techniques in combination with these advanced three-dimensional culture systems represent powerful tools for future brain-related studies. The potential to manipulate (epi)genetic and environmental factors in culture conditions, with the possibility to characterize cellular functions, electrophysiological properties and cellular connectivity of various neuronal subtypes in isolation, as well as organized in a multi-layered dimension, will be of great utility to enhance our current understanding on brain disorders and will undoubtedly contribute to the development of therapeutic interventions.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Dr. Sonia Guil and Edilene Siquiera Soares for all the fruitful discussions and insights on this manuscript. We thank CERCA Programme / Generalitat de Catalunya. Renzo J.M. Riemens is supported by Maastricht University (Maastricht, the Netherlands) and Julius Maximilians University (Wuerzburg, Germany). The work of the authors is supported by, among other institutions, the CELLEX Private Foundation, the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia, the Health Departments of the Government of Catalonia and the Joint Programme—Neurodegenerative Disease Research (JPND) for the EPIAD consortium. The project is supported through the following

funding organizations under the aegis of JPND; the Netherlands, The Netherlands Organisation for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 643417.

References

1. Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. Nature neuroscience. 2010;13(10):1161.
2. Akhtar A. The flaws and human harms of animal experimentation. Cambridge Quarterly of Healthcare Ethics. 2015;24(4):407-19.
3. Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. American journal of translational research. 2014;6(2):114.
4. Lewis DA. The human brain revisited: opportunities and challenges in postmortem studies of psychiatric disorders. Neuropsychopharmacology. 2002;26(2):143-54.
5. Badger J, Cordero-Llana O, Hartfield E, Wade-Martins R. Parkinson's disease in a dish–using stem cells as a molecular tool. Neuropharmacology. 2014;76:88-96.
6. Paşca SP, Panagiotakos G, Dolmetsch RE. Generating human neurons in vitro and using them to understand neuropsychiatric disease. Annual review of neuroscience. 2014;37:479-501.
7. Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nature reviews Drug discovery. 2017;16(2):115-30.
8. Yap MS, Nathan KR, Yeo Y, Lim LW, Poh CL, Richards M, et al. Neural differentiation of human pluripotent stem cells for nontherapeutic applications: toxicology, pharmacology, and in vitro disease modeling. Stem Cells International. 2015;2015.
9. Haston KM, Finkbeiner S. Clinical trials in a dish: the potential of pluripotent stem cells to develop therapies for neurodegenerative diseases. Annual review of pharmacology and toxicology. 2016;56:489-510.
10. Hou S, Lu P. Direct reprogramming of somatic cells into neural stem cells or neurons for neurological disorders. Neural Regeneration Research. 2016;11(1):28.
11. Riemens RJ, Soares ES, Esteller M, Delgado-Morales R. Stem Cell Technology for (Epi) genetic Brain Disorders. Neuroepigenomics in Aging and Disease: Springer; 2017. p. 443-75.
12. Young-Pearse TL, Morrow EM. Modeling developmental neuropsychiatric disorders with iPSC technology: challenges and opportunities. Current opinion in neurobiology. 2016;36:66-73.
13. Canfield SG, Stebbins MJ, Morales BS, Asai SW, Vatine GD, Svendsen CN, et al. An isogenic blood-brain barrier model comprising brain endothelial cells, astrocytes, and neurons derived from human induced pluripotent stem cells. J Neurochem. 2017;140(6):874-88.
14. Kirwan P, Turner-Bridger B, Peter M, Momoh A, Arambepola D, Robinson HP, et al. Development and function of human cerebral cortex neural networks from pluripotent stem cells in vitro. Development. 2015;142(18):3178-87.
15. Mertens J, Marchetto MC, Bardy C, Gage FH. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. Nature Reviews Neuroscience. 2016;17(7):424.
16. Maroof AM, Keros S, Tyson JA, Ying S-W, Ganat YM, Merkle FT, et al. Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. Cell stem cell. 2013;12(5):559-72.

17. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron. 2013;78(5):785-98.
18. Denham M, Dottori M. Neural differentiation of induced pluripotent stem cells. Neurodegeneration: Springer; 2011. p. 99-110.
19. Erceg S, Laínez S, Ronaghi M, Stojkovic P, Pérez-Aragó MA, Moreno-Manzano V, et al. Differentiation of human embryonic stem cells to regional specific neural precursors in chemically defined medium conditions. PLoS One. 2008;3(5):e2122.
20. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature. 2010;463(7284):1035-41.
21. Gifford CA, Ziller MJ, Gu HC, Trapnell C, Donaghey J, Tsankov A, et al. Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell. 2013;153(5):1149-63.
22. Lunyak VV, Rosenfeld MG. Epigenetic regulation of stem cell fate. Human molecular genetics. 2008;17(R1):R28-R36.
23. Qin H, Zhao A, Zhang C, Fu X. Epigenetic control of reprogramming and transdifferentiation by histone modifications. Stem Cell Reviews and Reports. 2016;12(6):708-20.
24. Imamura T, Uesaka M, Nakashima K. Epigenetic setting and reprogramming for neural cell fate determination and differentiation. Philos Trans R Soc Lond B Biol Sci. 2014;369(1652).
25. Feng JA, Fouse S, Fan GP. Epigenetic regulation of neural gene expression and neuronal function. Pediatr Res. 2007;61(5):58r-63r.
26. Fitzsimons CP, van Bodegraven E, Schouten M, Lardenoije R, Kompotis K, Kenis G, et al. Epigenetic regulation of adult neural stem cells: implications for Alzheimer's disease. Mol Neurodegener. 2014;9:25.
27. Kungulovski G, Jeltsch A. Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends in Genetics. 2016;32(2):101-13.
28. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA Methylation in the Mammalian Genome. Cell. 2016;167(1):233-47 e17.
29. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nature methods. 2016;13(2):127.
30. Low WC, Yau WWY, Stanton LW, Marcy G, Goh E, Chew SY. Directing neuronal differentiation of primary neural progenitor cells by gene knockdown approach. DNA and cell biology. 2012;31(7):1148-60.
31. Stappert L, Borghese L, Roese-Koerner B, Weinhold S, Koch P, Terstegge S, et al. MicroRNA-based promotion of human neuronal differentiation and subtype specification. PLoS One. 2013;8(3):e59011.
32. Jurkowski TP, Ravichandran M, Stepper P. Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity. Clinical Epigenetics. 2015;7.
33. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009;27(3):275-80.
34. Erceg S, Ronaghi M, Stojkovic M. Human embryonic stem cell differentiation toward regional specific neural precursors. Stem Cells. 2009;27(1):78-87.
35. Erceg S, Ronaghi M, Zipancic I, Lainez S, Rosello MG, Xiong C, et al. Efficient differentiation of human embryonic stem cells into functional cerebellar-like cells. Stem Cells Dev. 2010;19(11):1745-56.
36. Karumbayaram S, Novitch BG, Patterson M, Umbach JA, Richter L, Lindgren A, et al. Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. Stem Cells. 2009;27(4):806-11.
37. Shi G. Differentiation of human pluripotent stem cells into functional cells. OA Stem Cells. 2013:4.
38. Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol. 2003;21(2):183-6.

39. Denham M, Dottori M. Neural differentiation of induced pluripotent stem cells. Methods Mol Biol. 2011;793:99-110.
40. Karanfil I, Bagci-Onder T. Derivation of Neural Stem Cells from Mouse Induced Pluripotent Stem Cells. Methods Mol Biol. 2016;1357:329-38.
41. Itskovitz-Eldor J, Schuldiner M, Karsenti D, Eden A, Yanuka O, Amit M, et al. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Molecular medicine. 2000;6(2):88.
42. Abranches E, Silva M, Pradier L, Schulz H, Hummel O, Henrique D, et al. Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. PLoS One. 2009;4(7):e6286.
43. Günther K, Appelt-Menzel A, Kwok CK, Walles H, Metzger M, Edenhofer F. Rapid Monolayer Neural Induction of induced Pluripotent Stem Cells Yields Stably Proliferating Neural Stem Cells. Journal of Stem Cell Research & Therapy. 2016;2016.
44. Yan Y, Shin S, Jha BS, Liu Q, Sheng J, Li F, et al. Efficient and rapid derivation of primitive neural stem cells and generation of brain subtype neurons from human pluripotent stem cells. Stem cells translational medicine. 2013:sctm. 2013-0080.
45. Kawasaki H, Mizuseki K, Nishikawa S, Kaneko S, Kuwana Y, Nakanishi S, et al. Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. Neuron. 2000;28(1):31-40.
46. Lim MS, Shin MS, Lee SY, Minn YK, Hoh JK, Cho YH, et al. Noggin Over-Expressing Mouse Embryonic Fibroblasts and MS5 Stromal Cells Enhance Directed Differentiation of Dopaminergic Neurons from Human Embryonic Stem Cells. PLoS One. 2015;10(9):e0138460.
47. Fujimori K, Matsumoto T, Kisa F, Hattori N, Okano H, Akamatsu W. Escape from Pluripotency via Inhibition of TGF-beta/BMP and Activation of Wnt Signaling Accelerates Differentiation and Aging in hPSC Progeny Cells. Stem cell reports. 2017;9(5):1675-91.
48. Zhang Y, Li W, Laurent T, Ding S. Small molecules, big roles–the chemical manipulation of stem cell fate and somatic cell reprogramming. Journal of cell science. 2012;125(23):5609-20.
49. Davis-Dusenbery BN, Williams LA, Klim JR, Eggan K. How to make spinal motor neurons. Development. 2014;141(3):491-501.
50. Tao Y, Zhang SC. Neural Subtype Specification from Human Pluripotent Stem Cells. Cell Stem Cell. 2016;19(5):573-86.
51. Pasca SP, Panagiotakos G, Dolmetsch RE. Generating human neurons in vitro and using them to understand neuropsychiatric disease. Annu Rev Neurosci. 2014;37:479-501.
52. Patani R. Generating Diverse Spinal Motor Neuron Subtypes from Human Pluripotent Stem Cells. Stem Cells Int. 2016;2016:1036974.
53. Goulburn AL, Stanley EG, Elefanty AG, Anderson SA. Generating GABAergic cerebral cortical interneurons from mouse and human embryonic stem cells. Stem Cell Res. 2012;8(3):416-26.
54. Li W, Sun W, Zhang Y, Wei W, Ambasudhan R, Xia P, et al. Rapid induction and long-term self-renewal of primitive neural precursors from human embryonic stem cells by small molecule inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(20):8299-304.
55. Dhara SK, Stice SL. Neural differentiation of human embryonic stem cells. Journal of cellular biochemistry. 2008;105(3):633-40.
56. Wilson SI, Graziano E, Harland R, Jessell TM, Edlund T. An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. Curr Biol. 2000;10(8):421-9.
57. Eiraku M, Watanabe K, Matsuo-Takasaki M, Kawada M, Yonemura S, Matsumura M, et al. Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. Cell Stem Cell. 2008;3(5):519-32.

58. Lee H, Shamy GA, Elkabetz Y, Schofield CM, Harrsion NL, Panagiotakos G, et al. Directed differentiation and transplantation of human embryonic stem cell-derived motoneurons. Stem Cells. 2007;25(8):1931-9.
59. Lu J, Zhong X, Liu H, Hao L, Huang CT, Sherafat MA, et al. Generation of serotonin neurons from human pluripotent stem cells. Nat Biotechnol. 2016;34(1):89-94.
60. Yan Y, Yang D, Zarnowska ED, Du Z, Werbel B, Valliere C, et al. Directed differentiation of dopaminergic neuronal subtypes from human embryonic stem cells. Stem Cells. 2005;23(6):781-90.
61. Imaizumi K, Sone T, Ibata K, Fujimori K, Yuzaki M, Akamatsu W, et al. Controlling the Regional Identity of hPSC-Derived Neurons to Uncover Neuronal Subtype Specificity of Neurological Disease Phenotypes. Stem cell reports. 2015;5(6):1010-22.
62. Kirkeby A, Grealish S, Wolf DA, Nelander J, Wood J, Lundblad M, et al. Generation of regionally specified neural progenitors and functional neurons from human embryonic stem cells under defined conditions. Cell reports. 2012;1(6):703-14.
63. Liu H, Zhang SC. Specification of neuronal and glial subtypes from human pluripotent stem cells. Cell Mol Life Sci. 2011;68(24):3995-4008.
64. Gabilondo H, Stratmann J, Rubio-Ferrera I, Millan-Crespo I, Contero-Garcia P, Bahrampour S, et al. Neuronal Cell Fate Specification by the Convergence of Different Spatiotemporal Cues on a Common Terminal Selector Cascade. PLoS biology. 2016;14(5):e1002450.
65. Zhang SC, Wernig M, Duncan ID, Brustle O, Thomson JA. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. Nat Biotechnol. 2001;19(12):1129-33.
66. Zhang D, Jiang W. From One-Cell to Tissue: Reprogramming, Cell Differentiation and Tissue Engineering. BioScience. 2015:biv016.
67. Costa MR, Muller U. Specification of excitatory neurons in the developing cerebral cortex: progenitor diversity and environmental influences. Front Cell Neurosci. 2014;8:449.
68. Song H, Lee B, Pyun D, Guimera J, Son Y, Yoon J, et al. Ascl1 and Helt act combinatorially to specify thalamic neuronal identity by repressing Dlxs activation. Developmental biology. 2015;398(2):280-91.
69. Thomas Cheng H. Spinal cord mechanisms of chronic pain and clinical implications. Current pain and headache reports. 2010;14(3):213-20.
70. Li XJ, Zhang X, Johnson MA, Wang ZB, Lavaute T, Zhang SC. Coordination of sonic hedgehog and Wnt signaling determines ventral and dorsal telencephalic neuron types from human embryonic stem cells. Development. 2009;136(23):4055-63.
71. Gaspard N, Bouschet T, Hourez R, Dimidschstein J, Naeije G, van den Ameele J, et al. An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature. 2008;455(7211):351-7.
72. Bibel M, Richter J, Schrenk K, Tucker KL, Staiger V, Korte M, et al. Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat Neurosci. 2004;7(9):1003-9.
73. Espuny-Camacho I, Michelsen KA, Gall D, Linaro D, Hasche A, Bonnefont J, et al. Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron. 2013;77(3):440-56.
74. Kim JE, O'Sullivan ML, Sanchez CA, Hwang M, Israel MA, Brennand K, et al. Investigating synapse formation and function using human pluripotent stem cell-derived neurons. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(7):3005-10.
75. Shi Y, Kirwan P, Livesey FJ. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat Protoc. 2012;7(10):1836-46.
76. Shi Y, Kirwan P, Smith J, Robinson HP, Livesey FJ. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat Neurosci. 2012;15(3):477-86, S1.

77. Liu Y, Liu H, Sauvey C, Yao L, Zarnowska ED, Zhang SC. Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. Nat Protoc. 2013;8(9):1670-9.
78. Fishell G, Rudy B. Mechanisms of inhibition within the telencephalon:"where the wild things are". Annual review of neuroscience. 2011;34:535-67.
79. Nobrega-Pereira S, Kessaris N, Du T, Kimura S, Anderson SA, Marin O. Postmitotic Nkx2-1 controls the migration of telencephalic interneurons by direct repression of guidance receptors. Neuron. 2008;59(5):733-45.
80. Kriegstein AR, Noctor SC. Patterns of neuronal migration in the embryonic cortex. Trends in neurosciences. 2004;27(7):392-9.
81. Marin O, Rubenstein JL. Cell migration in the forebrain. Annu Rev Neurosci. 2003;26:441-83.
82. Sussel L, Marin O, Kimura S, Rubenstein JL. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. Development. 1999;126(15):3359-70.
83. Campbell K. Dorsal-ventral patterning in the mammalian telencephalon. Curr Opin Neurobiol. 2003;13(1):50-6.
84. Watanabe K, Kamiya D, Nishiyama A, Katayama T, Nozaki S, Kawasaki H, et al. Directed differentiation of telencephalic precursors from embryonic stem cells. Nature neuroscience. 2005;8(3):288-96.
85. Maroof AM, Brown K, Shi S-H, Studer L, Anderson SA. Prospective isolation of cortical interneuron precursors from mouse embryonic stem cells. The Journal of Neuroscience. 2010;30(13):4667-75.
86. Nicholas CR, Chen J, Tang Y, Southwell DG, Chalmers N, Vogt D, et al. Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development. Cell stem cell. 2013;12(5):573-86.
87. Carri AD, Onorati M, Castiglioni V, Faedo A, Camnasio S, Toselli M, et al. Human pluripotent stem cell differentiation into authentic striatal projection neurons. Stem Cell Reviews and Reports. 2013;9(4):461-74.
88. Lin L, Yuan J, Sander B, Golas MM. In Vitro Differentiation of Human Neural Progenitor Cells Into Striatal GABAergic Neurons. Stem Cells Transl Med. 2015;4(7):775-88.
89. Tang Y, Liu L, Li J, Yu L, Wang L, Shi J, et al. Induction and differentiation of human induced pluripotent stem cells into functional cardiomyocytes on a compartmented monolayer of gelatin nanofibers. Nanoscale. 2016;8(30):14530-40.
90. Chinta SJ, Andersen JK. Dopaminergic neurons. The international journal of biochemistry & cell biology. 2005;37(5):942-6.
91. Ono Y, Nakatani T, Sakamoto Y, Mizuhara E, Minaki Y, Kumai M, et al. Differences in neurogenic potential in floor plate cells along an anteroposterior location: midbrain dopaminergic neurons originate from mesencephalic floor plate cells. Development. 2007;134(17):3213-25.
92. Placzek M, Briscoe J. The floor plate: multiple cells, multiple signals. Nature reviews Neuroscience. 2005;6(3):230-40.
93. Chung S, Leung A, Han BS, Chang MY, Moon JI, Kim CH, et al. Wnt1-lmx1a forms a novel autoregulatory loop and controls midbrain dopaminergic differentiation synergistically with the SHH-FoxA2 pathway. Cell Stem Cell. 2009;5(6):646-58.
94. Xi J, Liu Y, Liu H, Chen H, Emborg ME, Zhang SC. Specification of midbrain dopamine neurons from primate pluripotent stem cells. Stem Cells. 2012;30(8):1655-63.
95. Prakash N, Brodski C, Naserke T, Puelles E, Gogoi R, Hall A, et al. A Wnt1-regulated genetic network controls the identity and fate of midbrain-dopaminergic progenitors in vivo. Development. 2006;133(1):89-98.

96. Andersson E, Tryggvason U, Deng Q, Friling S, Alekseenko Z, Robert B, et al. Identification of intrinsic determinants of midbrain dopamine neurons. Cell. 2006;124(2):393-405.
97. Barberi T, Klivenyi P, Calingasan NY, Lee H, Kawamata H, Loonam K, et al. Neural subtype specification of fertilization and nuclear transfer embryonic stem cells and application in parkinsonian mice. Nat Biotechnol. 2003;21(10):1200-7.
98. Lee SH, Lumelsky N, Studer L, Auerbach JM, McKay RD. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. Nat Biotechnol. 2000;18(6):675-9.
99. Ye W, Shimamura K, Rubenstein JL, Hynes MA, Rosenthal A. FGF and Shh signals control dopaminergic and serotonergic cell fate in the anterior neural plate. Cell. 1998;93(5):755-66.
100. Cai J, Yang M, Poremsky E, Kidd S, Schneider JS, Iacovitti L. Dopaminergic neurons derived from human induced pluripotent stem cells survive and integrate into 6-OHDA-lesioned rats. Stem Cells Dev. 2010;19(7):1017-23.
101. Swistowski A, Peng J, Liu Q, Mali P, Rao MS, Cheng L, et al. Efficient generation of functional dopaminergic neurons from human induced pluripotent stem cells under defined conditions. Stem cells. 2010;28(10):1893-904.
102. Arenas E, Denham M, Villaescusa JC. How to make a midbrain dopaminergic neuron. Development. 2015;142(11):1918-36.
103. Denham M, Bye C, Leung J, Conley BJ, Thompson LH, Dottori M. Glycogen Synthase Kinase 3β and Activin/Nodal Inhibition in Human Embryonic Stem Cells Induces a Pre-Neuroepithelial State That Is Required for Specification to a Floor Plate Cell Lineage. Stem Cells. 2012;30(11):2400-11.
104. Doi D, Samata B, Katsukawa M, Kikuchi T, Morizane A, Ono Y, et al. Isolation of human induced pluripotent stem cell-derived dopaminergic progenitors by cell sorting for successful transplantation. Stem cell reports. 2014;2(3):337-50.
105. Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, et al. Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature. 2011;480(7378):547-51.
106. Fernandez-Santiago R, Carballo-Carbajal I, Castellano G, Torrent R, Richaud Y, Sanchez-Danes A, et al. Aberrant epigenome in iPSC-derived dopaminergic neurons from Parkinson's disease patients. EMBO molecular medicine. 2015.
107. Woodard CM, Campos BA, Kuo SH, Nirenberg MJ, Nestor MW, Zimmer M, et al. iPSC-derived dopamine neurons reveal differences between monozygotic twins discordant for Parkinson's disease. Cell Rep. 2014;9(4):1173-82.
108. Deneris E, Gaspar P. Serotonin neuron development: shaping molecular and structural identities. Wiley Interdiscip Rev Dev Biol. 2018;7(1).
109. Alenina N, Bashammakh S, Bader M. Specification and differentiation of serotonergic neurons. Stem Cell Rev. 2006;2(1):5-10.
110. Vadodaria KC, Marchetto MC, Mertens J, Gage FH. Generating human serotonergic neurons in vitro: Methodological advances. BioEssays : news and reviews in molecular, cellular and developmental biology. 2016;38(11):1123-9.
111. Dias JM, Alekseenko Z, Applequist JM, Ericson J. Tgfbeta signaling regulates temporal neurogenesis and potency of neural stem cells in the CNS. Neuron. 2014;84(5):927-39.
112. Rhinn M, Brand M. The midbrain--hindbrain boundary organizer. Curr Opin Neurobiol. 2001;11(1):34-42.
113. Shimada T, Takai Y, Shinohara K, Yamasaki A, Tominaga-Yoshino K, Ogura A, et al. A simplified method to generate serotonergic neurons from mouse embryonic stem and induced pluripotent stem cells. J Neurochem. 2012;122(1):81-93.
114. Kumar M, Kaushalya SK, Gressens P, Maiti S, Mani S. Optimized derivation and functional characterization of 5-HT neurons from human embryonic stem cells. Stem Cells Dev. 2009;18(4):615-27.

115. Stock A, Kuzis K, Woodward WR, Nishi R, Eckenstein FP. Localization of acidic fibroblast growth factor in specific subcortical neuronal populations. J Neurosci. 1992;12(12):4688-700.
116. Migliarini S, Pacini G, Pelosi B, Lunardi G, Pasqualetti M. Lack of brain serotonin affects postnatal development and serotonergic neuronal circuitry formation. Mol Psychiatry. 2013;18(10):1106-18.
117. Vadodaria KC, Mertens J, Paquola A, Bardy C, Li X, Jappelli R, et al. Generation of functional human serotonergic neurons from fibroblasts. Mol Psychiatry. 2015.
118. Licinio J, Wong M. Serotonergic neurons derived from induced pluripotent stem cells (iPSCs): a new pathway for research on the biology and pharmacology of major depression. Molecular psychiatry. 2016;21(1):1-2.
119. Sierksma AS, van den Hove DL, Steinbusch HW, Prickaerts J. Major depression, cognitive dysfunction and Alzheimer's disease: is there a link? European journal of pharmacology. 2010;626(1):72-82.
120. Li X-J, Du Z-W, Zarnowska ED, Pankratz M, Hansen LO, Pearce RA, et al. Specification of motoneurons from human embryonic stem cells. Nature biotechnology. 2005;23(2):215-21.
121. Singh Roy N, Nakano T, Xuing L, Kang J, Nedergaard M, Goldman SA. Enhancer-specified GFP-based FACS purification of human spinal motor neurons from embryonic stem cells. Exp Neurol. 2005;196(2):224-34.
122. Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W, et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science. 2008;321(5893):1218-21.
123. Qu Q, Li D, Louis KR, Li X, Yang H, Sun Q, et al. High-efficiency motor neuron differentiation from human pluripotent stem cells and the function of Islet-1. Nature communications. 2014;5:3449.
124. Shimojo D, Onodera K, Doi-Torii Y, Ishihara Y, Hattori C, Miwa Y, et al. Rapid, efficient, and simple motor neuron differentiation from human pluripotent stem cells. Mol Brain. 2015;8(1):79.
125. Wichterle H, Lieberam I, Porter JA, Jessell TM. Directed differentiation of embryonic stem cells into motor neurons. Cell. 2002;110(3):385-97.
126. Amoroso MW, Croft GF, Williams DJ, O'Keeffe S, Carrasco MA, Davis AR, et al. Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. J Neurosci. 2013;33(2):574-86.
127. Vierbuchen T, Wernig M. Direct lineage conversions: unnatural but useful? Nat Biotechnol. 2011;29(10):892-907.
128. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell. 2006;126(4):663-76.
129. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318(5858):1917-20.
130. Schlaeger TM, Daheron L, Brickler TR, Entwisle S, Chan K, Cianci A, et al. A comparison of non-integrating reprogramming methods. Nature biotechnology. 2015;33(1):58-63.
131. Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. Nat Rev Mol Cell Biol. 2016;17(3):183-93.
132. Weintraub H, Tapscott SJ, Davis RL, Thayer MJ, Adam MA, Lassar AB, et al. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proceedings of the National Academy of Sciences. 1989;86(14):5434-8.
133. Mertens J, Paquola ACM, Ku M, Hatch E, Bohnke L, Ladjevardi S, et al. Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell. 2015;17(6):705-18.

134. Busskamp V, Lewis NE, Guye P, Ng AH, Shipman SL, Byrne SM, et al. Rapid neurogenesis through transcriptional activation in human stem cells. Mol Syst Biol. 2014;10:760.
135. Hou S, Lu P. Direct reprogramming of somatic cells into neural stem cells or neurons for neurological disorders. Neural Regen Res. 2016;11(1):28-31.
136. Raciti M, Granzotto M, Do Duc M, Fimiani C, Cellot G, Cherubini E, et al. Reprogramming fibroblasts to neural-precursor-like cells by structured overexpression of pallial patterning genes. Molecular and Cellular Neuroscience. 2013;57:42-53.
137. Matsui T, Takano M, Yoshida K, Ono S, Fujisaki C, Matsuzaki Y, et al. Neural stem cells directly differentiated from partially reprogrammed fibroblasts rapidly acquire gliogenic competency. Stem Cells. 2012;30(6):1109-19.
138. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature. 2010;463(7284):1035-41.
139. Pang ZP, Yang N, Vierbuchen T, Ostermeier A, Fuentes DR, Yang TQ, et al. Induction of human neuronal cells by defined transcription factors. Nature. 2011;476(7359):220-3.
140. Chanda S, Ang CE, Davila J, Pak C, Mall M, Lee QY, et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem cell reports. 2014;3(2):282-96.
141. Mall M, Kareta MS, Chanda S, Ahlenius H, Perotti N, Zhou B, et al. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. Nature. 2017;544(7649):245-9.
142. Heinrich C, Blum R, Gascon S, Masserdotti G, Tripathi P, Sanchez R, et al. Directing astroglia from the cerebral cortex into subtype specific functional neurons. PLoS biology. 2010;8(5):e1000373.
143. Thoma EC, Wischmeyer E, Offen N, Maurus K, Siren AL, Schartl M, et al. Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. PLoS One. 2012;7(6):e38651.
144. Colasante G, Lignani G, Rubio A, Medrihan L, Yekhlef L, Sessa A, et al. Rapid conversion of fibroblasts into functional forebrain GABAergic interneurons by direct genetic reprogramming. Cell stem cell. 2015;17(6):719-34.
145. Mong J, Panman L, Alekseenko Z, Kee N, Stanton LW, Ericson J, et al. Transcription Factor-Induced Lineage Programming of Noradrenaline and Motor Neurons from Embryonic Stem Cells. Stem Cells. 2014;32(3):609-22.
146. Pfisterer U, Kirkeby A, Torper O, Wood J, Nelander J, Dufour A, et al. Direct conversion of human fibroblasts to dopaminergic neurons. Proceedings of the National Academy of Sciences. 2011;108(25):10343-8.
147. Son EY, Ichida JK, Wainger BJ, Toma JS, Rafuse VF, Woolf CJ, et al. Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell stem cell. 2011;9(3):205-18.
148. Xu Z, Jiang H, Zhong P, Yan Z, Chen S, Feng J. Direct conversion of human fibroblasts to induced serotonergic neurons. Mol Psychiatry. 2015.
149. Hester ME, Murtha MJ, Song S, Rao M, Miranda CJ, Meyer K, et al. Rapid and efficient generation of functional motor neurons from human pluripotent stem cells using gene delivered transcription factor codes. Molecular therapy. 2011;19(10):1905-12.
150. Kim J, Su SC, Wang H, Cheng AW, Cassady JP, Lodato MA, et al. Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. Cell stem cell. 2011;9(5):413-9.
151. Gascon S, Masserdotti G, Russo GL, Gotz M. Direct Neuronal Reprogramming: Achievements, Hurdles, and New Roads to Success. Cell Stem Cell. 2017;21(1):18-34.
152. Xie X, Fu Y, Liu J. Chemical reprogramming and transdifferentiation. Curr Opin Genet Dev. 2017;46:104-13.
153. Caiazzo M, Dell'Anno MT, Dvoretskova E, Lazarevic D, Taverna S, Leo D, et al. Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature. 2011;476(7359):224-7.

154. Addis RC, Hsu F-C, Wright RL, Dichter MA, Coulter DA, Gearhart JD. Efficient conversion of astrocytes to functional midbrain dopaminergic neurons using a single polycistronic vector. PloS one. 2011;6(12):e28719.
155. Theka I, Caiazzo M, Dvoretskova E, Leo D, Ungaro F, Curreli S, et al. Rapid generation of functional dopaminergic neurons from human induced pluripotent stem cells through a single-step procedure using cell lineage transcription factors. Stem cells translational medicine. 2013;2(6):473-9.
156. Wasko NJ. Direct Reprogramming of Mouse Embryonic Fibroblasts Into GABAergic Neurons. 2013.
157. Mazzoni EO, Mahony S, Closser M, Morrison CA, Nedelec S, Williams DJ, et al. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. Nature neuroscience. 2013;16(9):1219-27.
158. Barrero MJ, Boué S, Belmonte JCI. Epigenetic mechanisms that regulate cell identity. Cell stem cell. 2010;7(5):565-70.
159. Delgado-Morales R, Esteller M. Opening up the DNA methylome of dementia. Molecular psychiatry. 2017;22(4):485-96.
160. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128(4):635-8.
161. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131:21-64.
162. Avgustinova A, Benitah SA. Epigenetic control of adult stem cell function. Nat Rev Mol Cell Bio. 2016;17(10):643.
163. Coskun V, Tsoa R, Sun YE. Epigenetic regulation of stem cells differentiating along the neural lineage. Curr Opin Neurobiol. 2012;22(5):762-7.
164. Lilja T, Heldring N, Hermanson O. Like a rolling histone: epigenetic regulation of neural stem cells and brain development by factors controlling histone acetylation and methylation. Biochimica et Biophysica Acta (BBA)-General Subjects. 2013;1830(2):2354-60.
165. Olynik BM, Rastegar M. The genetic and epigenetic journey of embryonic stem cells into mature neural cells. Front Genet. 2012;3:81.
166. Liu J, Githinji J, McLaughlin B, Wilczek K, Nolta J. Role of miRNAs in neuronal differentiation from human embryonic stem cell-derived neural stem cells. Stem Cell Rev. 2012;8(4):1129-37.
167. Lukovic D, Moreno-Manzano V, Klabusay M, Stojkovic M, Bhattacharya SS, Erceg S. Non-coding RNAs in pluripotency and neural differentiation of human pluripotent stem cells. Frontiers in genetics. 2014;5.
168. Esteller M. Non-coding RNAs in human disease. Nature reviews genetics. 2011;12(12):861-74.
169. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. Nature structural & molecular biology. 2012;19(11):1068-75.
170. Smith ZD, Sindhu C, Meissner A. Molecular features of cellular reprogramming and development. Nat Rev Mol Cell Bio. 2016;17(3):139-54.
171. Hirabayashi Y, Gotoh Y. Epigenetic control of neural precursor cell fate during development. Nature reviews Neuroscience. 2010;11(6):377-88.
172. Huang C, Wu JC. Epigenetic modulations of induced pluripotent stem cells: novel therapies and disease models. Drug Discovery Today: Disease Models. 2012;9(4):e153-e60.
173. van den Hurk M, Kenis G, Bardy C, van den Hove DL, Gage FH, Steinbusch HW, et al. Transcriptional and epigenetic mechanisms of cellular reprogramming to induced pluripotency. Epigenomics. 2016;8(8):1131-49.
174. Watanabe A, Yamada Y, Yamanaka S. Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013;368(1609):20120292.

175. Firas J, Liu X, Lim SM, Polo JM. Transcription factor-mediated reprogramming: epigenetics and therapeutic potential. Immunology and cell biology. 2015;93(3):284-9.
176. Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature Reviews Genetics. 2014;15(2):93-106.
177. Cheng L, Gao L, Guan W, Mao J, Hu W, Qiu B, et al. Direct conversion of astrocytes into neuronal cells by drug cocktail. Cell research. 2015;25(11):1269-72.
178. Cheng L, Hu W, Qiu B, Zhao J, Yu Y, Guan W, et al. Erratum: Generation of neural progenitor cells by chemical cocktails and hypoxia. Cell research. 2015;25(5):645-6.
179. Black JB, Adler AF, Wang HG, D'Ippolito AM, Hutchinson HA, Reddy TE, et al. Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell. 2016;19(3):406-14.
180. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, E PRI, et al. Highly efficient Cas9-mediated transcriptional programming. Nat Methods. 2015;12(4):326-8.
181. Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development. 2014;141(1):219-23.
182. Victor MB, Richner M, Hermanstyne TO, Ransdell JL, Sobieski C, Deng PY, et al. Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. Neuron. 2014;84(2):311-23.
183. Laufer BI, Singh SM. Strategies for precision modulation of gene expression by epigenome editing: an overview. Epigenetics & chromatin. 2015;8(1):1.
184. Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N, Leong KW. A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem cell reports. 2014;3(6):940-7.
185. Wei S, Zou Q, Lai S, Zhang Q, Li L, Yan Q, et al. Conversion of embryonic stem cells into extraembryonic lineages by CRISPR-mediated activators. Sci Rep. 2016;6:19648.
186. Gao X, Yang J, Tsang JC, Ooi J, Wu D, Liu P. Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. Stem cell reports. 2013;1(2):183-97.
187. Balboa D, Weltner J, Eurola S, Trokovic R, Wartiovaara K, Otonkoski T. Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. Stem cell reports. 2015;5(3):448-59.
188. Cano-Rodriguez D, Rots MG. Epigenetic editing: on the verge of reprogramming gene expression at will. Current genetic medicine reports. 2016;4(4):170-9.
189. Fu L, Zhu X, Yi F, Liu G-H, Belmonte JCI. Regenerative medicine: transdifferentiation in vivo. Cell Research. 2014;24(2):141-2.
190. Gascon S, Murenu E, Masserdotti G, Ortega F, Russo GL, Petrik D, et al. Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. Cell Stem Cell. 2016;18(3):396-409.
191. Guo Z, Zhang L, Wu Z, Chen Y, Wang F, Chen G. In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell Stem Cell. 2014;14(2):188-202.
192. Torper O, Pfisterer U, Wolf DA, Pereira M, Lau S, Jakobsson J, et al. Generation of induced neurons via direct conversion in vivo. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(17):7038-43.
193. Niu W, Zang T, Zou Y, Fang S, Smith DK, Bachoo R, et al. In vivo reprogramming of astrocytes to neuroblasts in the adult brain. Nature cell biology. 2013;15(10).
194. Ohori Y, Yamamoto S, Nagao M, Sugimori M, Yamamoto N, Nakamura K, et al. Growth factor treatment and genetic manipulation stimulate neurogenesis and oligodendrogenesis by endogenous neural progenitors in the injured adult spinal cord. J Neurosci. 2006;26(46):11948-60.

195. Li H, Chen G. In Vivo Reprogramming for CNS Repair: Regenerating Neurons from Endogenous Glial Cells. Neuron. 2016;91(4):728-38.
196. Srivastava D, DeWitt N. In Vivo Cellular Reprogramming: The Next Generation. Cell. 2016;166(6):1386-96.
197. Ong SG, Lee WH, Kodo K, Wu JC. MicroRNA-mediated regulation of differentiation and trans-differentiation in stem cells. Adv Drug Deliv Rev. 2015;88:3-15.
198. Stappert L, Roese-Koerner B, Brustle O. The role of microRNAs in human neural stem cells, neuronal differentiation and subtype specification. Cell Tissue Res. 2015;359(1):47-64.
199. Lopez-Ramirez MA, Nicoli S. Role of miRNAs and epigenetics in neural stem cell fate determination. Epigenetics. 2014;9(1):90-100.
200. He M, Liu Y, Wang X, Zhang MQ, Hannon GJ, Huang ZJ. Cell-type-based analysis of microRNA profiles in the mouse brain. Neuron. 2012;73(1):35-48.
201. Davis S, Lollo B, Freier S, Esau C. Improved targeting of miRNA with antisense oligonucleotides. Nucleic acids research. 2006;34(8):2294-304.
202. Yang N. An overview of viral and nonviral delivery systems for microRNA. International journal of pharmaceutical investigation. 2015;5(4):179.
203. Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, Murchison E, et al. A MicroRNA feedback circuit in midbrain dopamine neurons. Science. 2007;317(5842):1220-4.
204. Yang D, Li T, Wang Y, Tang Y, Cui H, Tang Y, et al. miR-132 regulates the differentiation of dopamine neurons by directly targeting Nurr1 expression. J Cell Sci. 2012;125(Pt 7):1673-82.
205. Abernathy DG, Kim WK, McCoy MJ, Lake AM, Ouwenga R, Lee SW, et al. MicroRNAs Induce a Permissive Chromatin Environment that Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts. Cell Stem Cell. 2017;21(3):332-48 e9.
206. Ambasudhan R, Talantova M, Coleman R, Yuan X, Zhu S, Lipton SA, et al. Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. Cell stem cell. 2011;9(2):113-8.
207. Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li YL, et al. MicroRNA-mediated conversion of human fibroblasts to neurons. Nature. 2011;476(7359):228-U123.
208. Åkerblom M, Jakobsson J. MicroRNAs as neuronal fate determinants. The Neuroscientist. 2014;20(3):235-42.
209. Coolen M, Katz S, Bally-Cuif L. miR-9: a versatile regulator of neurogenesis. Frontiers in cellular neuroscience. 2013;7:220.
210. Meza-Sosa KF, Pedraza-Alva G, Pérez-Martínez L. microRNAs: key triggers of neuronal cell fate. Frontiers in cellular neuroscience. 2014;8:175.
211. Mateo F, Arenas E, Aguilar H, Serra-Musach J, De Garibay GR, Boni J, et al. Stem cell-like transcriptional reprogramming mediates metastatic resistance to mTOR inhibition. Oncogene. 2017;36(19):2737-49.
212. Medvedev S, Shevchenko A, Zakian S. Induced pluripotent stem cells: problems and advantages when applying them in regenerative medicine. Acta Naturae (англоязычная версия). 2010;2(2 (5)).
213. Nishizawa M, Chonabayashi K, Nomura M, Tanaka A, Nakamura M, Inagaki A, et al. Epigenetic variation between human induced pluripotent stem cell lines is an indicator of differentiation capacity. Cell stem cell. 2016;19(3):341-54.
214. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010;467(7313):285-90.
215. Kim H, Lee G, Ganat Y, Papapetrou EP, Lipchina I, Socci ND, et al. miR-371-3 expression predicts neural differentiation propensity in human pluripotent stem cells. Cell Stem Cell. 2011;8(6):695-706.
216. Masserdotti G, Gascón S, Götz M. Direct neuronal reprogramming: learning from and for development. Development. 2016;143(14):2494-510.

217. Rouaux C, Arlotta P. Direct lineage reprogramming of post-mitotic callosal neurons into corticofugal neurons in vivo. Nature cell biology. 2013;15(2):214-21.
218. Kee N, Volakakis N, Kirkeby A, Dahl L, Storvall H, Nolbrant S, et al. Single-Cell Analysis Reveals a Close Relationship between Differentiating Dopamine and Subthalamic Nucleus Neuronal Lineages. Cell Stem Cell. 2017;20(1):29-40.
219. Kirkeby A, Nolbrant S, Tiklova K, Heuer A, Kee N, Cardoso T, et al. Predictive Markers Guide Differentiation to Improve Graft Outcome in Clinical Translation of hESC-Based Therapy for Parkinson's Disease. Cell Stem Cell. 2017;20(1):135-48.
220. Fatehullah A, Tan SH, Barker N. Organoids as an in vitro model of human development and disease. Nature cell biology. 2016;18(3):246-54.
221. Lancaster MA, Knoblich JA. Organogenesis in a dish: modeling development and disease using organoid technologies. Science. 2014;345(6194).
222. Murphy SV, Atala A. 3D bioprinting of tissues and organs. Nature biotechnology. 2014;32(8):773-85.
223. Dezonne RS, Sartore RC, Nascimento JM, Saia-Cereda VM, Romão LF, Alves-Leon SV, et al. Derivation of functional human astrocytes from cerebral organoids. Scientific reports. 2017;7:45091.
224. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. Nature. 2013;501(7467):373-9.
225. Li Y, Muffat J, Omer A, Bosch I, Lancaster MA, Sur M, et al. Induction of expansion and folding in human cerebral organoids. Cell stem cell. 2017;20(3):385-96. e3.
226. Nascimento JM, Martins-de-Souza D. The proteome of schizophrenia. npj Schizophrenia. 2015;1(1):1-11.
227. Fu R-H, Wang Y-C, Liu S-P, Shih T-R, Lin H-L, Chen Y-M, et al. Decellularization and recellularization technologies in tissue engineering. Cell transplantation. 2014;23(4-5):621-30.

Supplementary material

Supplementary Table 1. Overview of the assays that have been used for the phenotypic characterization of the chemically defined differentiation systems for neuronal induction and differentiation.

					Phenotypic characterization							
References	**Phenotypes**	**Species**	**Starting cell types**	**Cell sorting**	**(Quantitative) Polymerase chain reaction**	**Genome-wide expression analysis**	**Western-Blot**	**Immunohistochemistry**	**Electrophysiological assessment**	**Biochemical assessment**	**Transplantations/Co-culture assay**	**Other bioassays and assessments**
Bibel *et al.* (2008)	Glutamatergic neurons	Mouse	ESCs	No	*App, p75, Trka, Trkb* and *Trkc*	No	APP, GLUR1, SYP, p75 and TRKS	GABA, ISL1, NES, PAX6, RC2, TH, TUJ1, VGLUT1 and VGAT	Whole cell patch clamp	No	No	No
Gaspard *et al.* 2008	Glutamatergic neurons	Mouse	ESCs (*TAU::GFP*)	No	*Bf1, Dlx1, Dlx5, Emx1, Emx2, Lhx6, Nkx2-2* and *Shh*	No	No	ASCL1, BrdU, CHAT, COUP-TFI, COUP-TFII, CTIP2, CUX1, CALB2, CASP3, Cytokeratin, EN1, ETV1, FOXP2,GABRA6, GFAP, GFP, GSX2, HNF4, MAP2, MATH1, MKI67, Myosin, NKX2-1, NKX2-2, NES, OCT3/4, OTX1, OTX2, P73, PAX6, PECAM1, RET, RELN, SATB2, SOX5, TBR1, TBR2, TH, TLE4, TUJ1, VGAT, VGLUT1 and VGLUT2	Whole cell patch clamp	No	Transplantation into prenatal mice (*in utero*), transplantation into neonatal mice and postnatal mouse brain slice co-culture assay	BrdU proliferation assay (Birth-date analysis) and clonal analyses

Eiraku *et al.* (2008)	Glutamatergic neurons	Mouse and human	ESCs (*Sox1::GFP* and *Bf1::Venus*)	FACS of Bf1::Venus+ cells	No	No	No	BF1, BLBP, BrdU, CALB2, CAMKIIA, CDH2, COUP-TFI, CTIP2, CUX1,EMX1, TUBG1, GAD67, GFP, GRIA1, MAP2, OTX2, P73, Phosphorylated-H3, PROM1, PRKCA, RELN, SATB2, SOX1, SYP, TAU, TBR1, tBR2, TTR, TUJ1, VGLUT1 and ZO1	No	Calcium Imaging	Transplantation into neonatal mice and embryonic mouse forebrain tissue co-culture assay	BrdU proliferation assay (Birth-date analysis)
Li *et al.* (2009)	Glutamatergic neurons	Human	ESCs	No	*BF1, DLX2, EMX1, EN1, GAPDH, GLI3, HOXB4, NKX2-1, PAX6* and *WNT3A*	Microarray (Affymetrix Human GENOME U133 Plus 2.0 Array)	ACTB, GLI3, TBR1, DARPP32 and GAD65/67	BF1, CTIP2, DARPP32, GABA, GAD65/67, HOXB4, NKX2-1, PAX6, TBR1, TUJ1 and VGLUT1	Whole cell patch clamp	No	No	No
Espuny-Camacho *et al.* (2013)	Glutamatergic neurons	Human	ESCs (*ACTB::GFP*) and iPSCs	No	*BF1, BLBP, TDGF1, MYC, CTIP2, CUX1, CUX2, DLX2, DNMT3B, EMX1, EMX2, ETV1, FOXP1, GAPDH,*	Microarray (Affymetrix Human GENOME U133 Plus 2.0 Array)	No	α-SMA, AFP, BF1, BLBP, BrdU, BRN2, CALR, CHAT, COUP-TFI, CTIP2, CUX1, FOXP2, GAD67, GFAP,GFP, Homer1, HuNu clone 235-1, MAP2, NANOG, NES, NKX2-1, OCT3/4, OTX1/2, PAX6, SAFB2, SOX2, SSEA3,	Whole cell patch clamp (*in vitro* and *ex vivo*)	Calcium Imaging	Transplantation into neonatal mice	Axonal tracing

					GBX2, GSX2, IRX3, KLF4, NANOG, NKX2-2, OCT3/4, TOX1, PAX6, RELN, REX1, SATB2, SOX1, SOX2, TBR2, TBR1, TERT, VGLUT1, VGLUT2 and *YWHAZ*			SSEA4, SYP, TBR1, TBR2, TRA-1-60, TRA-1-81, TH, TUJ1, VGAT and VGLUT1				
Kim *et al.* (2011c)	Glutamatergic neurons	Human	ESCs and iPSCs	No	*AFP, ASCL1, BF1, BMP4, DACH1, EMX2, GAPDH, GFAP, GLI3, HOXA4, HOXB4, MAP2, NANOG, NCAM1, NKX2-1, OCT3/4, PAX6, PSD95, SHH, SIX3, SOX1, SOX2,*	No	No	Alkaline phosphatase, CALR, FLAG M2, FORSE-1, GFAP, GFP, NANOG, NES, OCT3/4, SYN1, TRA-1-81 and TUJ1	Whole cell patch clamp	No	No	Artificial synapse formation assay and electron microscopy

					SYT1 and *T/BRACHYURY*							
Shi *et al.* (2012a; 2012b)	Glutamatergic neurons	Human	ESCs and iPSCs	No	*BF1, DLX1, EMX1, HOXB4, ISL1, NKX2-1, OTX1, PAX6, SOX1, SOX2* and *TBR2*	No	No	Acetylated Tubulin, Atypical PKC, BF1, BrdU, BRN2, CD133, CDH2, CTIP2, CUX1, DCX, GAD67, GFAP, GFP, Gamma Tubulin, Homer1, MAP2, MKI67, MUNC13-1, NCAM, SOX2, NES, OTX1/2, PAX6, PSD95, Phosphorylated-H3, Phosphorylated-Ser55, Phosphorylated-Vimentin, S100, SATB2, SYP, TBR1, TBR2, Transferrin Receptor, TUJ1, VGLUT1 and ZO1	Whole cell patch clamp	No	Embryonic mouse cortical brain slice co-culture assay	BrdU proliferation assay (Birth-date analysis)
Barberi *et al.* (2003)	GABAergic neurons	Mouse	ESCs	No	*Bf1, Bmpr1b, Chat, Chrd, Dat, Dbx1, Dlx-1, Eaac1, En1, Fgf8, Fgfr3, Fst, Gad67, Gapdh,*	No	No	5-HT, CNP, EN1, GABA, GFAP, HB9, MBP, MSI1, NCAM, NES, NG2, O1, O4, TH and TUJ1	No	No	No	No

					Gfap, Gli1, Gli3, Hb9, Irx3, Isl1, Lmx1b, Mag, Map2, Mbp, Msi1, Nes, Ngn2, Nkx2-2, Nkx6-1, Nurr1, Oct3/4, Olig2, Osp, Otx2, Pax2, Pax5, Pax6, Pax7, Pitx3, Shh, Sox10, SOX2, Tuj1 and *Th*							
Maroof *et al.* (2010)	GABAergic neurons	Mouse	ESCs (*Lhx6::GFP*)	FACS of Lhx6-GFP+ cells	No	No	No	BF1, CALB2, DLX2, GABA, GFP, KV3.1, LHX6, MKI67, NKX2-1, NPY, PV, SST and TBR1	Whole cell patch clamp	No	Transplantation into neonatal mice	No
Watanabe *et al.* (2005)	GABAergic neurons	Mouse	ESCs (*Sox1::GFP*)	FACS of Sox1-GFP+, CDH1+ and OCT3/4+ cells	*Dlx5, Fgf5, Foxp1, Nodal, Oct3/4, Sox1, T/Brachyury,*	No	SMAD1, SMAD5, SMAD, Phosphorylated-SMAD1, Phosphorylated-	5-HT, ASCL1, BrdU, CDH1, GABA, GAD65, GAD67, GFP, HOXB4, ISL1/2, NES, NKX2-1, NKX2-2, OCT3/4, OTX1, PAX2,	No	Activity-dependent GABA release (HPLC)	No	Apoptosis analysis (TUNEL and annexin V method) and BrdU proliferation assay (Birth-date analysis)

					Tdgf1, Wnt1, Wnt10b, Wnt2b, Wnt3, Wnt3a, Rex1 and *Zfp503*		SMAD5 and Phosphorylated-SMAD8	PAX6, RC2, TH, TUJ1 and Phosphorylated-histone H3				
Li *et al.* (2009)	GABAergic neurons	Human	ESCs	No	*BF1, DLX2, EMX1, EN1, GAPDH, GLI3, HOXB4, NKX2-1, PAX6* and *WNT3A*	No	ACTB, DARPP32, GAD65/67, GLI3 and TBR1	BF1, CTIP2, DARPP32, GABA, GAD65/67, HOXB4, NKX2-1, PAX6, TBR1, TUJ1 and VGLUT1	Whole cell patch clamp	No	No	No
Liu *et al.* (2013)	GABAergic neurons	Human	ESCs and iPSCs	No	No	No	No	BF1, CALB1, GABA, NNOS, NKX2-1, PV, SST and TUJ1	Whole cell patch clamp	No	No	No
Carri *et al.* (2013)	GABAergic neurons	Human	ESCs and iPSCs	No	*ARPP-21, ASCL1, DACH1, DARPP-32, DLX5, DLX6, DRD1, EAR, EMX2, FOXG1, FOXP1, FOXP2, GAPDH, GBX2, ISL1, LIX1, LMO3, MSX1,*	No	No	5-HT, A2A, BF1, CALB1, CALR, CHAT, CTIP2, DARPP-32, DRD2, FOXP1, FOXP2, GABA, GAD65/67, GATA4, GFAP, GSX2, MAP2ab, NES, NPY, OCT4, OTX2, PAX6, PV, SOX1, SST, SYP, TH, TUJ1, VGLUT1, p75 and α-SMA	Whole cell patch clamp	No	No	No

					NKX2-1, OTX1, OTX2, PAX6 and *SIX3*							
Nicholas *et al.* (2013)	GABAergic neurons	Human	ESCs (*NKX2-1::GFP*) and iPSCs	FACS of NKX2-1-GFP+ cells	*ASCL1, BF1, CALB2, DLX2, GAD67, GAPDH, LHX6, NKX2-1, NPY, OLIG2, PV, POU5F1, SST* and *TUJ1*	Microarray (Illumina Human HT-12 BeadChip)	No	ASCL1, BF1, CALB1, CDH2, CR, CHAT, CTIP2, CXCR4, DARPP32, DCX, DLX2, ETV1, GABA, GFAP, GFP, HuNu, ISL1, LHX6, MKI67, NEUN, NKX2-1, OLIG2, PAX6, PV, RAX, RFP, SST, TBR1, TH, TUJ1 and VGAT	Whole cell patch clamp	No	Transplantation into CB.17 SCID mice and mouse cortical glia cell co-culture	No
Maroof *et al.* (2013)	GABAergic neurons	Human	ESCs (*NKX2-1::GFP*) and iPSCs	FACS of NKX2-1-GFP+ cells	No	Microarray (Illumina Human HT-12 BeadChip)	GAPDH, LHX6 and RAX	ASCL1, BF1, CALB1, CALB2, CHAT, DLX2, DCX, FOXA2, GABA, GAPDH, Gephyrin, GFAP, GFP, LHX6, MAP2, MKI67, NES, NKX2-1, NNOS, OLIG2, PV, PAX6, PSD95, RAX, SATB2, SC121, SST, TBR1, TH, TUJ1, VGAT, VGLUT1 and VIP	Whole cell patch clamp	No	Transplantation into neonatal mice, mouse brain slice co-culture assay, slice migration analysis and human cortical projection neuron co-culture assay	Operetta – automated high content image analysis
Lin *et al.* (2015)	GABAergic neurons	Human	Neural precursor (ReNcell VM)	No	*ADORA2A, ARPP-21, ASCL1, BAX, BCL2,*	No	No	5-HT, ASCL1, BF1, CALB1, CALB2, CHAT, CTIP2, DARPP32, DRD2, GABA, GFAP, MAP2,	No	Calcium Imaging and GABA release analysis (ELISA)	No	No

					CALB1, CDC20, CTIP2, DARPP32, DRD1, DRD2, EMX2, FOXP1, FOXP2, GBX2, GFAP, GRIA1, GRIK5, GRIN2B, HTT, MAP2, MBP, MEIS2, MYC, NES, NKX2-1, PAX6, REST, RPL19, SOX2 and *TUJ1*			MEIS2, NES, NKX2-1, NPY, O1, PAX6, PV, SOX2, SST, TH, TUJ1 and VGLUT1				
Barberi *et al.* (2003)	Dopaminergic neurons	Mouse	ESCs	No	*Bf1, Bmpr1b, Chat, Chrd, Dat, Dbx1, Dlx-1, Eaac1, En1, Fgf8, Fgfr3, Fst, Gad67, Gapdh, Gfap, Gli1,*	No	No	5-HT, CNP, EN1, GABA, GFAP, HB9, MBP, MSI1, NCAM, NES, NG2, O1, O4, TH and TUJ1	Whole cell patch clamp	Activity-dependent dopamine release (HPLC)	Transplantation into 6-OHDA lesioned mice	Electron microscopy

					Gli3, Hb9, Irx3, Isl1, Lmx1b, Mag, Map2, Mbp, Msi1, Nes, Ngn2, Nkx2-2, Nkx6-1, Nurr1, Oct3/4, Olig2, Osp, Otx2, Pax2, Pax5, Pax6, Pax7, Pitx3, Shh, Sox10, SOX2, Tuj1 and *Th*							
Kawasaki *et al.* (2000)	Dopaminergic neurons	Mouse	ESCs	No	*Gapdh, Nurr1* and *Pitx3*	No	No	5-HT, CDH1, DβH, FLK1, GAD, GFAP, KRT14, MF20, NCAM, NES, PDGFRA, RC2, SYP, TH, TUJ1 and VACHT	No	Activity-dependent dopamine release (HPLC)	Transplantation into 6-OHDA lesioned mice	No
Lee *et al.* (2000)	Dopaminergic neurons	Mouse	ESCs	No	*Actb, En1, Fgfr3, Fgf8, Gli1, Nes, Nurr1, Otx1, Otx2,*	No	No	5-HT, DβH, GABA, TH and TUJ1	Whole cell patch clamp	Activity-dependent dopamine release (HPLC)	No	No

					Pax2, Pax5, Ret, Shh, Smo and *Wnt1*							
Lim *et al.* (2015)	Dopaminergic neurons	Human	ESCs	No	*EN1, GAPDH, GIRK2, LMX1B, NURR1, OCT3/4, PITX3, SMAD1, SMAD5, SMAD8, TH* and *TUJ1*	No	ACTB, LMNB1, Phosphorylated-SMAD1/5/8 and SMAD1	ACTB, MKI67, NES, SMAD1/5/8, SMAD2/3, TH and TUJ1	No	No	No	Cytosolic and nuclear fractionation
Kirkeby *et al.* (2012)	Dopaminergic neurons	Human	ESCs	No	*ACTB, BF1, CORIN, DLX2, EN1, EN2, FOXA1, FOXA2, GAPDH, GBX2, GDF7, HOXA2, HOXA4, IRX3, LEF1, LHX1, LHX2, LMX1A, LMX1B, MAP2, NANOG, NKX2-1, NKX6-1, NTN1, OCT3/4, OTX2,*	No	No	5-HT, AADC, CTNNB1, CORIN, EN1, FOXA2, GABA, GIRK2, NCAM, HuNu, LMX1A, MAP2, NES, NURR1, PCNA and TH	Whole cell patch clamp	Activity-dependent dopamine release (FAST)	Transplantation into 6-OHDA lesioned rats	No

					PAX5, PAX6, PAX7, SHH, SIM1, SIX3, SYP, TH and *WNT1*							
Yan *et al.* (2005)	Dopaminergic neurons	Human	ESCs	No	*BF1, EN1, GAPDH, GBX2, LMX1B, NKX6-1, NURR1, PAX2, PITX3, SHH* and *WNT1*	No	No	AADC, BF1, CALB1, CCK8, DβH, EN1, GABA, NES, OCT3/4, OTX2, PAX6, PNMT, RET, SSEA4, SOX1, SYP, TH, TUJ1 and VMAT2	Whole cell patch clamp	Activity-dependent dopamine release (HPLC)	No	No
Swistowski *et al.* (2010)	Dopaminergic neurons	Human	iPSCs	No	*AADC, DAT, EN1, GIRK2, LMX1B, MSX1, NURR1, OTX2, TH* and *VMAT*	Microarray (Illumina Human HT-12 BeadChip)	ACTB, CTNNB1, OCT3/4, MYC and NR5A2	GALC, GFAP, GIRK2, HuNu, MSI1, NES, OCT3/4, SOX1, TH, TH clone TH-16 and TUJ1	No	No	Transplantation into 6-OHDA lesioned rats	No
Cai *et al.* (2010)	Dopaminergic neurons	Human	iPSCs	No	*ALDH1A1, CHAT, DβH, FOXA2, GAD67, GAPDH, GLAST, LMX1A, MSX1, NURR1, PITX3, TH* and *TPH1*	No	No	ALDH1A1, LMX1A, MKI67, NES, HuNu, SOX2, TH and TRKB	No	Activity-dependent dopamine release (HPLC)	Transplantation into 6-OHDA lesioned rats	No

Doi *et al.* (2014)	Dopaminergic neurons	Human	iPSCs	FACS of CORIN+ cells	*EN1, FOXA2, GAPDH, GBX2, GSC, LMX1A, MAP2, NANOG, NURR1, OCT3/4, OTX2, SIX3, SOX1, SOX17, T/BRACHYURY* and *TH*	Microarray (Affymetrix GeneChip Whole-Transcript Expression Arrays)	No	5-HT, AADC, CORIN, FOXA2, GIRK2, HuNu, LMX1A, MKI67, NES, NEUN, NURR1, OTX2, PITX3, TH and TUJ1	Whole cell patch clamp	Activity-dependent dopamine release (HPLC)	Transplantation into 6-OHDA lesioned rats	No
Kriks *et al.* (2011)	Dopaminergic neurons	Human	ESCs and iPSCs	No	*ASCL1, FOXA2, LMX1A, NR4A2* and *PITX3*	Microarray (Illumina Human HT-12 BeadChip)	No	5-HT, BF1, CALB1, Collagen, CTNNB1, Cytokeratin, DAT, DCX, ED1, FOXA2, GABA, GFAP, GFP, GIRK2, NCAM, HuNu, Human cell adhesion molecule, Human specific cytoplasm (SC-121), IBA1, LMX1A, MKI67, NES, NURR1, OCT3/4, OTX2, PAX6, PITX3, SYN1, TH, TUJ1 and VMAT2	Whole cell patch clamp	Activity-dependent dopamine release (HPLC)	6-OHDA lesioned mice and rats, and MPTP lesioned rhesus monkeys	No
Xi *et al.* (2012)	Dopaminergic neurons	Human and primate	ESCs and iPSCs	No	No	No	No	5-HT, ALDH1A1, ASCL1, BF1, CORIN, EN1, FOXA2, FOXP1, GABA, GIRK2, HOXA2, HOXB1, LMX1A, LMX1B,	Whole cell patch clamp	No	No	No

								NES, NGN2, NKX6-1, NURR1, PAX6, TH, TUJ1 and OTX2				
Barberi *et al.* (2003)	Serotonergic neurons	Mouse	ESCs	No	*Bf1, Bmpr1b, Chat, Chrd, Dat, Dbx1, Dlx-1, Eaac1, En1, Fgf8, Fgfr3, Fst, Gad67, Gapdh, Gfap, Gli1, Gli3, Hb9, Irx3, Isl1, Lmx1b, Mag, Map2, Mbp, Msi1, Nes, Ngn2, Nkx2-2, Nkx6-1, Nurr1, Oct3/4, Olig2, Osp, Otx2, Pax2, Pax5, Pax6, Pax7, Pitx3, Shh, Sox10,*	No	No	5-HT, CNP, EN1, GABA, GFAP, HB9, MBP, MSI1, NCAM, NES, NG2, O1, O4, TH and TUJ1	No	No	No	No

					SOX2, Tuj1 and *Th*							
Lee *et al.* (2000)	Serotonergic neurons	Mouse	ESCs	No	*Actb, En1, Fgfr3, Fgf8, Gli1, Nes, Nurr1, Otx1, Otx2, Pax2, Pax5, Ret, Shh, Smo* and *Wnt1*	No	No	5-HT, DβH, GABA, TH and TUJ1	No	No	No	No
Shimada *et al.* (2012)	Serotonergic neurons	Mouse	ESCs (*ePet::EGFP*) and iPSCs	FACS of ePet-EGFP+ cells.	*Actb, Ascl1, Chat, Gfap, Gad67, Map2, Mbp, Nes, Nkx2-2, Pet-1, Phox2b, Sert, Shh, Th* and *Tph2*	No	No	5-HT, GABA, MAP2, NES, TH and TUJ1	Whole cell patch clamp	Activity-dependent 5-HT release (HPLC)	No	No
Kumar *et al.* (2009)	Serotonergic neurons	Human	ESCs	No	*ASCL1, FEV, GAPDH, GATA2, LMX1B, NKX2-2* and *SERT*	No	No	5-HT, BLBP, GFAP, MAP2, NES, TPH, TUJ1 and Vimentin	No	Activity-dependent 5-HT release	No	5-HT imaging (Three-photon microscopy) and analysis of the relative 5-HT content
Vadodaria *et al.* (2015)	Serotonergic neurons	Human	ESCs (*TPH2::GFP*,	FACS of TPH2-GFP+	*FEV, GATA2, LHX6,*	No	No	5-HT, GABA, MAP2ab, TH, TPH and TPH2	No	No	No	No

			TPH2::TdT and *SYN1::dsRed*)	and SYN1-dsRed+ cells	*LMX1B, NKX2-2* and *TPH2*							
Lu *et al.* (2016)	Serotonergic neurons	Human	ESCs and iPSCs	FACS OF NKX2-2+ cells.	*BF1, CDH2, CDX2, DBX1, EGR2, EMX1, EN1, FEV, GAPDH, GBX2, HOXA1, HOXA2, HOXA3, HOXA4, HOXB1, HOXB2, HOXB3, HOXB4, HOXB6, HOXB8, HOXC5, HOXC6, IRX3, LHX1, LMX1A, LMX1B, NANOG, NKX2-1, NKX2-2, NKX6-1, OCT3/4, OLIG2, OTX2, PAX6, PAX7, SERT, SIM1, SIX3, SOX1, SOX17,*	No	FEV, GAPDH and SERT	5-HT, AADC, BF1, EN1, FEV, FOXA2, GABA, GAPDH, GATA2, GATA3, GFAP, HOXA2, HOXA3, HOXB1, HOXB4, HTR1A, MKI67, NES, NKX2-2, NKX6-1, OCT3/4, OLIG2, OTX2, PAX3/7, PAX6, PHOX2B, SERT, SOX1, SOX2, TH, TPH2, TUJ1 and VMAT2	Whole cell patch clamp	Activity-dependent 5-HT release (UPLC-ESI-MS/MS)	No	No

					T/BRACHYURY, TBR2 and *TPH2*							
Barberi *et al.* (2003)	Cholinergic/Motor neurons	Mouse	ESCs	No	*Bf1, Bmpr1b, Chat, Chrd, Dat, Dbx1, Dlx-1, Eaac1, En1, Fgf8, Fgfr3, Fst, Gad67, Gapdh, Gfap, Gli1, Gli3, Hb9, Irx3, Isl1, Lmx1b, Mag, Map2, Mbp, Msi1, Nes, Ngn2, Nkx2-2, Nkx6-1, Nurr1, Oct3/4, Olig2, Osp, Otx2, Pax2, Pax5, Pax6, Pax7, Pitx3, Shh,*	No	No	5-HT, CNP, EN1, GABA, GFAP, HB9, MBP, MSI1, NCAM, NES, NG2, O1, O4, TH and TUJ1	No	No	No	No

					Sox10, SOX2, Tuj1 and Th							
Wichterle *et al.* (2002)	Cholinergic/Motor neurons	Mouse	ESCs (*Hb9:GFP*)	FACS of Hb9-GFP+ cells	No	No	No	CHAT, CHX10, DBX1, EN1, GFP, HB9, HOXC5, HOXC6, IRX3, ISL1, LHX3, LIM1, LIM2, NCAM, NEUN, Neurofilament, NKX2-2, NKX6-1, OLIG2, OTX2, PAX6, PAX7, SC1, SOX1, SYB, SYN, TUJ1 and VACHT	No	No	Chick spinal cords (*in ovo*)	BrdU proliferation assay (Birth-date analysis)
Li *et al.* (2005)	Cholinergic/Motor neurons	Human	ESCs	No	*GAPDH, HOXB1, HOXB6, HOXC5, HOXC8, IRX3, NKX2-2, NKX6-1, OLIG2, OTX2, PAX6, SHH* and *SOX1*	No	No	CHAT, HB9, ISL1/2, LHX3, NES, OLIG2, OTX2, PAX6, PAX7, SOX1, TUJ1 and VACHT	Whole cell patch clamp	No	Mouse myocyte co-culture	No
Singh Roy *et al.* (2005)	Cholinergic/Motor neurons	Human	ESCs (*Hb9:EGFP*)	FACS of Hb9-EGFP+ cells	*EN1, ISL1, OCT3/4* and *PAX7*	No	No	CHAT, Desmin, GFP, HB9, ISL1, SSEA4, SV2, TRA 1-81 and TUJ1	Whole cell patch clamp	No	Rat muscle co-culture assay	No
Dimos *et al.* (2008)	Cholinergic/Motor neurons	Human	ESCs and iPSCs	No	*MYC, FOXD3, GAPDH, KLF4, NANOG, OCT3/4, REX1,*	No	No	α-SMA, AFP, CHAT, Desmin, GFAP, HB9, ISL1/2, NANOG, SSEA-3, SSEA4, TE-7, TRA-1-60,	No	No	No	No

					SOD1, SOX3, TDGF1 and *TERT*			TRA-1-81 and TUJ1				
Qu *et al.* (2014)	Cholinergic/Motor neurons	Human	ESCs and iPSCs	No	*ARX, BCL2, BRN2, CACNA1A, CEP290, CHAT, CHX10, CNTNAP2, COUP-TFI, COUP-TFII, CREBBP, CTNNA2, EGFL7, EGR2, EMX2, EP300, EVI1, FEZF1, FEZF2, FOXA2, GAS1, GBX2, GNPAT, HB9, HES3, HESX1, HESX1, HOXC4, HSPA5, KAT2B, LEF1, LHX2, MAFB, MTPN,*	No	GAPDH, NANOG, OCT3/4, PAX6, SOX1 and ZIC1	CHAT, CHX10, HB9, ISL1, MAP2, NANOG, OCT3/4, SMI-32, SV2, SYN1 and TUJ1	Whole cell patch clamp	No	Mouse myocyte co-culture	Synaptic structure and neuromuscular junction analysis (Multiphoton microscopy)

					NCOA1, NCOA2, NCOA3, NEUROD1, NKX6-1, OTX1, POU3F1, RARA, RARB, RARG, RAX, RORA, RXRA, RXRB, RXRG, SDF4, SIX3, SMO, SOX3, ULK1, WNT7B and *ZEB2*							
Karumbayaram *et al.* (2009)	Cholinergic/Motor neurons	Human	ESCs and iPSCs (*Hb9:EGFP*)	No	No	No	No	BRN2, CHAT, HOXA3, HOXA5, HOXC6, HOXC8, ISL1, LHX3, NKX6-1, OLIG2, PAX6, PAX7, NES, SOX1, SOX2, SOX3 and TUJ1	Whole cell patch clamp	No	No	No
Amoroso et al. (2013)	Cholinergic/Motor neurons	Human	ESCs and iPSCs (*Hb9:GFP*)	FACS of Hb9-GFP+ cells	*CHT1, FOXP1, HOXA5, HOXC6, HOXC8, HOXD9, LHX3* and *RALDH2*	Whole transcriptome RNA sequencing (Illumina HiSeq)	No	BRN3A, FOXP1, GFP, HB9, ISL1, LHX3, NF-H, RALDH2 and TUJ1	Whole cell patch clamp	Calcium Imaging	Chick spinal cords (*in ovo*)	No

Shimojo *et al.* (2015)	Cholinergic/Motor neurons	Human	ESCs and iPSCs (*HB9::Venus*)	FACS of HB9-Venus+ cells.	*ACTB, CHAT, EBNA-1, HB9, ISL1, KLF4, MYCL, LIN28, NANOG, NGN2, NKX2-2, OCT3/4, OLIG2, PAX6, SOX1* and *SOX2*	No	ACTB, CHAT, HB9 and ISL1	CHAT, GFP, HB9, ISL1, MYHC and TUJ1	No	No	Human myocyte co-culture	No
Lee *et al.* (2007)	Cholinergic/Motor neurons	Human and primate	ESCs (*RUES1-EGFP*)	No	*BF1, CHAT, GAPDH, HOXB4, ISL1, NGN2, NKX2-2, NKX6-1, OLIG2, OLIG3, PAX6, PAX7, SOX1* and *VACHT*	Microarray (Affymetrix Human GENOME U133 Plus 2.0 Array)	No	BF1, CHAT, HB9, HOXB4, HOXC6, HOXC8, HuNu, ISL1, LHX3, NCAM, NKX2-2, OLIG2, OTX2, PAX6, PAX7, SOX1 and TUJ1	Whole cell patch clamp	Activity-dependent acetylcholine release (HPLC)	Chick eggs (*in ovo*) and rat spinal cords (*in vivo*)	No
Neuronal phenotypes, species and starting cell types are summarized for each protocol according to their reference. Cell sorting techniques and the phenotypic characterization, consisting of assays such as (quantitative) polymerase chain reaction, genome-wide expression analysis, western-blot, immunohistochemistry, electrophysiological assessment, biochemical assessment, transplantation in to animal models, co-culture assays or other bioassays, are broadly outlined if applicable.												

Supplementary Table 2. Overview of the assays that have been used for the phenotypic characterization of the transcription factor-mediated neuronal reprogramming protocols.

References	Phenotypes	Species	Starting cell types	Cell sorting	Phenotypic characterization							
					(Quantitative) Polymerase chain reaction	Genome-wide expression analysis	Western-Blot	Immunohistochemistry	Electrophysiological assessment	Biochemical assessment	Transplantations /Co-culture assay	Other bioassays and assessments
Vierbuchen *et al.* (2010)	Glutamatergic neurons	Mouse	Fibroblasts (*Tau::EGFP*)	FACS of Tau-EGFP+ cells	*Actb, Sox1 and Sox10*	No	No	5-HT, BrdU, CHAT, CALB2, GABA, GAD65, GFAP, MAP2, NEUN, NKX2-2, OLIG1, PAX3, PAX6, PAX7, PRPH, SOX2, SYN1, TH, TUJ1 and VGLUT1	Whole cell patch clamp	No	Neonatal mouse neuron co-culture assay and postnatal mouse astrocyte co-culture assay	BrdU proliferation assay (Birth-date analysis)
Heinrich *et al.* (2010)	Glutamatergic neurons	Mouse	Astocytes	No	*Aldh1l1, Emx1, Emx2, Gapdh, Gfap, Glt1, Glu1, Ngn2, S100β, Sox2* and *Tuj1*	No	No	CALR, CAMKIIα, GAD1, GFAP, GFP, MAP2, RFP, SYN1, TBR1, TBR2, TUJ1, VGAT and VGLUT1	Whole cell patch clamp	Calcium imaging	No	No
Thoma *et al.* (2012)	Glutamatergic neurons	Mouse	ESCs	No	*Afp, Dcx, Ef1a1, Gad1, Insulin, Math3, Myod1, Nanog, Neun, Ngn2, Ngn2 (Transient), Ngn2 (Inducible), Oct3/4,*	No	No	MAP2, MAP2ab, MYC, NANOG, NR1, STAT3, SYN1, TAU, TH, TUJ1 and VGLUT1	Whole cell patch clamp	No	Mouse hippocampal neuron co-culture assay	No

					Olig2, Pax6, Pu1, Sox1, Th, Vglu2 and *Vglut1*							
Chanda *et al.* (2014)	Glutamatergic neurons	Mouse and human	Fibroblasts (Tau::EGFP) and ESCs	No	*Ascl1, Ngn2, Myt1l, Brn2* and *Gapdh*	Single-cell gene expression analysis (Fluidigm dynamic array)	No	GFP, MAP2, NEUN, SYN1, TUJ1, VGAT and VGLUT1	Whole cell patch clamp	No	Mouse hippocampal neuron co-culture assay	No
Pang *et al.* (2011)	Glutamatergic neurons	Human	Fibroblasts (*Tau::EGFP*), ESCs and iPSCs	No	*ACTB, ASCL1, Ascl1, BRN1, BRN2, Brn2, CAMKIIB, DCX, GAD65, GAD67, GAPDH, GFAP, MAP2, MYT1L, Myt1l, NCAM, NES, NEUROD1, Neurod1, OLIG2, P75, PAX6, PV, SOX10, SOX9, SYN1, TH, TUJ1, VGAT, VGLUT1, VGLUT2*	Single-cell gene expression analysis (Fluidigm dynamic array)	No	ASCL1, BrdU, BRN2, GAD65, GFAP, LU5, MAP2, NEUN, Neurofilament, PRPH, SOX10, SYN1, SYT1, TBR1, TH, TUJ1 and VGLUT1	Whole cell patch clamp	No	Mouse cortical neuron co-culture assay	No

					and *VGLUT3*							
Zhang et al. (2013)	Glutamatergic neurons	Human	ESCs and iPSCs	No	No	Single-cell gene expression analysis (Fluidigm dynamic array)	ACTB, CPX1/2, DLX1, MUNC18-1, NES, NEUN, OCT3/4, PSD-95, SNAP 25, SOX2, SYB2, SYN1, SYNT1A/B, SYT1 and TUJ1	GAD6, GFAP, MAP2, NANOG, NEUN, OCT3/4, SOX2, SYN1, TUJ1 and VGLUT1	Whole cell patch clamp	Calcium imaging	Glia-neuron co-culture assay and transplantation into mice striatum	Optogenetic based presynaptic assesment
Addis *et al.* (2011)	Dopaminergic neurons	Mouse	Astrocytes (*MAP2::CD4* and *MAP2::GCaMP3*) and fibroblasts	MACS of MAP2-CD4+ cells	*Aldh1a7, Brn2, Cacna1g, Calb1, Dat, Ddc, En1, Foxa2, Gabra1, Gad65, Grin1, Kcnn3, Lmx1a, Msx1, Ngn2, Pax2, Pax5, Pitx3, Scn3a, Syn2, Tau, Th, Vglut1* and *Vmat2*	No	No	GIRK2, OTX2, SYP, TH, TUJ1 and V5	Whole cell patch clamp	Activity dependent dopamine release (HPLC) and calcium Imaging	No	No
Kim *et al.* (2011b)	Dopaminergic neurons	Mouse	Fibroblasts (*Pitx3::EGFP*)	FACS of Pitx3-EGFP+ cells	*Aadc, Dat, Th* and *Vmat2*	No	No	AADC, DAT, EGFP, MAP2, PITX3, TH and TUJ1	Whole cell patch clamp	Activity dependent dopamine release (HPLC)	Transplantation into 6-OHDA lesioned mice	Southern blot

Caiazzo *et al.* (2011)	Dopaminergic neurons	Mouse and human	Mouse fibroblasts (TH::GFP) and human fibroblasts	FACS of TH-GFP+ cells	*ALDH1A1, Aldh1a1, Ascl1 (Viral), DAT, Dat, Drd2, En1, GAPDH, Gapdh, Lmx1a, Lmx1a (Viral), Lmx1b, Ngn2, Nurr1, Nurr1 (Viral), Otx2, Pitx3, Sox2, TH, Th, VMAT2* and *Vmat2*	Microarray (Affymetrix Gene-Chip Mouse Gene ST1.0)	No	AADC, ALDH1A1, BrdU, CALB1, DAT, DRD2, GFP, MAP2, OTX2, SYN1, SYT1, TH, TUJ1 and VMAT2	Whole cell patch clamp (*in vitro* and *in vivo*)	Activity dependent dopamine release (HPLC) and FM4-64 assay	Transplantation into postnatal mice	Evoked catecholamine exocytosis (Amperometric recording), bisulfite promotor sequencing (*Th* and *Vmat2*) and electron microscopy
Pfisterer *et al.* (2011)	Dopaminergic neurons	Human	Fibroblasts	No	*ACTB, Ascl1 (Viral), BLBP, Brn2 (Viral), En1 (Viral), Foxa2 (Viral), GAPDH, Gli1 (Viral), Lmx1a (Viral), Lmx1b (Viral), MAP2, Msx1 (Viral),*	No	No	5-HT, AADC, BLBP, Collagen I, Collagen II, GABA, Glutamate, MAP2, NES, NGN2, NKX2-2, NURR1, P75, PAX6, PAX7, PRPH, SOX1, SOX2, SYP, TE-7, TH and TUJ1	Whole cell patch clamp	No	No	No

					Myt1l (Viral), NGN2, Nurr1 (Viral), Otx2 (Viral), Pax2 (Viral), Pax5 (Viral), PAX6, PAX7, SOX1, SOX2 and *SOX10*							
Theka *et al.* (2013)	Dopaminergic neurons	Human	iPSCs	No	*ALDH1A1, ASCL1, Ascl1 (Viral), CALB1, CORIN, DAT, DDC, DRD2, EN1, FOXA2, GAPDH, GIRK2, LMX1B, LMX1A, Lmx1a (Viral), NURR1, Nurr1 (Viral), OTX2, PITX3, TH* and *VMAT2*	No	No	ALDH1A1, CALB1, DDC, FOXA2, GFP, GIRK2, MAP2, NANOG, OCT3/4, SOX2, SYT1, TH, TRA-1-60, TUJ1 and VMAT2	Whole cell patch clamp	Activity dependent dopamine release (HPLC)	Transplantation into postnatal mice	No
Wasko (2013)	GABAergic neurons	Mouse	Fibroblasts	No	No	No	No	GABA, MAP2, NEUN, TUJ1 and VGLUT1	No	No	No	No

Heinrich *et al.* (2010)	GABAergic neurons	Mouse	Astocytes	No	*Aldh1l1, Emx1, Emx2, Gapdh, Gfap, Glt1, Glu1, Ngn2, S100ß, Sox2* and *Tuj1*	No	No	CALR, CAMKIIα, GAD1, GFAP, GFP, MAP2, RFP, SYN1, TBR1, TBR2, TUJ1, VGAT and VGLUT1	Whole cell patch clamp	Calcium imaging	No	No
Colasante *et al.* (2015)	GABAergic neurons	Mouse and human	Mouse fibroblasts (*GAD67::GFP*), human firbroblasts and human iPSCs	FACS of GAD67-GFP+ cells	*18s, ACTB, Arx, Ascl1, Bf1, Calb1, Calb2, Chat, Col24a1, Col9a2, Coup-tfI, Coup-tfII, Darpp32, DLX1, DLX2,Dlx2, Dlx5, Dlx6, Erbb4, Etv1, GAD67, GAD65, Gbx2, Gfap, Helt, Isl1, Kv3.1, Lhx6, Lhx8, Mafb, Mef2c, Nkx2-1, Olig2, PV, Ptf1a, Reln,*	Whole transcriptome RNA sequencing (NEBNext Ultra Directional RNA Library Prep Kit and Illumina HiSeq 2500)	ASCL1, Sox2 and V5	ARX, CALB1, CALB2, GABA, GAD65/67, GFP, HOMER1, KV3.1B, MAP2, NPY, NAV, PV, RFP, SST, SYT1, TUJ1 and VGAT	Whole cell patch clamp	GABA concentration measurements (ELISA)	Transplantation into adult mice hippocampus (*in vivo)* and rat hippocampal co-culture assay	Functional GABAergic neurotransmission (optogenetics), pre-embedding immunogold localization experiment (electron microscopy) and chromatin immunoprecipitation (for ASCL1, BF1 and SOX2)

					SST, Satb1, Sox2, Sox6, Sp8, Tbx1, Tbx2, VGLUT1, Vglut1 and *VIP*							
Xu *et al.* (2015)	Serotonergic neurons	Human	Fibroblasts	No	*AADC, ALDH1A1, ASCL1, ASCL1 (Viral), CACNA1C, FEV, FEV (Viral), FOXA2, FOXA2 (Viral), GAPDH, LMX1B, LMX1B (Viral), MAP2, hp53shRNA (Viral), PCLO, SCN1A, SERT, SYN1, TPH1, TPH2, TUJ1* and *VMAT2*	No	No	5-HT, AADC, ALDH1A1, MAP2, NEUN, SERT, STX1A, TPH2, TUJ1 and VMAT2	Whole cell patch clamp	Activity dependent 5-HT release and 5-HT uptake analysis (both HPLC)	No	No
Vadodaria *et al.* (2015)	Serotonergic neurons	Human	Fibroblasts (TPH2::GFP and SYN1::dsRed)	FACS of TPH2-GFP+ and SYN1-dsRed+ cells	*FEV, GATA2, LHX6, LMX1B, NKX2-2* and *TPH2*	Whole transcriptome RNA sequencing (Illumina HiSeq 2500)	FEV, GAPDH and NKX2-2	5-HT, FEV, GABA, GATA2, GFAP, GFP, LMX1B, MAP2ab, NEUN, NKX2-2, SERT, TAU, TH,	Whole cell patch clamp	Calcium imaging and 5-HT concentration measurem	Rat astrocyte co-culture	Drug assay (SSRIs)

								TPH, TPH2, TUJ1 and Vimentin		ents (ELISA)		
Mazzoni *et al.* (2013)	Cholinergic/Motor neurons	Mouse	ESCs (*Hb9::GFP*)	FACS of Hb9-GFP+ cells	*Actb, Chat, Hb9, Isl1, Lhx3, Ngn2, Phox2a* and *Vacht*	Microarray (Affymetrix Mouse Gene 1.0 ST)	V5	CHAT, HB9, ISL1, MKI67, NCAM, NEUN, NGN2, OLIG2, PHOX2B, SV2A, TUJ1, V5 and VACHT	Whole cell patch clamp	No	No	Chromatin immunoprecipitation (for Isl1 and V5), co-immunoprecipitation and DNA motif analysis
Mong *et al.* (2014)	Cholinergic/Motor neurons	Mouse and human	ESCs	MACS of PSA-NCAM+ cells	*CHAT, DβH, L19, ISL1, Isl1, PHOX2A, Phox2a, PHOX2B, Phox2b, PRPH, NET, Net, Tbx20, Th, Tlx3* and *Tuj1*	Whole transcriptome sequencing (Affymetrix Gene-Chip Mouse Exon ST1.1)	No	NET, PHOX2A, PHOX2B, PSA-NCAM, PRPH, TH and TUJ1	No	No	No	High-throughput imaging drug assay
Son *et al.* (2011)	Cholinergic/Motor neurons	Mouse and human	Mouse fibroblasts (*Hb9::GFP*) and human fibroblasts	FACS of Hb9-GFP+ cells	*Ascl1, Brn2, Hb9, Isl1, LTR (Viral), Lhx3, Myt1l* and *Ngn2*	No	No	ACHR, CHX10, HB9, ISL1, SYN1, TH and VACHT	Whole cell patch clamp	No	Mouse C2C12 myoblast co-culture assay, chick myotube co-culture assay, chick eggs (*in ovo)*, glia-neuron co-culture assay	BrdU proliferation assay (Birth-date analysis)
Hester *et al.* (2011)	Cholinergic/Motor neurons	Human	ESCs (*Hb9::GFP*) and iPSCs	No	*ACTB, CHAT, HB9, ISL1, LHX3, NANOG, NEUROD1, OCT3/4, OLIG2, PAX6, REX1* and *SOX2*	No	No	ACTA1, CHAT, HB9, HOXC6, ISL1, LHX3, NANOG, NGN2, OCT3/4, OLIG2, OTX2, PAX6, RHD-BTX, TUJ1, SMI31, SOX2, SSEA4 and SV2A	Whole cell patch clamp	No	Mouse C2C12 myotube co-culture assay	No

Mong *et al.* (2014)	Noradrenergic neurons	Mouse and human	ESCs	MACS of PSA-NCAM+ cells	*CHAT, DβH, L19, ISL1, Isl1, PHOX2A, Phox2a, PHOX2B, Phox2b, PRPH, NET, Net, Tbx20, Th, Tlx3* and *Tuj1*	Microarray (Affymetrix Gene-Chip Mouse Exon ST1.1)	No	NET, PHOX2A, PHOX2B, PSA-NCAM, PRPH, TH and TUJ1	No	No	No	High-throughput imaging drug assay

Neuronal phenotypes, species and starting cell types are summarized for each protocol according to their reference. Cell sorting techniques and the phenotypic characterization, consisting of assays such as (quantitative) polymerase chain reaction, genome-wide expression analysis, western-blot, immunohistochemistry, electrophysiological assessment, biochemical assessment, transplantation in to animal models, co-culture assays or other bioassays, are broadly outlined if applicable.

Supplementary Table 3. Overview of the assays that have been used for the phenotypic characterization of the epigenetic-based neuronal conversion approaches.

					Phenotypic characterization							
References	**Phenotypes**	**Species**	**Starting cell types**	**Cell sorting**	**(Quantitative) Polymerase chain reaction**	**Genome-wide expression analysis**	**Western-Blot**	**Immunohistochemistry**	**Electrophysiological assessment**	**Biochemical assessment**	**Transplantations /Co-culture assay**	**Other bioassays and assessments**
Raciti *et al.* (2013)	NPCs	Mouse	Fibroblasts (*Trp53+/+, Trp53−/−, Sox1+/+, Sox1EGFP/+, Tau+/+* and *TauEGFP/+*)	FACS of Sox1-EGFP+ and Tau-EGFP+ cells	*Bf1, Emx1, Emx2, Gapdh, Hes5, Ngn2* and *Pax6*	No	No	BF1, BrdU, GFAP, MAP2, NANOG, NEUN, PAX6, SOX2 and TUJ1	Whole cell patch clamp	No	Transplantation into neonatal mice	BrdU proliferation assay (Birth-date analysis) and NPC-differentiation into neurons
Cheng *et al.* (2015b)	NPCs	Mouse and human	Mouse fibroblasts and human urinary cells	No	*Ascl1, Blbp, Brn2, NANOG, Nanog, NES, OCT3/4, Oct3/4, PAX6, Pax6,*	Microarray (8 × 60K array from Agilent Technologies)	No	GAD67, GFAP, Glutamate, MAP2, MBP, MKI67, NES, NEUN, OLIG2, PAX6, SOX2, SYN1 and TUJ1	Whole cell patch clamp	Alkaline phosphatase analysis	Transplantation into embryonic mice	NPC-differentiation into neurons and glia

					SOX1, SOX2, Sox2, HPRT and *Hprt*							
Cheng *et al.* (2015a)	Dopaminergic, GABAergic, glutamatergic and cholinergic motor neurons	Mouse	Astrocytes (*GFAP*::*GFP* and *Neurod1*::*GFP*)	No	*Asl1, Dll1, Dll3, Dll4, Dlx2, Hes1, Hes5, Hprt1, Jag1, Jag2, Neurod1, Ngn2, Notch1, Notch2, Notch3, Notch4, Pax6* and *Sox2*	No	No	CHAT, DCX, GAD67, GFP, MAP2, NEUN, SOX2, TH, TUJ1, VLGUT1 and VMAT2	Whole cell patch clamp	No	No	No
Black *et al.* (2016)	Neurons (Unspecified)	Mouse	Fibroblasts (*SYN1*::*RFP*)	No	*ASCL1, Ascl1 (Endogenous), Ascl1 (Total), Brn2 (Endogenous), Brn2 (Total), Myt1l (Endogenous), Myt1l (Total), Tuj1, gRNA, GAPDH* and *Gapdh*	No	ASCL1, BRN2, FLAG, GAPDH and MYT1L	ASCL1, BRN2, MAP2 and TUJ1	Whole cell patch clamp	No	No	ChiP-seq and ChIP-qPCR (*Brn2, Ascl1* and *Myt1l* - *H3K27ac* and *H3K4me3*)
Chavez *et al.* (2015)	Neurons (Unspecified)	Human	iPSCs	No	*ACTB, ACTC1, ASCL1,*	No	No	Neurofilament 200 and TUJ1	No	No	No	No

					MIAT, NEUROD 1, NGN2, RHOXF2, TTN and *VEGF*							
Low *et al.* (2012)	Neurons (Unspecified)	Mouse	NPCs	No	*Actb, Gfap, Map2, Nes, Rest, Rip* and *Tuj1*	No	ACTB and REST	GFAP, MAP2, NES, O4 and TUJ1	No	No	No	Examination of neurite length and outgrowth
Yang *et al.* (2012)	Dopaminergic neurons	Mouse	ESCs (*TH::GFP*)	FACS of TH-GFP+ cells	*Actb, Bdnf, Gdnf, Nurr1, Pitx3, Pre-miR-132* and *Th*	RNA expression analysis (TaqMan Low Density Assay v2.0)	ACTB and NURR1	MAP2, NES and TH	No	No	No	No
Kim *et al.* (2007)	Dopaminergic neurons	Mouse	ESCs	FACS of PITX3+ cells	*ACTB, Actb, miR133b* (Human) and *miR133b* (Mouse)	No	No	BrdU, GABA, miR-133, TH and TUJ1	No	Activity-dependent dopamine release (HPLC)	No	Apoptosis analysis (TUNEL) and *Pitx3* luciferase assay
Stappert *et al.* (2013)	Dopaminergic neurons	Human	NPCs	No	*18S, DAT, GAD67, NURR1* and *TH*	MiRNA expression analysis (ABI microRNA multiplex TaqMan assay)	No	BrdU, GAD65/67, TH and TUJ1	No	No	No	BrdU proliferation assay (Birth-date analysis) and non-radioactive northern blotting
Victor *et al.* (2014)	GABAergic neurons	Human	Fibroblasts (*SYN1::EGFP*)	FACS	*ANK2, ASCL1, Bcl-xL, BDNF, CALB2, CHAT, CHRM4, CTIP2, DβH,*	Single-cell gene expression analysis (Fluidigm dynamic array)	No	Ankyrin G, CTIP2, DARPP32, DLX5, FOXP1, GABA, GAD67, MAP2, NEUN, SCN1A, SYN1, TH, TUJ1 and VGLUT1	Whole cell patch clamp (*in vitro* and *ex vivo*)	No	Transplantation into neonatal NSG mice	No

					DCX, DDC, DLX1, DLX1 3'UTR, DLX2, DLX2 3'UTR, DRD1, DRD2, GAD65, GAD67, GAPDH, GPR6, GRP, HPRT1, HTR2C, HTR3A, LHX6, MAP2, MYT1L, NCAM1, NES, NEUROD 1, NKX2-1, OPRM1, OTX1, OTX2, P75, PAX6, PCP2, PDYN, PENK, POU3F1, PPP1R1B, Pri-miR-9 Loci 3, Pri-miR-124 Loci 1, PRPH, PSD95, PVALB, RARB,							

					RPS18, SCN2A, SCN3A, SHANK3, SLC17A6, SLC17A8, SLC6A3, SLC6A4, SST, TAC1, TAU, TBR1, TBR2, TH, TPH2, Transgenic Pri-miR9124 and *TUJ1*							
Yoo *et al.* (2011)	Glutamatergic and GABAergic neurons	Human	Fibroblasts	FACS of MAP2+/TUJ1+ cells	*18S, BDNF, BSN, CACNA1C, CRIM1, CTIP2, CUX1, DCX, DDC, DKK3, DLX1, DLX5, ETV1, FOXP1, FOXP2, GAD67, GAPDH, GRP, HTR2C, LXN, MEF2C, NCAM, NR4A3, NSE, OMA1, PCLO,*	Single-cell gene expression analysis (Fluidigm dynamic array)	BAF53A and GAPDH	BAF45B, BAF45C, BAF53B, GABA, GRIN1, MAP2, Neurofilament 200, SCN1A, SV2, SYN1, TUJ1 and VGLUT1	Whole cell patch clamp	Calcium imaging and FM1-43 imaging	No	EdU proliferation assay (Birth-date analysis)

					PCP2, PERIPHERIN, PLXND1, PPP1R1B-DARPP32F, S100A10, SATB2, SCN1A, SCN2A, SCN3A, SCN8A, SEMA3E, SHANK3, SLC1A2, SOX5, SYN1, SYT9, TAU, TBR1, TH, TIS21-BTG2, TLE1, TLE4, TPM2, TUJ1, GRM5, UNC5D and *VLGUT1*							
Ambasudhan *et al.* (2011)	Glutamatergic and GABAergic neurons	Human	Fibroblasts	No	*BAF53B, BRN2, BRN2 (Viral), KRT1, miR-124, MYT1L, MYT1L (Viral), P4HA1, P75, PAX6,*	No	No	GABA, GFAP, KRT1, MAP2, NEUN, NKX2-2, P75, PAX6, PERIPHERIN, PHD4, SOX2, SYN1, TH, TUJ1, VGAT and VGLUT1	Whole cell patch clamp	No	No	EdU proliferation assay (Birth-date analysis)

					SOX1 and *SOX10*							
Abernathy et al. (2017)	Cholinergic motor neurons	Human	Fibroblasts	No	*CHAT, COL13A1, HB9, HOXA7, HOXA9, HOXB4, HOXB7, HOXC10, HOXC11, HOXC6, HOXC8, HOXC9, HOXD11, HOXD8, HOXD9, HPRT, hsa-miR-218 S100A4, RNU-44, SLC18A3* and *VIM*	Microarray (Ilumina TotalPrep kits for Agilent Human 4x44Kv1), RNA-seq and TRAP RNA-seq	No	FSP1, MAP2, KI67, TUJ1, NEUN, NCAM, MNX1, SCN1A, ANKG, SV2, CHAT, SMI-32	Whole cell patch clamp	No	Human myotue co-culture assay	ATAC-seq, DREM analysis, MeDIP-seq, MRE-seq and GO enrichment analysis
Neuronal phenotypes, species and starting cell types are summarized for each protocol according to their reference. Cell sorting techniques and the phenotypic characterization, consisting of assays such as (quantitative) polymerase chain reaction, genome-wide expression analysis, western-blot, immunohistochemistry, electrophysiological assessment, biochemical assessment, transplantation in to animal models, co-culture assays or other bioassays, are broadly outlined if applicable.												

Chapter 8

Establishment and characterization of a human neuronal *in vitro* model system for Alzheimer's disease using induced pluripotent stem cells: An exploratory approach

Renzo J.M. Riemens[1,2], Katherine Basil[1], Cengiz Akbulut[3], Leon J. Schurgers[3], Daniël L.A. van den Hove[1,4], Jos C. Kleinjans[5], Twan van den Beucken[5†], Gunter Kenis[1†]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Biochemistry, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, the Netherlands.

† Authors contributed equally to this work

Abstract
In recent years, the establishment of induced pluripotent stem cells (iPSCs) has offered an alternative *in vitro* model system for Alzheimer's disease (AD). While still in their infancy, these developing models hold considerable promise to push forward the efforts to disentangle AD's pathophysiology. However, before the potential of these models can be fully realized, it is imperative to set up a robust neuronal differentiation protocol relevant for AD, as this represents a milestone that needs to be achieved in order to subsequently develop an adequate *in vitro* model. Here we present preliminary data on the characterization of a cortical forebrain differentiation protocol from iPSCs. Based on the step-wise application of neuralizing factors, iPSCs were differentiated towards neural progenitor cells (NPCs) and subsequently cortical neurons and glia, which represents cellular populations known to be affected in the pathophysiology of AD. Based on assessing the expression of key-lineage transcription factors at each stage of the differentiation process we confirmed that the cells adopted towards a cortical forebrain fate. Furthermore, we assessed the expression of Tenascin XB (TNXB) and the oxytocin receptor (OXTR), as we have previously identified *TNXB* and oxytocin (*OXT*)-signaling to be associated with AD in genome-wide methylation screenings. Both genes were detectable after 14 days and their expression maintained up until 56 days of NPC differentiation. Taken together, we successfully established a differentiation protocol for cortical forebrain cells that will enable us to perform mechanistic investigation of candidate genes for AD.

Keywords
Induced pluripotent stem cells (iPSCs), Alzheimer's disease, cortical neurons, directed differentiation.

Introduction

Research into AD has changed rapidly over the last years as iPSCs derived from patients are now being widely explored as an alternative model system to study the disease in a more humanized *in vitro* setting. One could claim that the establishment of this technology represents as yet one of the most innovative biomedical advances of this century, mainly because these patient-specific cells contain the molecular information from the donors, and therefore might be able to provide new insights into the disease that could contribute to the development of (personalized) therapeutics. Another major innovative character of this approach is that, at least in theory, all major brain cell types can be obtained by directed differentiation of the iPSCs, which allows for cell-specific studies *in vitro* and the establishment of increasingly complex human co-culture systems relevant for AD [1]. By applying this technology, pioneering studies have furthermore already demonstrated that cellular AD hallmarks can be found in patient-derived neural cells differentiated from iPSCs when compared to those derived from healthy controls [2-7]. Thus, while still in their relative infancy, iPSC-based neural models are thought to hold considerable promise to push AD research forward.

Although significant advances have been made in generating various neural cells from iPSCs, *in vitro* neuronal differentiation is actually not a process that is fully disciplined yet, since protocols for specific neural subtypes vary substantially in time scale and efficiency, both within and between different laboratories [1]. On that account, setting up a robust neuronal differentiation protocol relevant for AD represents a milestone that needs to be achieved *a priori* in order to subsequently develop adequate AD models. In the present study, we therefore present preliminary data on the characterization of a cortical forebrain differentiation protocol from iPSCs. Based on the step-wise application of neural patterning factors, we differentiated iPSCs towards neural progenitor cells (NPCs) and subsequently cortical neurons and glia, which represent cellular populations known to be affected in the pathophysiology of AD [8]. We postulate that the application of this differentiation protocol on both patient- and healthy control-derived iPSCs might provide an appealing model to study AD. We furthermore assessed the expression of Tenascin XB (TNXB) and the oxytocin receptor (OXTR) in the differentiated neural cells. TNXB and oxytocin (OXT)-signaling have previously been associated with AD pathophysiology, and iPSC-based studies could help in determining their specific molecular mechanisms in relation to the disease [9-11]. All in all, the pilot data presented here represents an exploratory approach characterizing a cortical forebrain differentiation protocol in iPSC with the eventual aim to apply this for the establishment of an AD-related disease model.

Materials and methods

Sample and ethics statement

Peripheral blood samples were obtained from a healthy Southern European male that was in his late twenties at time of donation. Mononuclear blood cells isolated from the peripheral blood samples were reprogrammed towards iPSCs using the Epi5™ Episomal iPSC Reprogramming Kit (Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA) in combination with a Nucleofector™ Device (Lonza, Bazel, Switzerland). All laboratory procedures for iPSC reprogramming were performed at the Cardiovascular Research Institute Maastricht (CARIM; Maastricht University, Maastricht, the Netherlands). Clones of growing iPSC colonies were picked, expanded and characterized according to general quality control guidelines, including tests such as the evaluation of pluripotency marker expression, assessment of transgene silencing, karyotyping by G-banding, embryoid body (EB) induction and directed differentiation into the three embryonic germ layers [12]. A single clone, *i.e.* CARIMi001-A, which surpassed the quality control criteria, was selected and used in the present study. For further details on the used iPSC line, please refer to https://hpscreg.eu/cell-line/CARIMi001-A.

Induced pluripotent stem cell maintenance

iPSC colonies were maintained in Essential 8 flex medium (Gibco™, Thermofisher Scientific) on plates coated with human embryonic stem cell-qualified Geltrex™ (Gibco™, Thermofisher Scientific) and regularly passaged upon confluence using UltraPure™ EDTA (Invitrogen™, Thermo Fisher Scientific). Passaged iPSCs were incubated overnight with medium containing 10 µM ROCK Inhibitor Y-27632 (Stemcell Technologies, Vancouver, Canada) in order to enhance cell survival and medium was refreshed every other day thereafter. Cells were regularly checked and verified to be free of mycoplasma contamination.

Neural progenitor cell induction

NPCs were differentiated based on a previously described cortical forebrain differentiation protocol, with small adaptations [13]. In brief, clumps of iPSC colonies were sparsely plated and grown up until profound single rounded colonies were present in the culture. The colonies were then enzymatically dissociated using collagenase IV (Gibco™, Thermofisher Scientific) and transferred to ultra-low-attachment plates (Costar®, Corning, New York, NY, USA) in order to generate free-floating EBs. The generating EBs were maintained in neuronal induction medium (NIM) on an orbital shaker for the first 24 hours and then as static culture for a total time of two weeks. The NIM consisted of DMEM/F-12 GlutaMAX™ (Gibco™, Thermofisher Scientific) with N-2™ (1X; Gibco™,

Thermofisher Scientific) and B-27™ (1X; Gibco™, Thermofisher Scientific), supplemented with 100 ng/ml recombinant human Noggin (Peprotech, Cranbury, NJ, USA) 100 nM LDN193189 (StemMacs™, Miltenyi Biotec, Bergisch Gladbach, Germany) and 10 nM SB431542 (StemMacs™, Miltenyi Biotec), *i.e.* dual SMAD inhibition [14]. After two weeks of incubation, EBs were plated onto poly-L-ornithine- (PLO; Sigma Aldrich, Saint Louis, MO, USA) and mouse laminin- (Gibco™, Thermofisher Scientific) coated plates in NIM containing 20 ng/ml human fibroblast growth factors 2 (FGF2; Miltenyi Biotec) and 1 µg/ml laminin. After 48 hours, adhered EBs displaying outgrowth of neural projections were picked, dissociated using Accutase® (Corning) and plated onto PLO/laminin-coated plates in order to generate neuronal progenitor cells (NPCs). The NPCs were maintained at high density as monolayers in NIM containing 20 ng/ml human FGF2 and regularly passaged using Accutase®. Passaged NPCs were incubated overnight with 10 µM ROCK Inhibitor Y-27632 in order to enhance cell survival and medium was refreshed every other day thereafter.

Cortical forebrain differentiation

For terminal differentiation towards cortical neural cells, NPC were plated at lower densities (25,000 to 50,000 cells/cm^2) on PLO/laminin-coated plates, while maintained in neuronal differentiation medium (NDM) for up to 2, 4 or 8 weeks. NDM consisted of N-2 and B-27 supplemented DMEM/F12:Neurobasal™ media (1:1; Gibco™, Thermofisher Scientific) with 20 ng/ml recombinant human brain-derived neurotrophic factor (BDNF; Peprotech), 20 ng/ml human recombinant glial cell-derived neurotrophic factor (GDNF; Peprotech), 1mM dibutyrl cyclic-AMP (dcAMP; Sigma-Aldrich), 200 nM ascorbic acid (AA; Stemcell Technologies), 1 µg/ml laminin, non-essential amino acids (NEAA (1X); Gibco™, Thermofisher Scientific) and antibiotic-antimycotic solution (1X; Gibco™, Thermofisher Scientific). Plated NPCs were incubated overnight with 10 µM ROCK Inhibitor Y-27632 in order to enhance cell survival and medium was refreshed every other day thereafter.

Immunocytochemisty and cell quantification

Cells were washed three times with 1X phosphate buffered saline (PBS) and subsequently fixated for 15 minutes at room temperature using 4% paraformaldehyde in PBS. The cells were then washed three times for 5 minutes each with 1X PBS and permeabilized using 0.1% Triton-X100 in PBS for 5 minutes. Following three subsequent washing steps with PBS lasting 5 minutes each, the cell were blocked for one hour at room temperature using 10% of bovine serum albumin (BSA) diluted in PBS. After blocking, cells were incubated at 4°C overnight with primary antibodies diluted in 10% BSA blocking solution. The next day, the cells were washed three times for 10 minutes using 1X PBS. The cells

were then incubated with the corresponding Alexa Fluor® conjugated secondary antibodies diluted in 10% BSA for one hour at room temperature. Following incubation, the cells were washed three times for 10 minutes using 1X PBS and counterstained with Hoechst (1:500) diluted in PBS. Please refer to Supplementary Table S1 for an overview of used antibodies in this study. Fluorescent images were taken using an iX81 microscope (Olympus Life Science, Olympus Corporation, Waltham, MA, USA) with the Stereo Investigator software (MBF, Bioscience, Williston, VT, USA). For NPC cell counting, 15,000 cells were plated per well of a 24-well plate and fixated after 48 hours in culture. Images of the cells in five visual fields were randomly taken and loaded into the FIJI ImageJ software (https://imagej.nih.gov/ij/index.html) [15]. On average, 318 ± 39 cells per staining were manually counted using the ImageJ point tool function.

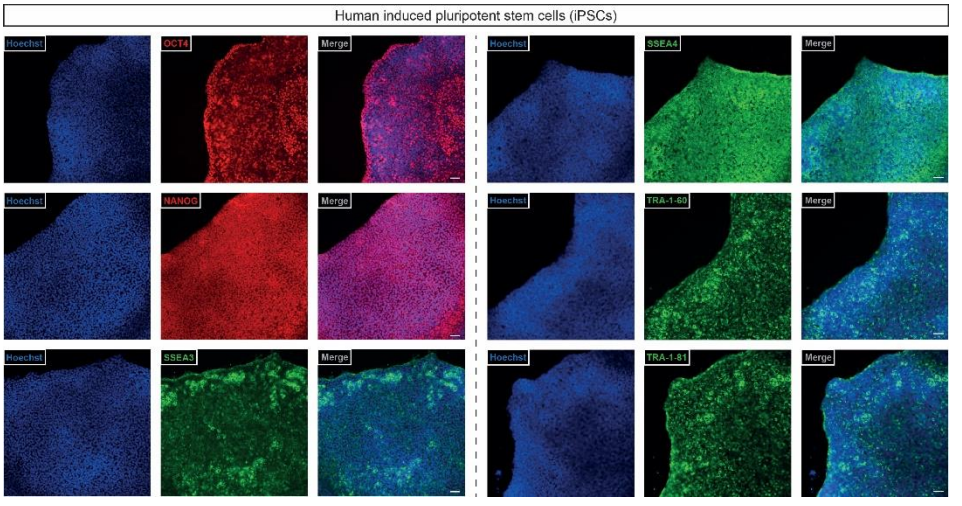


Figure 1. Characterization of stemness markers in human induced pluripotent stem cells (iPSCs). The human iPSC line, *i.e.* CARIMi001-A, was stained for the stemness markers OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 and TRA-1-81. Images are displayed for the nuclear counterstaining with Hoechst, the respective stemness marker and the overlay of both images (Merge). Scale bars represent 100 µM.

Results

Characterization of human iPSCs and derived cortical forebrain neurons and glia

First, the iPSC line used in the present study, *i.e.* CARIM-i001A, was stained for several commonly used iPSC markers. As shown in Figure 1, growing iPSC colonies exhibited a typical round morphology with well-defined sharp edges and tightly packed cells that stained positive for OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 and TRA-1-81. Next, iPSCs were differentiated to NPCs via a

combination of dual SMAD inhibition and EB formation (see Material and methods, as well as Figure 2 for details). After 4 weeks of NPC differentiation, cells were characterized for expression of NPC-specific markers. As shown in Figure 3, the NPCs stained positive for NESTIN (85.89±7.14%), PAX6 (93.10±5.85%) and SOX2 (91.28±6.39%). Double labeling with each of these markers revealed that 82.13±5.81% of the Hoechst-positive cells stained for both NESTIN and PAX6, 85.69±8.62% for NESTIN and SOX2, and 88.44±8.06% for SOX2 and PAX6, indicating that most of the iPSCs differentiated into NPCs.

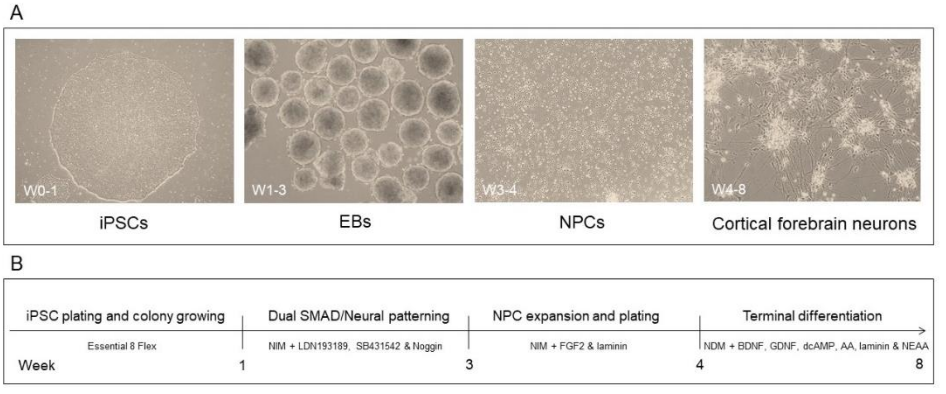


Figure 2. Overview of the cortical forebrain differentiation protocol in induced pluripotent stem cells (iPSCs). **A.** Brightfield images represent examples of the cells at different stages in the process of neuronal differentiation, from iPSCs to embryoid bodies (EBs), to neural progenitor cells (NPCs), to cortical forebrain neurons. **B.** The timeline displays the overview of the method for generating human cortical forebrain neurons by inhibition of dual SMAD signaling combined with EB formation, the use of neuronal induction medium (NIM) supplemented with fibroblast growth factor-2 (FGF2) for NPC maintenance and subsequently terminal differentiation towards neurons by using neuronal differentiation medium (NDM) supplemented with, brain-derived neurotrophic factor (BDNF), glial cell-derived neurotrophic factor (GDNF), dibutyrl cyclic-AMP (dcAMP), ascorbic acid (AA), 1 µg/ml laminin, non-essential amino acids (NEAA). The differentiation process starting from the iPSCs towards the cortical forebrain neurons takes approximately 8 weeks.

NPCs were subsequently differentiated towards cortical forebrain cells (see Material and methods, as well as Figure 2 for details) for a total of 4 weeks. While after 2 weeks of differentiation the cells had lost the expression of the pluripotency gene OCT4, expression of the neuronal marker TUJ could be detected (Figure 4A). A more elaborate characterization after 4 weeks of differentiation furthermore confirmed the expression of the neuronal markers MAP2AB, DCX and TUJ1, as well as the glial marker GFAP (Figure 4B). In addition, after 4 weeks of differentiation the neural cells expressed cortical forebrain markers, including CTIP2, a cortical pyramidal neuronal marker found in layer VI-V, and the telencephalic marker FOXG1 (Figure 4C) [16]. Overall, these data show that we successfully generated neurons, as well as glia cells, from these iPSCs with

a cortical forebrain identity that was confirmed based on the expression of the aforementioned key lineage-specific markers.

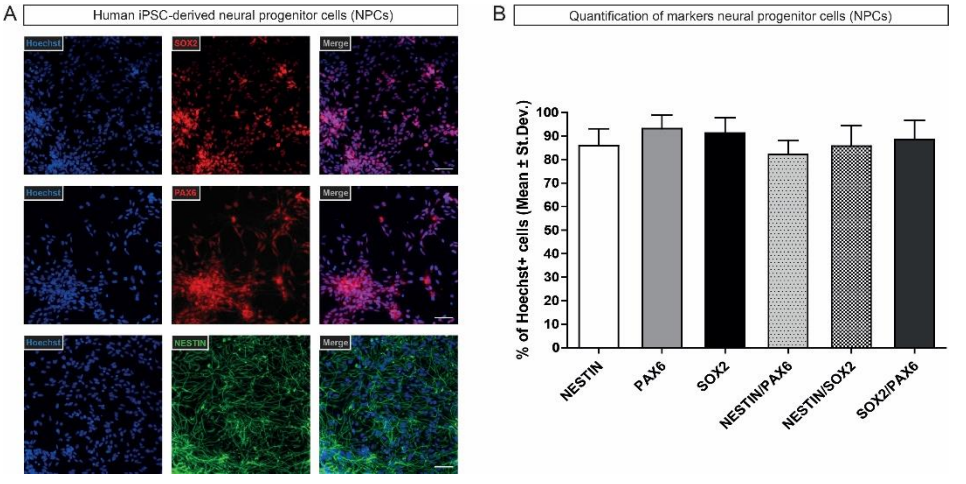


Figure 3. Characterization of the neural progenitor cells (NPCs) derived from the human induced pluripotent stem cells (iPSCs). **A.** Representative fluorescent images of the NPCs staining positive for markers such as NESTIN, SOX2 and PAX6. Images are displayed for the nuclear counterstaining with Hoechst, the respective NPC marker and the overlay of both images (Merge). Scale bars represent 100 µM. **B.** Quantification of these markers demonstrated that most of the iPSC-derived NPCs stained double positive for NESTIN, PAX6 and SOX2.

Cortical forebrain neural cells express TNXB and OXTR

We furthermore assessed the expression of TNXB and OXTR in the differentiated neural cells. Deregulation in the *TNXB* and *OXT*, as well as OXT-signaling have previously been associated with AD and, specifically, OXT administration has therefore been proposed as a potential therapeutic approach for dementia [9, 10, 17-20]. In order to use the iPSC-derived cortical forebrain cells for mechanistic studies targeting these candidates, it is crucial to confirm that the cells described here express the related proteins. Expression of both TNXB and OXTR was assessed at three time-intervals, *i.e.* after 14 days, 28 days and 56 days of NPC-differentiation, in order to determine potential timely differences in their expression during the differentiation and maturation of the neural cells. As shown in Figure 5, expression of both TNXB and OXTR could be identified in the differentiated cortical forebrain neurons, as early as 14 days and maintained up to 56 days of NPC-differentiation. Immunoreactivity of TNXB could be identified extracellularly, whereas OXTR expression was detected at the soma of both TUJ1-postivite and TUJ1-negative cells. These data therefore suggest that the cortical forebrain protocol described here might represent an appealing model system to study processes related to TNXB and OXT signaling in the context of AD.

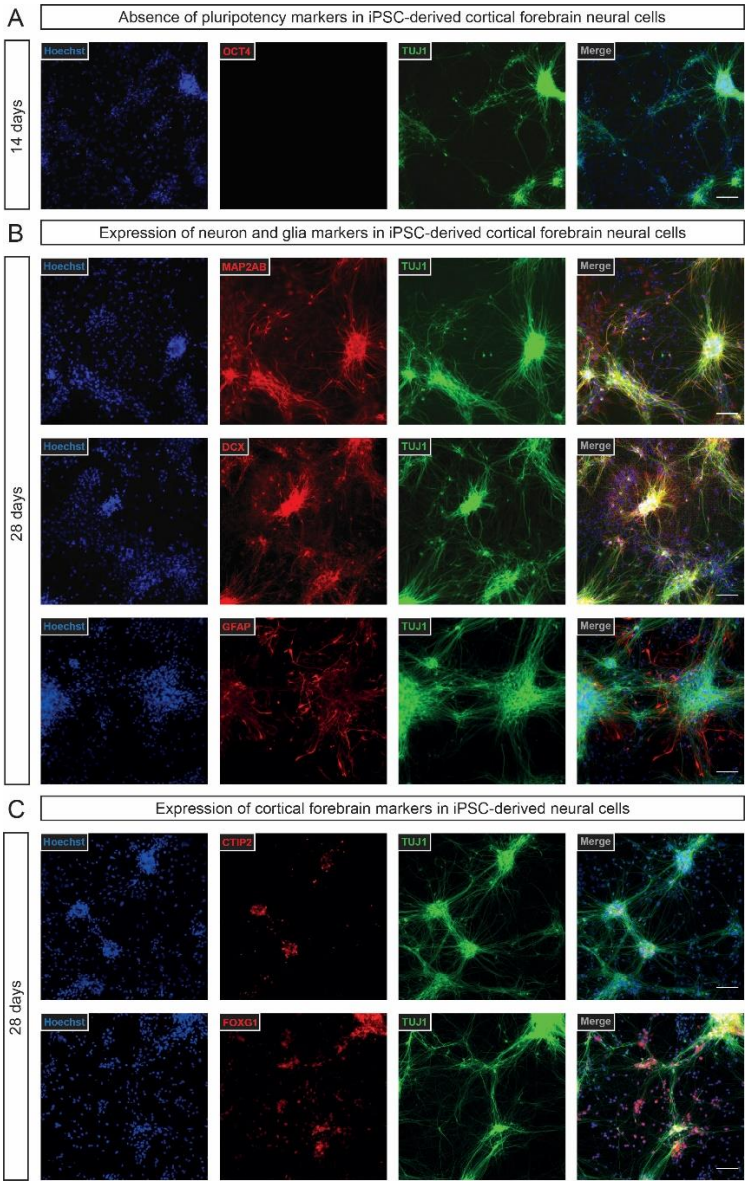


Figure 4. Characterization of the cortical forebrain neurons and glia derived from the human induced pluripotent stem cells (iPSCs). **A.** Expression of pluripotency markers such as OCT4 was not detected in the differentiated cortical forebrain cells 14 days after the start of neural progenitor cell (NPC) differentiation. **B.** Representative fluorescent images of the cells staining positive for the neuronal markers MAP2AB, TUJ1, DCX, as well as the glial marker GFAP 28 days after start of NPC differentiation. Images are displayed for the nuclear counterstaining with Hoechst, the respective marker and the overlay of both images (Merge). **C.** Cells furthermore expressed cortical forebrain markers, including CTIP2, which is cortical pyramidal neuronal marker found in layer VI-V, as well as the telencephalic marker FOXG1. Scale bars represent 100 µM.

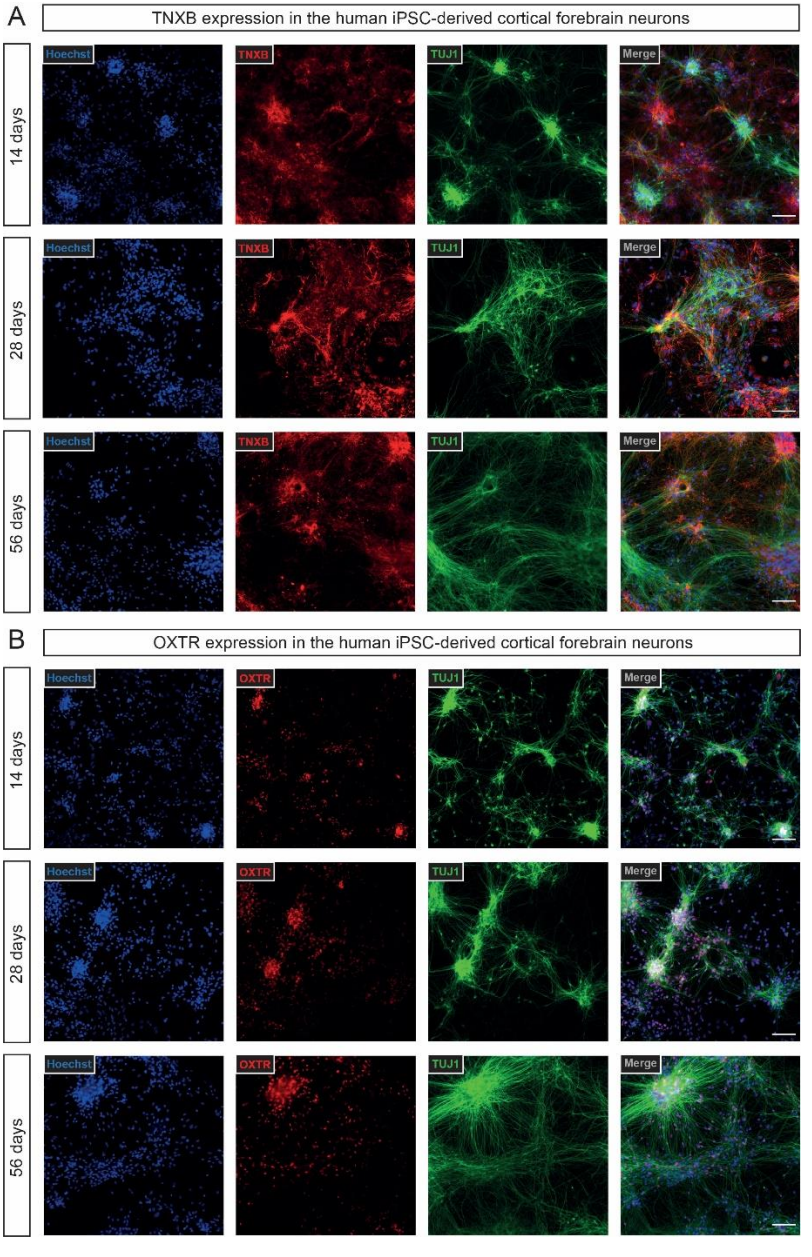


Figure 5. Expression of Tenascin XB (TNXB) and the oxytocin receptor (OXTR) in the cortical forebrain neurons and glia. Images show the expression of both TNXB (**A**) and OXTR (**B**), which were assessed at three time-intervals, *i.e.* after 14 days, 28 days and 56 days of neural progenitor cell (NPC)-differentiation. Images are displayed for the nuclear counterstaining with Hoechst, TNXB or OXTR and the overlay of both images (Merge). Expression of both TNXB and OXTR could be identified in the differentiated cortical forebrain neurons after 14 days and maintained up until 56 days of differentiation. Scale bars represent 100 µM.

Discussion

In the present study, we described the first steps in establishing and characterizing a cortical forebrain differentiation protocol in iPSCs. The findings described here represent preliminary data of a currently ongoing project aimed at developing *in vitro* model systems for the study of AD. For this purpose, the stemness of the iPSC-line, which was used for neuronal differentiation, was first confirmed based on assessing the expression of pluripotency markers such as OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 and TRA-1-81. Next, cells were differentiated based on a chemical stimulation with neuralizing factors known to be involved in guiding development towards a cortical forebrain fate [13]. At each step of the differentiation process, the cells were monitored for their expression of key-lineage-specific markers. As such, we demonstrated that the cells successfully adopted to a forebrain fate and differentiated into NPCs, and subsequently both neurons and glia. Successful derivation of NPCs was confirmed by the presence of markers such as NESTIN, SOX2 and PAX6 (Figure 3). Quantification of these markers furthermore revealed that most of the iPSCs differentiated towards NPCs (>82%) with a comparable efficiency as described previously [13]. Immunofluorescent examination of the differentiated NPCs verified the presence of the neuronal markers such as MAP2AB, DCX and TUJ1, and the astrocytic marker GFAP. Additionally, the presence of cortical forebrain markers, including CTIP2 and FOXG1, were identified. Overall, these preliminary findings confirmed the successful differentiation of the iPSCs in a heterogeneous mixed population of cortical forebrain neurons and glia.

As a next step in the characterization process of the iPSC-derived neuronal cells, it would be imperative to identify the neurotransmitter subtypes that are present in the differentiated cultures. In general, iPSCs that are differentiated *in vitro* towards a cortical forebrain fate produce neurons that predominantly express glutamatergic and gamma-aminobutyric acid (GABA)-ergic markers [1]. For instance, in a study using a similar differentiation protocol it was found that the majority of the iPSC-derived neurons (~80%) were glutamatergic and a smaller subset (~20%) was GABAergic [13]. The eventual derivation of these subtypes is, however, dependent on a variety of technological and biological factors, and variation between, as well as within laboratories, is commonly observed while using the exact same protocol [1]. Furthermore, previous studies have demonstrated that the differentiation propensity for specific neurotransmitter subtypes also depends on the iPSC line used, as processes such as epigenetic memory and complete reprogramming of the cells might affect differentiation efficiency [1, 21]. It would, therefore, be interesting to explore the ratios between these subtypes, as well as other neuronal cells, in the cultures. In addition, next to the expression of both neuronal and glial markers, it would be important to

confirm that the neuronal cells derived from the iPSC also display the functional characteristics of their *in vivo* equivalents. As such, another important milestone to be achieved, would be to confirm that the derived neurons are electrophysiologically active, assessed either by means of calcium imaging or by patch-clamp recordings. Overall, such measures, as well as the ability to replicate these finding in different batches and other iPSC lines, are crucial to address the robustness of the differentiation protocol.

After further confirmation of the aforementioned characteristics, the protocol described here could potentially be used for the establishment of a cortical *in vitro* model, allowing mechanistic studies related to AD. In an attempt to address this potential already at an early stage, we then quantified the expression of TNXB and OXTR in the iPSC-derived cortical cultures. As previously demonstrated in Chapter 4 of this thesis, epigenetic and genetic deregulation of TNXB have been associated with AD pathophysiology in multiple regions of the AD brain, including the cortex [9, 17-19, 22]. In addition, epigenetic deregulation of the OXT gene and alterations in OXT signaling have previously been implicated in AD or associated disease phenotypes [10, 20, 23, 24]. Taken together, these studies demonstrated that both genes and related signaling pathways might be functionally involved in AD; hence, the availability of iPSC-derived cortical neurons from both AD patients and healthy controls might represent an appealing approach to study their role in the disease at the molecular level. Here, as a crucial first step, we confirm the presence of TNXB and OXTR protein in the iPSC-derived cortical cells at different time intervals during the differentiation; both proteins could be detected in the differentiated neural cells already at 14 days and maintained up until 56 days of NPC-differentiation. TNXB is a member of the tenascin family of extracellular matrix glycoproteins [25], and, in line with that, we found TNXB immunoreactivity extracellularly in our differentiated cultures. OXTR expression on the other hand, has mainly been observed in neural cell somas with little expression in the neurites [26]. Although a further detailed characterization remains necessary, we also detected OXTR immunoreactivity in the somas of the iPSC-derived neural cells. Thus, the data presented here demonstrate that the cortical forebrain cells might represent an appealing model system to study mechanistic processes related to TNXB and OXT signaling in the context of AD.

As a next step in validating these cultures for the use of an AD disease model it would be highly informative to explore the toxicological responses of the neural cell towards disease-related environmental insults, including amyloid beta proteins. In combination with an MTT assay one would be able to assess cell viability and to determine a dose-toxicity curve towards these peptides. A

subsequent analysis using both patient- and healthy control-derived cortical neural cells differentiated from iPSCs might then potentially demonstrate whether toxicological responses could differ in these cultures. Within this context, it would furthermore be interesting to interrogate whether genes such as TNXB and OXT or related signaling pathways could mediate resilience or susceptibility towards such insults. Although for TNXB little is known yet about its mechanistic role in the disease, administration of OXT has *e.g.* recently been shown to reverse amyloid beta-induced synaptic impairments in mouse hippocampal slices [27]. As such, it would be interesting to address whether similar effects can be identified in human iPSC-derived neural cultures and to apply this platform for further interrogation of the underlying molecular mechanisms mediating these effects.

Aside from the possibilities addressed above, the iPSC-derived cortical neurons obtained by the protocol described in the present manuscript, offer a plethora of other future research opportunities for AD. For instance, it is increasingly recognized that environmental risk factors with complex gene-environment interactions, play a crucial role in reinforcing AD pathogenesis. However, the exact molecular nature of these interactions, as well as their temporal relationship with the development of AD, remain largely unknown. Furthermore, studies have shown that exposure to risk factors can bring about epigenomic, transcriptomic, proteomic and metabolomic changes related to AD, which can bring about sustained alterations in molecular processes leading into the manifestation of the full-blown disease. These observations that advocate for the role of environmental risk factors in combination with genetic and epigenetic factors, could nowadays be modeled and, hence, studied more thoroughly in iPSC-derived neuronal cells. Within this framework, the recent availability of editing systems based on CRISPR-Cas9, as well as the use of isogenic lines even adds another dimension to these possibilities. All in all, it is therefore anticipated that future efforts specifically focusing on the associations between environmental factors and the aforementioned multi-omics modalities will aid in developing a better understanding about their involvement in AD. Furthermore, studies aimed at comparing both multi-omics profiles between iPSC-derived and *ex vivo* neuronal sub-populations, obtained from both AD patients and healthy controls, would represent another fascinating example that might not only provide us with new knowledge about molecular disease mechanisms, but also could aid in further mapping the strengths and weaknesses of iPSC-derived models.

Finally, next to the cortical neurons, it would also be interesting to explore whether other neural subtypes, *e.g.* serotonergic neurons, can be obtained from the iPSCs using adapted protocols based on alternative patterning factors. Hence, this would allow one to explore whether the aforementioned environmental factors

and molecular alterations relevant to AD behave similarly or differently when studying them in other neural cells derived from iPSCs. Moreover, the availability of different iPSC-derived neural populations also allows for future co-cultures or complex 3D models using scaffolds-based and scaffold-free systems, *e.g.* gels and spheroids, respectively [28]. These latter complex cultures are emerging as an innovative and advanced alternative driven by their resemblance to *in vivo* environmental and architectural characteristics. In fact, these model systems allow for complex intercellular communication and the establishment of intricate spatial organizations that provide physical and chemical cues, which might even be essential for the study of human brain diseases at both cellular and molecular levels. Furthermore, it has even been shown that 3D environments can promote better neuronal differentiation and neural network formation [28-32]. Alternatively, a combination of both *in vitro* differentiation and subsequent transplantation into animal models might represent another opportunity for AD modelling [33]. Such approaches might even capture pathological phenotypes that could not be detected when cells are only cultured *in vitro*. All in all, the availability of iPSC-derived neurons represents an exciting new area to explore for future AD studies.

Acknowledgements

The authors would like to thank Dr. Theodora Saridaki and Daria Kaloujskaia for their assistance during the project. This research was made possible by BReIN (Brightlands e-infrastructure for Neurohealth), an initiative which is co-funded by the Province of Limburg, Maastricht University and Maastricht University Medical Centre + in the Netherlands.

References

1. Riemens RJ, van den Hove DL, Esteller M, Delgado-Morales R. Directing neuronal cell fate in vitro: achievements and challenges. Progress in neurobiology. 2018;168:42-68.
2. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012;482(7384):216-20.
3. Yagi T, Ito D, Okada Y, Akamatsu W, Nihei Y, Yoshizaki T, et al. Modeling familial Alzheimer's disease with induced pluripotent stem cells. Hum Mol Genet. 2011;20(23):4530-9.
4. Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, et al. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. Cell Stem Cell. 2013;12(4):487-96.
5. Muratore CR, Rice HC, Srikanth P, Callahan DG, Shin T, Benjamin LN, et al. The familial Alzheimer's disease APPV717I mutation alters APP processing and Tau expression in iPSC-derived neurons. Hum Mol Genet. 2014;23(13):3523-36.
6. Young JE, Boulanger-Weill J, Williams DA, Woodruff G, Buen F, Revilla AC, et al. Elucidating molecular phenotypes caused by the SORL1 Alzheimer's disease genetic risk factor using human induced pluripotent stem cells. Cell Stem Cell. 2015;16(4):373-85.

7. Sproul AA, Jacob S, Pre D, Kim SH, Nestor MW, Navarro-Sobrino M, et al. Characterization and molecular profiling of PSEN1 familial Alzheimer's disease iPSC-derived neural progenitors. PLoS One. 2014;9(1):e84547.
8. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics & chromatin. 2018;11(1):41.
9. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.
10. Lardenoije R, Roubroeks JA, Pishva E, Leber M, Wagner H, Iatrou A, et al. Alzheimer's disease-associated (hydroxy) methylomic changes in the brain and blood. Clinical epigenetics. 2019;11(1):164.
11. Riemens RJM, Pishva E, Iatrou A, Roubroeks J, Nolz J, Lardenoije R, et al. Brain-region- and cell type-specific epigenetic profiling strongly implicates a role for dysregulation of TNXB and other loci in the brainstem in Alzheimer's disease. 2021.
12. Sullivan S, Stacey GN, Akazawa C, Aoyama N, Baptista R, Bedford P, et al. Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. Regenerative medicine. 2018;13(7):859-66.
13. Vadodaria KC, Ji Y, Skime M, Paquola A, Nelson T, Hall-Flavin D, et al. Serotonin-induced hyperactivity in SSRI-resistant major depressive disorder patient-derived neurons. Molecular psychiatry. 2019;24(6):795-807.
14. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009;27(3):275-80.
15. Rasband WS. National Institutes of Health, Bethesda, Maryland, USA. http://imagej nih gov/ij/. 2011.
16. Espuny-Camacho I, Michelsen KA, Gall D, Linaro D, Hasche A, Bonnefont J, et al. Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron. 2013;77(3):440-56.
17. Chang T, editor Tau Network Genes in a Genome Wide Association Study of Progressive Supranuclear Palsy. MOVEMENT DISORDERS; 2017: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.
18. Chapuis J, Hot D, Hansmannel F, Kerdraon O, Ferreira S, Hubans C, et al. Transcriptomic and genetic studies identify IL-33 as a candidate gene for Alzheimer's disease. Molecular psychiatry. 2009;14(11):1004-16.
19. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. Alzheimer's research & therapy. 2019;11(1):5.
20. Haas BW, Filkowski MM, Cochran RN, Denison L, Ishak A, Nishitani S, et al. Epigenetic modification of OXT and human sociability. Proceedings of the National Academy of Sciences. 2016;113(27):E3816-E23.
21. Kim JE, O'Sullivan ML, Sanchez CA, Hwang M, Israel MA, Brennand K, et al. Investigating synapse formation and function using human pluripotent stem cell-derived neurons. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(7):3005-10.
22. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. Journal of Alzheimer's Disease. 2011;23(2):349-59.
23. Cong L, Jia J, Qin W, Ren Y, Sun Y. Genome-wide analysis of DNA methylation in an APP/PS1 mouse model of Alzheimer's disease. Acta neurologica Belgica. 2014;114(3):195-206.

24. Tampi RR, Maksimowski M, Ahmed M, Tampi DJ. Oxytocin for frontotemporal dementia: a systematic review. Therapeutic advances in psychopharmacology. 2017;7(1):48-53.
25. Aktar R, Peiris M, Fikree A, Cibert-Goton V, Walmsley M, Tough IR, et al. The extracellular matrix glycoprotein tenascin-X regulates peripheral sensory and motor neurones. The Journal of physiology. 2018;596(17):4237-51.
26. Tan Y, Singhal SM, Harden SW, Cahill KM, Nguyen D-TM, Colon-Perez LM, et al. Oxytocin receptors are expressed by glutamatergic prefrontal cortical neurons that selectively modulate social recognition. Journal of Neuroscience. 2019;39(17):3249-63.
27. Takahashi J, Yamada D, Ueta Y, Iwai T, Koga E, Tanabe M, et al. Oxytocin reverses Aβ-induced impairment of hippocampal synaptic plasticity in mice. Biochemical and Biophysical Research Communications. 2020;528(1):174-8.
28. Hernández-Sapiéns MA, Reza-Zaldívar EE, Cevallos RR, Márquez-Aguirre AL, Gazarian K, Canales-Aguirre AA. A three-dimensional Alzheimer's disease cell culture model using iPSC-Derived neurons carrying A246E mutation in PSEN1. Frontiers in Cellular Neuroscience. 2020;14.
29. Choi SH, Kim YH, Hebisch M, Sliwinski C, Lee S, D'Avanzo C, et al. A three-dimensional human neural cell culture model of Alzheimer's disease. Nature. 2014;515(7526):274-8.
30. Ravi M, Paramesh V, Kaviya S, Anuradha E, Solomon FP. 3D cell culture systems: advantages and applications. Journal of cellular physiology. 2015;230(1):16-26.
31. Zhang D, Pekkanen-Mattila M, Shahsavani M, Falk A, Teixeira AI, Herland A. A 3D Alzheimer's disease culture model and the induction of P21-activated kinase mediated sensing in iPSC derived neurons. Biomaterials. 2014;35(5):1420-8.
32. Fang Y, Eglen RM. Three-dimensional cell cultures in drug discovery and development. Slas discovery: Advancing Life Sciences R&D. 2017;22(5):456-72.
33. McQuade A, Kang YJ, Hasselmann J, Jairaman A, Sotelo A, Coburn M, et al. Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. Nature communications. 2020;11(1):1-17.

Supplementary material

Supplementary table 1. Overview antibodies				
Primary antibodies				
Antibody	**Host**	**Catalogue Number**	**Company**	**Dilution**
OCT4	Goat	ab27985	Abcam	1:1000
NANOG	Rabbit	ab5731	Sigma Aldrich	1:200
SSEA3	Rat	330302	BioLegend	1:200
SSEA4	Mouse	330402	BioLegend	1:200
TRA-1-60	Mouse	330602	BioLegend	1:1000
TRA-1-81	Mouse	330702	BioLegend	1:500
PAX6	Rabbit	ab5790	Abcam	1:200
NESTIN	Mouse	MAB5326	Merck Millipore	1:200
SOX2	Goat	EB07378	Everest Biotech	1:200
MAP2AB	Chicken	CH22103	Neuromics	1:200
TUJ1	Mouse	MAB1637	Sigma Aldrich	1:500
DCX	Goat	ab113435	Abcam	1:200
CTIP2	Rat	ab18465	Abcam	1:200
FOXG1	Rabbit	ab18259	Abcam	1:200
GFAP	Rabbit	ab7260	Abcam	1:200
OXTR	Rabbit	ab217212	Abcam	1:200
TNXB	Rabbit	AB19011	Merck Millipore	1:200
Secundary antibodies				
Antibody	**Host**	**Catalogue Number**	**Company**	**Dilution**
Anti-Goat Alexa 647	Donkey	A-21447	ThermoFisher Scientific, Invitrogen	1:100
Anti-Rabbit Alexa 488	Donkey	A-21206	ThermoFisher Scientific, Invitrogen	1:100
Anti-Rabbit Alexa 594	Donkey	A-21207	ThermoFisher Scientific, Invitrogen	1:100
Anti-Mouse Alexa 488	Donkey	A-21202	ThermoFisher Scientific, Invitrogen	1:100
Anti-Rat Alexa 488	Donkey	A-21208	ThermoFisher Scientific, Invitrogen	1:100
Anti-Rat Alexa 594	Donkey	A-21209	ThermoFisher Scientific, Invitrogen	1:100
Anti-Chicken Alexa 647	Donkey	703-605-155	Jackson ImmunoResearch Europe Ltd	1:100
Overview of the used antibodies in the present study. Displayed are the target of the antibody, the host species, the catalogue number, the company and dilution that was used for each respective antibody.				

Chapter 9

General discussion

In this section, the major aims, key findings, strengths, limitations and future perspectives of the research collected in this thesis will be discussed.

Discussion of major aims and key findings

In this thesis, the importance of epigenetic mechanisms, with a major focus on DNA methylation and hydroxymethylation, in the development and course of sporadic Alzheimer's disease (sAD) were addressed. As emphasized throughout the dissertation, an individual's epigenotype at the level of both aforementioned DNA modifications can be altered as a response towards (*e.g.* adverse) environmental stimuli, leading to changes in gene expression patterns that can affect molecular and cellular processes [1]. These epigenetic changes can occur throughout life and may contribute directly to the development of sAD pathology or interact with the embedded effects of other contributing factors, such as genetic risk factors, to both induce and influence sAD development and progression [2]. In recent years, an increasing number of epigenome-wide association studies (EWAS) have therefore been performed in sAD (reviewed in Chapter 2), where state-of-the-art Illumina microarray technology was applied in combination with more targeted approaches using pyrosequencing on various patient-derived (typically cortical) brain tissues and blood samples to identify disease-associated changes at the level of DNA (hydroxy)methylation [3-12]. As such, these studies have identified distinct and overlapping loci that are dysregulated in sAD patients' brain and/or blood, emphasizing their crucial involvement in the disease. Even though further validation and functional interrogation of these epigenetic signatures remains necessary, the identification of these loci is highly valuable for sAD research, as they provide novel molecular leads that deepen our knowledge on underlying disease mechanisms that subsequently aid in the development of future therapeutic interventions and potential disease biomarker assays.

In spite of the importance of these findings for sAD research, it is well-known that neuroepigenomic studies still come along with numerous challenges. These include the limited genomic coverage of the commonly applied Illumina arrays, the relative low statistical power of EWAS, the (a)specificity of current methodologies, issues related to cell-type specificity of epigenetic marks, variations in target tissue compositions, a lack of understanding in terms of causality and a current scarcity in multi-omics approaches (reviewed in Chapter 2), some of which will also be addressed in more detail in the section 'Strengths and limitations' that can be found below. Altogether, these issues challenge the correct interpretation of current epigenetic findings in relation to sAD, thereby limiting our understanding of their exact functional implications and causal relationships in the etiopathogenesis of AD. It has therefore been debated that

future endeavors aimed at incorporating these considerations into their study designs can substantially increase the reliability and significance of the derived epigenetic data. Hence, the first part of this thesis (Chapters 3 and 4) builds upon this notion and aims at overcoming reoccurring methodological challenges in neuroepigenomic studies, thereby allowing for the production of more conclusive data. Issues that were tackled included the (a)specificity of current (bisulfite) methodologies, as well as the tissue and cell-type specificity of epigenetic marks, which will be discussed in the more detail in the following sections.

It is well-known that classical bisulfite treatment does not allow for the discrimination between DNA methylation and hydroxymethylation, *i.e.* between 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC), respectively [13, 14]. In fact, Illumina bead arrays and other sequencing-based approaches, such as pyrosequencing, which have relied on this technique, have previously acquired cumulative measures on the levels of both modifications rather than each of them individually. For this reason, earlier studies in sAD utilizing this classical bisulfite approach might have overestimated, underestimated, or even overlooked relevant effects related to these epigenetic marks. Thus, in an attempt to overcome this issue, Chapter 3 presented an alternative approach for pyrosequencing based on a highly selective chemical oxidation using KRuO4, which converts 5-hmC into 5-formylcytosine (5-fC) bases [13]. While after classical bisulfite treatment and subsequent polymerase chain reaction (PCR)-based amplification, both 5-mC and 5-hmC are read as a cytosine, 5-fC bases are read as a thymine at the pyrosequencing step, hence allowing one to distinguish between 5-mC and 5-hmC in the context of single cytosine-guanine-phosphate (CpG) sites [13]. Oxidative bisulfite pyrosequencing – the technology described in this chapter – therefore provides a more accurate readout of the levels of 5-mC as compared to classical bisulfite pyrosequencing, while at the same time allowing the user to estimate levels of 5-hmC after quantitative subtraction of the bisulfite and oxidative bisulfite data obtained for the same DNA sample.

In order to demonstrate the efficacy of this modified approach, three spike-in pyrosequencing controls were developed. These can be added to any given DNA sample prior to (oxidative) bisulfite treatment. By utilizing these standards, it was demonstrated that the oxidative bisulfite pyrosequencing approach can be successfully applied for the detection of 5-mC and 5-hmC, as well as unmodified cytosine (5-uC) bases, the latter of which can be ascertained by subtracting the bisulfite signal from 100% [12]. In order to further address the sensitivity of the method, the 5-hmC conversion efficiency was estimated by using the spike-in pyrosequencing controls, revealing that a 5-hmC conversion efficiency >88% per

CpG site could be obtained, which is similar to previous studies using the oxidative bisulfite technology with other sequencing approaches, *e.g.* Sanger sequencing [14]. In addition, factors such as an insufficient bisulfite incubation time were demonstrated to negatively influence the 5-hmC conversion efficiency. Next, it was shown that levels of all three aforementioned modifications could be detected following this approach targeting *OXT* and *DNAJB13* in both post-mortem brain tissue and cultured induced pluripotent stem cells (iPSCs), which are tissues known to be enriched in all three modifications [15-18]. Finally, it was confirmed that the 5-hmC spike-in control could be used as internal pyrosequencing control that does not interfere with the analysis of the accompanying sample, and *vice versa*. Thus, the oxidative bisulfite pyrosequencing protocol presented in this thesis offers a standardized approach for the targeted detection of 5-mC, 5-hmC and 5-uC that could greatly benefit future neuroepigenomic studies, even outside the field of sAD research.

In addition to the (a)specificity of current (bisulfite) methodologies, the second issue that was tackled in the first part of the thesis was the reoccurring challenge of tissue and cell-type specificity of epigenetic marks, obtained from heterogeneous bulk tissues used for epigenetic profiling (see Chapter 4). Aside from inducing noise in epigenetic profiles, the use of bulk tissues, as commonly applied in sAD studies, could mask cell-type specific epigenetic modifications related to the disease, as changes in one cell type could dilute or oppose changes in another, thereby obscuring important cell-specific alterations when analyzed in bulk tissue [19]. By utilizing a combination of laser capture microdissection (LCM) and limiting dilution bisulfite pyrosequencing (LDBSP) [20], Chapter 4 described a novel approach for targeted methylation analysis of single alleles derived from small pools of individually isolated neurons. As such, this approach offers a workable solution for the aforementioned issue, as it allows for targeted methylation profiling in a limited number of isolated brain cells.

For this purpose, dorsal raphe nucleus (DRN) tissue sections were stained for a neuronal subtype-specific marker (*i.e.* serotonin (5-HT)), after which immuno-positive neurons were individually isolated using LCM and subsequently processed in pools of 50 cells for LDBSP analysis, targeting *DNAJB13*, *PGLYRP1*, *OXT*, *RHBDF2*, *C3* and *LMX1B* simultaneously. In contrast with previous studies [20-25], LDBSP on pools of 50 neurons rendered a considerable amount (8.77% on average per gene) of the downstream PCR reactions with more than one target allele. In order to correct the LDBSP data for these multi-allele reactions, a novel data analysis approach was then developed. This novel approach, which compromises a CpG-site calling procedure combined with an integrated in-depth analysis of the raw CpG methylation values, aims at avoiding

unintentionally induced bias due to the blunt exclusion of reactions that seem to contain more than one target allele. When comparing this novel approach to the traditional method for the downstream data analysis, which is based on excluding multi-allele reactions, significant differences in the combined CpG methylation rates for *LMX1B*, as well as for individual CpG sites in *LMX1B*, *RHBDF2* and *OXT*, were identified. These findings emphasized that the derived LDBSP methylation data can differ significantly depending on the method that is applied for the downstream analysis. This implies that it is crucial to correct for multi-allele reactions instead of excluding them when encountered utilizing LDBSP, as this could potentially influence experimental data negatively, either by inducing bias or by reducing or reinforcing effect sizes, for example. Overall, the approach proposed here provides the user with a more accurate estimation of the DNA methylation status of each target gene in the analyzed cell pools, thereby adding further validity to the data. In conclusion, LDBSP in combination with LCM offers a novel and alternative strategy to single cell(type) bisulfite sequencing techniques that can be applied for the study of DNA methylation marks in the human brain. It is anticipated that similar strategies using a limited number of isolated cells in combination with LDBSP will be increasingly valuable for future neuroepigenomic studies, also outside the scope of sAD.

Another major focus in the first part of this thesis (see Chapter 5) was centered on the importance and (unmet) need for profiling epigenetic signatures in early affected regions of the sAD brain, including regions of the brainstem such as the DRN and locus coeruleus (LC), which has not been performed previously. When looking at the bulk of EWAS that has been conducted in sAD to date (see Chapter 2), one would realize that these studies have primarily assessed epigenetic profiles in either cortical regions or the hippocampus. Even though one cannot deny (later) involvement of limbic and cortical areas of the brain in relation to the cognitive symptoms, recent evidence suggest that early dysregulation in both the aforementioned brainstem nuclei accounts for the earliest, largely non-cognitive symptomatology [26-28]. Furthermore, studies have shown that tauopathy may begin earlier than previously thought and possibly in the brainstem rather than in transentorhinal regions [26-29], suggesting a potential causal role in the pathogenesis of sAD. In line with this view, it is currently thought that from these brainstem nuclei pathology could spread to subcortical areas, including the hippocampus, and, subsequently, to areas of the neocortex, marking the clinical stages of the disease. Other evidence for a central role of the brainstem in sAD comes from magnetic resonance imaging studies, demonstrating significant volume reductions and structural deformations in these brainstem regions of sAD patients [30, 31]. Thus, interrogating the epigenetic landscape in early affected regions such as the DRN and LC could potentially improve our understanding on

the role of epigenetic mechanisms involved in the early stages, as well as during the course of the disease.

Chapter 5 therefore presented the first large-scale epigenetic study in the DRN and LC, consisting of two EWAS combined with a similar oxidative bisulfite approach as described in Chapter 3. Furthermore, these discovery EWAS were complemented with a targeted pyrosequencing validation study in both brainstem regions using the same patient cohort, as well as a cell-subtype specific validation analysis in the DRN using an independent cohort (relying on the combinatory LCM-LDBSP approach described in Chapter 4). During the EWAS, Braak stage-associated alterations at the level of 5-mC, 5-hmC and 5-uC were quantified in bulk tissues derived from both the DRN and the LC by using the Illumina Infinium MethylationEPIC BeadChip array. A strong overlapping dysregulation in the Tenascin XB (*TNXB*) gene was identified in both brainstem regions assessed, next to previously identified and novel epigenetic loci, which might play a pivotal role in the early development and course of sAD. In the subsequent validation study using bisulfite pyrosequencing in the same patient cohort, it was confirmed that demethylation of *TNXB* was associated with increasing Braak stage in both the DRN and LC. These findings were then validated in an independent patient cohort targeting *TNXB* in single serotonergic and non-serotonergic cells isolated from the DRN by using the LCM-LDBSP approach and an adapted bisulfite pyrosequencing protocol for each of the cell types, respectively. When comparing the bisulfite methylation levels of *TNXB* between sAD patients and controls in serotonergic and non-serotonergic cells, a significant interaction between cell-type and experimental condition was identified. Strikingly, sAD-associated methylation profiles in *TNXB* were opposite in serotonergic neurons and non-serotonergic cells, the latter of which resembled the initially acquired EWAS and pyrosequencing data.

Taken together, these results strongly advocate for an important role of epigenetic dysregulation in *TNXB* in the (etio)pathophysiology of the disease. Interestingly, epigenetic dysregulation in *TNXB* has previously also been annotated in another sAD EWAS [5], next to genetic variation in this locus that has been associated with a risk of developing the disease [32-34], as well as with another tau-related disorder [35]. *TNXB* expresses a glycoprotein with anti-adhesive properties, but its exact physiological role in the brain is not completely understood [36]. As such, more research on the exact function of *TNXB*, the interplay between genetic and epigenetic variation in this locus, as well as further functional interrogation of this sAD neuropathology-associated epigenetic signature, remains necessary in order to develop a better understanding of its exact role in sAD. Another major outcome of this study was that the associated

dysregulation in *TNXB* in the DRN was both dependent on the disease phenotype and the cell type analyzed, which supports the notion that epigenetic data derived from heterogeneous post-mortem bulk tissue should be interpreted with caution, as changes in one cell type could negate or mask changes in another. Overall, this finding has crucial implications for future planned epigenetic studies in AD, as it warrants the need for single cell (-type) neuroepigenetic analyses, opposite to the more common bulk tissue analyses that have been performed to date. Future efforts aimed at dissecting intercellular differences of epigenetic marks are therefore anticipated to aid in developing a more complete understanding of their role in the disease and will be increasingly important for neuroepigenomic studies into sAD.

While still dedicated to the pathophysiology of sAD, the second part of the thesis (Chapters 6, 7 and 8) deviates from profiling epigenetic signatures in patient-derived brain tissues or cells and addresses the potential of iPSC-based models for studying the disease. The capacity to differentiate iPSCs into disease-relevant neural subtypes has empowered them as an exciting tool for the generation of patient-specific 'brain in a dish' models that could fill the gap between pre-clinical and clinical sAD research and may, eventually, largely replace experimental animal studies [37-39]. The second part of the thesis therefore commences with a review paper presented in Chapter 6, summarizing the pioneering studies that have explored the use of these stem cells for sAD modelling. Recent findings have demonstrated that cells harboring patient-specific backgrounds do manifest pathological phenotypes *in vitro*, such as increased levels of amyloid beta peptides and hyperphosphorylation of tau protein, amongst others [40-44]. For this reason, patient-derived iPSC models are thought to be extremely valuable for fundamental research into sAD, as they allow one to interrogate functional effects of genetic, epigenetic and transcriptional variants linked to risk, as well as to protective, environmental factors. Within the field of neuroepigenomics, iPSC-based models in combination with epigenetic editing systems might even offer appealing avenues to dissect causality of epigenetic signatures in view of disease pathology, and *vice versa* [19]. Such endeavors might significantly contribute to elaborate our understanding on the mechanisms underlying the disease and could aid in the development of therapeutic interventions.

In spite of their potential, these aforementioned studies have also proven that a high degree of variability in terms of detectable disease hallmarks is common when comparing differentiated neural cells from different patient-derived iPSC lines. While some of this might be explained by the heterogeneous nature of sAD, it is likely that other sources of variability related to the models play a role in this as well. Given that iPSC models in sAD represent a relative nascent area of

research, it is therefore crucial to dissect each of these sources of variability and to assess how these could impact the interpretation of the findings in relation to sAD. Examples of such sources (that were also reviewed in Chapter 6) include genetic alterations induced in response to iPSC reprogramming, cell type origin-dependent neural differentiation bias, loss of potential aging traits in iPSC lines due to reprogramming processes of the epigenome and a lack of consensus in neural differentiation protocols between different studies. In addition, much work is still needed in order to define and establish adequate reference panels for control cell lines, as well as to assess the impact of somatic cell mosaicism that can affect disease modelling, and clonal selectivity of donor cells for iPSC reprogramming.

Expanding on the potential of patient-specific *in vitro* disease models, Chapter 7 follows with a thorough review on directed and direct neural differentiation protocols using iPSCs and other somatic cells. The progress that has been made to date in generating human and mouse anatomically-specified, *i.e.* attributable to specific regions of the brain, neural precursor cells (NPCs) and differentiated neuronal subtypes are being summarized, including glutamatergic-, GABAergic, dopaminergic, serotonergic- and cholinergic- neurons. Overall, over the last years, increasing knowledge on developmental signaling pathways and gene regulatory networks has guided the design of current *in vitro* neuronal differentiation and cellular reprogramming strategies. Moreover, recent studies have demonstrated that within these networks of patterning molecules and transcription factors that orchestrate neuronal induction and differentiation, the epigenetic machinery is essential for fine-tuning genetic programs that coordinate distinct developmental processes, as well as shaping neuronal identities at a phenotypic resolution [45, 46]. For this reason, different kinds of methods relying on chemically defined systems using neuralizing patterning factors, transcription factor-mediated reprogramming strategies and epigenetic-based approaches for the derivation of the aforementioned subtypes are currently being developed. Overall, patient-derived neurons obtained via directed differentiation of iPSCs or direct reprogramming of somatic cells promise to fill an important niche between studies in humans and animal models in deciphering disease mechanisms and identifying therapeutic avenues for neurodegenerative- and other psychiatric diseases.

Although recent scientific efforts begin to tap into this potential, pioneering studies have also highlighted challenges in directing neuronal differentiation that must be overcome for it to be fully realized. In fact, *in vitro* neuronal differentiation is actually not a process that is fully disciplined yet, given the great variability in protocol efficiencies and the production of unwanted cell types in the obtained

neural cultures that are observed both within and between laboratories. Furthermore, differentiation protocols for multiple neuronal subtypes remain unestablished, especially when considering very specific anatomical subtypes, or they remain unstable and do not allow one to obtain the terminally differentiated neuron with its entire functional mature characteristics. As such, currently, emphasis is payed toward efforts to increase the efficiency of current protocols, including combinatory approaches and the use of cell sorting techniques, among others (reviewed in Chapter 7). Taken together, the establishment of robust protocols to obtain functional neurons *in vitro* is essential for studying human brain cell functions, as well as disease modelling, drug discovery and regenerative medicine, both within and outside the scope of sAD research. Future efforts aimed at overcoming these aforementioned limitations are therefore anticipated to aid in realizing the full potential of this platform.

Finally, in Chapter 8, a pilot study is presented that combines efforts described in previous sections and aims at characterizing a cortical forebrain differentiation protocol from iPSCs, with the ultimate goal to apply this approach for the establishment of a sAD-related *in vitro* model. For this purpose, the iPSC-line used for neuronal differentiation was first characterized for the expression of pluripotency markers such as OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 and TRA-1-81, in order to confirm their stemness. Next, the cells were differentiated towards NPCs based on the formation of free-floating embryoid bodies (EBs) combined with the inhibition of dual SMAD signaling [47], which favors development into the neural lineage. These obtained NPCs were then phenotypically characterized by means of immunofluorescence markers for e.g. NESTIN, PAX6 and SOX2. Double labeling with each of these markers revealed that most of the iPSCs differentiated into NPCs (>82%). These cells were then further differentiated towards a cortical forebrain fate by maintaining them in medium supplemented with both neural patterning- and growth factors. Following this 8-week differentiation protocol, the expression of neuronal markers, including MAP2AB, DCX and TUJ1, and the glial marker GFAP, was confirmed in the cultures. Moreover, the cells lost the expression of pluripotency markers such as OCT4 and typically expressed cortical forebrain markers such as CTIP2 and FOXG1. Furthermore, as TNXB and oxytocin (OXT)-signaling have previously been associated with sAD based on EWAS screenings, expression of both *TNXB* and the oxytocin receptor (*OXTR)* was confirmed in the differentiated neural cells by a immunofluorescence staining, suggesting that the cultures might represent an appealing model system to study processes related to both genes in the context of sAD. Overall, these data showed that the applied protocol successfully generated neurons, as well as glial cells, from these iPSCs, with a cortical

forebrain identity that will enable future mechanistic investigation of candidate genes for sAD.

As a next step in the characterization, it would be interesting to explore the (ratios of) neurotransmitter subtypes that are present in the differentiated cultures, such as glutamatergic and GABAergic neurons, and to assess whether these cells also display the functional characteristics of their *in vivo* equivalents, such as their electrophysiological properties. These measures and the ability to replicate the presented findings in different batches and other iPSC lines will be crucial to address the robustness of the differentiation protocol. In addition, to further address the potential for these cultures for the use as a disease model, it would be highly informative to explore the responses of the neural cell towards sAD-related environmental stimuli, including amyloid beta, and assess whether potential toxicological responses might differ in these cultures when comparing healthy control and patient-derived neural cells differentiated from iPSCs. Within this context, it would furthermore be interesting to interrogate whether genes such as *TNXB* and *OXT* or related signaling pathways could mediate susceptibility or resilience towards such exposures. In addition, by using alternative differentiation protocols, or by adapting the one proposed in this chapter by including additional patterning factors, allowing for the derivation of serotonergic- or noradrenergic neurons that are found in the DRN and LC, respectively, it would be highly fascinating to assess whether similar or different cellular and molecular responses can be identified in different cell types. In conclusion, the differentiation method proposed in this chapter provides an important stepping-stone for future functional studies using iPSC-derived brain cells for sAD research.

Strengths and limitations

The research collected in the first part of this thesis aims at interrogating epigenetic signatures in the (etio)pathophysiology of sAD, with a major focus on identifying dysregulation in early affected brainstem regions that might reflect changes associated with the most incipient stages of the disease. Aside from gaining new insights into underlying disease mechanisms of sAD, a major strength of the projects presented here is the establishment and application of state-of-the-art methodologies. To our knowledge, the method presented in Chapter 3 represents the first fully optimized approach combining pyrosequencing with oxidative bisulfite technology, allowing one to distinguish between ‘true’ DNA methylation and DNA hydroxymethylation, *i.e.* 5-mC and 5-hmC, respectively, in a locus-specific manner. Moreover, the novel developed DNA spike-in controls that can be applied in conjunction with the described protocol enables to confirm the successful conversion of the oxidative bisulfite-converted DNA, hence ensuring methodological consistency of the described

procedure. This approach in combination with these novel spike-in controls therefore provides future opportunities for more standardized, thorough and affordable interrogation of both DNA modifications using pyrosequencing in a wide variety of pathological conditions, including sAD. Before the introduction of this oxidative bisulfite conversion step, only the combined signal of methylation and hydroxymethylation could be obtained. As repeatedly addressed in this thesis, this has been a major issue in neuroepigenomic studies to date, especially when applied on brain tissue, where DNA hydroxymethylation is known to be more abundant as compared to other bodily tissues. It should therefore be emphasized that previous (brain-related) studies relying on the classical bisulfite approach have been unable to discriminate between the two and, hence, should be interpreted with caution.

In relation to the aforementioned strength, Chapter 5 describes the first study to date that targets DRN and LC bulk tissues derived from the brainstem of sAD patients using the Illumina Infinium MethylationEPIC BeadChip array in combination with an oxidative bisulfite treatment of the isolated target DNA from these samples. Aside from representing one of the pioneering studies applying the oxidative bisulfite approach for genome-wide profiling in sAD, the use of the novel EPIC array technology furthermore provides a higher genomic coverage when compared to previous studies that have commonly relied on its predecessor, *i.e.* the Illumina BeadChip 450K array. In fact, the EPIC array technology has nearly doubled the number of target CpG sites across the genome (*i.e.* from 485.512 to 866.836 probes [48]), now also including previously untargeted enhancer regions, hence allowing for the interrogation of additional sites in relation to the disease. In addition to the value of dealing with one of the most devastating disorders world-wide, the effort to interrogate epigenetic signatures in previously overlooked brainstem structures, *i.e.* the DRN and LC, makes the presented study stand out from similar endeavors targeting other brain regions that are affected in more advanced stages of sAD, which is a major asset of this thesis. To this end, overlapping and unique epigenomic profiles associated with disease progression described by Braak stage were identified, which have set the pillars for future replication studies and meta-analyses, as well as for mechanistic and functional research into their role in the disease.

Another significant strength from the first part of this thesis is the approach described in Chapter 4 and applied in Chapter 5, which relies on a combination of LCM and LDBSP and allows for targeted (bisulfite) methylation profiling of single alleles derived from a limited number of isolated neurons. Initially, the input requirements for targeted approaches such as pyrosequencing, as well as for genome-wide methods such as arrays, required hundreds of thousands or even

millions of cells. Owing to technological developments as shown in Chapter 4, this is changing with numerous epigenetic features now assayable at the single-cell level. Understanding of DNA methylation signatures in the past was therefore mostly based on studies that utilized these techniques for quantifying the average epigenetic states of bulk cell populations and tissues, similar to what was performed in Chapter 5 where the EPIC array was applied on heterogeneous brainstem tissue homogenates. Even though these studies using bulk tissues are still highly informative and have revealed meaningful disease-associated signatures, such studies generally come along with a small signal-to-noise ratio, while the accompanying measurements that represent average tissue profiles cannot always be extrapolated to single cell populations or certain cell type niches within or between different brain regions, *e.g.* cells that display disease-related pathology compared to those that do not. Different cell types are known to differ in their epigenetic profiles, independent of disease-related changes, and manifestation of the latter could easily occur in a cell-dependent manner, as demonstrated in Chapter 5, or only under very specific circumstances, such as within specific time intervals or only upon exposure to certain environmental cues. Evidently, for many biological questions, interrogation of epigenetic marks is most informative when studied at a single-cell level, where intercellular differences can be dissected leading to a more refined understanding of their contribution to the disease. All in all, such cell-specific approaches are therefore of immense added value and are anticipated to be increasingly considered in future neuroepigenomic studies, both within and outside the scope of sAD.

The second part of the thesis is focused on addressing the potential of iPSC-based models in sAD. A clear strength of this platform, which is also addressed in Chapters 6, 7 and 8, lies in the fact that patient-derived cells contain a pathogenic background and, hence, it is anticipated that these cells manifest pathological processes *in vitro*, thereby offering a promising avenue for sAD modeling. In fact, the use of disease-relevant neural cells by differentiating iPSCs along the neural lineage, as demonstrated in Chapters 7 and 8, offers an alternative approach to study the underlying neuropathological mechanisms in a humanized, personalized, and cell subtype-specific manner. As such, iPSC-derived neuronal populations generated from sAD patients with known pathogenic or susceptible backgrounds can be studied, (epi)genetically probed, exposed to environmental factors and treated with drug libraries to investigate their molecular and cellular effects. For these reasons, there has been a growing body of research over the past years to adopt rapidly improving iPSC-derived model systems of sAD for fundamental research applications, as well as for the assessment of drugs prior to the initiation of clinical trials [49-53]. Another clear advantage of these model systems is their potential to reduce the need for animal

experimentation and hopefully increase the translatability of preclinical studies towards clinical interventions. Although still a relative nascent area of investigation, it is anticipated that future efforts in this field will broaden our knowledge on the underlying pathophysiology of many brain disorders, including sAD.

In spite of the aforementioned strengths, the work described in this thesis is also subject to both general and more specific limitations. While the oxidative bisulfite technology used enables the detection of 'true' DNA methylation and hydroxymethylation, it should be recognized that this approach does still not take into consideration other (less prevalent, but functionally relevant) CpG modifications such as 5-fC and 5-carboxylcytosine (5-caC). In reality, the signals of both 5-fC and 5-caC are entwined with the signal ascertained for 5-uC, which is currently considered inversely proportional to the bisulfite read-out and can be obtained by quantitative subtraction of the bisulfite signal from 100%, as demonstrated in Chapter 3. In fact, a bisulfite treatment of 5-uC, 5-fC and 5-caC bases results in a conversion to uracil and, subsequently, to thymine after the PCR amplification step, which makes it impossible to distinguish between the three modifications. As the abundancy of 5-fC and 5-caC in the brain is known to be relatively low [18], even when compared to 5-hmC, it is not expected that they will have a profound impact on the obtained data. This could therefore be regarded as a minor limitation, even though still important to consider given that it affects both the pyrosequencing approach described in Chapter 3 and the array data in Chapter 5.

While both of the previous chapters allow for the detection of 5-mC and 5-hmC, the method described in Chapter 4 relies on the combination of LCM and LDBSP and hence still provides the unsegregated signals, as oxidative bisulfite treatment had not been implemented yet. In fact, it is important to acknowledge that for LDBSP, which relies on the analysis of single alleles, it is impossible to obtain a 5-hmC measurement based on this technology. Even when implementing the oxidative module, one can not to run both the bisulfite and oxidative bisulfite approach on a single DNA molecule in parallel, which would be necessary to derive a readout for 5-hmC. It would therefore require other methodological advances and alternative strategies (such as for example proposed in [54]) in order to obtain a direct 5-hmC read-out from a single DNA molecule, opposite to the indirect method that is used for the oxidative bisulfite technology. Nonetheless, future implementation of the oxidative module for LDBSP could still allow for the derivation of the 5-mC signal only, thereby providing a more detailed profile of 'true' DNA methylation patterns on a single allele-level.

Another limitation stemming from the indirect method to determine 5-hmC and 5-uC levels using the oxidative bisulfite technology is related to the 'naïve' subtraction of the oxidative- and the bisulfite signals that is described in Chapter 3. Due to technological variation and the need for paired measurements derived from the same DNA target sample, the oxidative bisulfite method is subject to noise that can affect the correct readout for (relative small levels of) 5-hmC. This issue is further exemplified by two kinds of overshooting values when estimating the 5-hmC and 5-uC levels following quantitative subtraction. As demonstrated in Chapter 3, subtraction of the oxidative bisulfite signal from the combined bisulfite signal for estimating 5-hmC can result in the derivation of negative values, which are biologically meaningless and represent false differences in scores between the paired bisulfite-only and the oxidative bisulfite data sets, due to background noise. On the other hand, when combining the levels of 5-uC, 5-mC and 5-hmC after quantitative subtraction, levels can also exceed 100% due to similar experimental noise. These overshooting sites may even constitute a substantial proportion of total CpG sites, depending on the technical variation and the degree of 5-hmC content in the analyzed samples. In order to exclude negative values, strict detection thresholds have commonly been applied, leading into a substantial dropout of CpG sites from related (Illumina array) data sets. Alternatively, novel methods such as a maximum likelihood estimation (MLE) can nowadays be used to determine the levels of these modifications when dealing with oxidative bisulfite-based 5-mC, 5-hmC and 5-uC data [55, 56], as also demonstrated in Chapter 5. This approach is consistent in that 5-hmC levels are non-negative and all three modifications together never sum over 100%, consequently leading to less CpG dropouts from the data set. Thus, this alternative therefore provides a more suitable approach for estimating the levels of these modifications, even though still subject to experimental variation and noise that would require further methodological developments to be overcome.

Another limitation associated with data analysis of Illumina arrays is related to the identification of differentially modified regions, in which DNA methylation and hydroxymethylation are treated equally. However, next to their distinct biological roles, these two modifications are not equally abundant in the brain. In fact, DNA hydroxymethylation is not as enriched as DNA methylation, and, as such, by using identical data analysis strategies for the identification of differentially hydroxymethylated regions (DHRs) and differentially methylated regions (DMRs), current approaches tend to contempt the former. This is demonstrated from the analysis performed in Chapter 5, where neither for the DRN nor for the LC a DHR associated with Braak stage was identified. These findings therefore suggest that the spatial correlation analysis for DHRs might need to be optimized, *e.g.* by adjusting the sliding window, if one wishes to identify meaningful altered regions

at the level of 5-hmC. Similarly, next to treating the different modifications equally, current methods also asses each of these modifications independently, while alterations at a locus at the level of multiple modifications can also have functional consequences. Future methods should therefore also be aimed at integrating changes within specific sliding windows at the level of multiple modifications, rather than looking at each of these modifications individually.

Two more specific minor limitations faced in Chapter 4 are related to the (novel) data analysis method that is used for LDBSP on pools of neurons isolated with LCM. While the novel approach presented here reduces data bias by omitting the need for excluding multi-allele reactions (as conducted in the traditional analysis pipeline), it is limited in detecting only a maximum of three DNA molecules present in a single reaction. In fact, it extrapolates the number of target alleles present based on arbitrary thresholds that were set for the raw CpG methylation values obtained after pyrosequencing. Within the boundaries of these thresholds, *i.e.* (1) ≤8.33% and ≥91.33%, (2) 50±8.33%, and (3) 33.33±8.33% and 66.66±8.33%, one can make a relative safe estimation on whether a reaction contains one, two or three alleles, respectively. However, with increasing numbers, the estimation based on CpG methylation values becomes rather ambiguous, as one also needs to consider potential technical variation that might affect these methylation values. The method proposed in this chapter will most likely classify these reactions as representing three alleles. Despite the fact that the methylation data obtained from these reactions would now be partially corrected by this novel data analysis pipeline (when compared to the traditional pipeline), this correction is still imperfect in such a scenario. Hence, overall, it is strongly advised to dilute the DNA sample properly in the first place to ensure that the proportion of reactions that contain more than three alleles is either negligible or, preferable, zero. In addition to the aforementioned notion and independent of the performed downstream analysis, LDBSP does also not allow one to identify multi-allele reactions that show an identical pattern of CpG methylation. In this specific scenario, and irrespective of whether a reaction contains one, two, or three or more alleles, all CpG methylation values will approach 0% or 100%, for unmethylated and methylated sites, respectively. These reactions therefore display a typical binary methylation pattern that is expected for a single DNA molecule, while in reality more alleles were present. In general, this becomes a bigger challenge when, due to biological factors, a target locus is completely methylated or unmethylated. It is therefore advisable to always target a substantial number of CpG sites per target gene (5 or more) in order to increase the chance of obtaining, and hence detecting, at least one site with a ‘non-binary’ methylation status on one of the target alleles.

Several other limitations that should be addressed regarding the first part of this thesis are related to the brainstem EWAS described in Chapter 5, of which the most common ones are also summarized in Chapter 2. A recurring limitation of these studies is the small to moderate sample sizes that are being used, which limits the statistical power to detect significant differences. This poses a problem particularly when aiming to identify differentially modified positions at the level of 5-uC, 5-mC and 5-hmC, which generally show very small effect sizes. In addition, the significance thresholds for assigning these positions after multiple testing correction is also extremely conservative, as a $p < 9.0E-08$ has been established for achieving genome-wide significance when utilizing the Illumina Infinium MethylationEPIC BeadChip array [57]. This is exemplified in the EWAS performed in Chapter 5, where all the top-ranked differentially modified positions for both brainstem regions were nominally significant at $p < 0.001$, but none of them reached the genome-wide significance level after multiple testing correction. As the collection of bigger sample sizes remains an ongoing challenge, it would therefore require further development of alternative bioinformatics approaches to strengthen these results. In addition, it will be imperative to conduct replication studies in independent patient cohorts and perform meta-analyses to confirm the obtained findings. Recently, the first cross-cortex meta-analysis in sAD using data from 1,408 donors from six independent studies identified 220 CpG sites associated with neuropathology, annotated to 121 genes, of which 96 genes had not been previously reported at experiment-wide significance, which further emphasizes the importance of these type of analyses [58].

While the Illumina Infinium MethylationEPIC BeadChip provides a higher coverage compared to its predecessor, the probes present on the chip still only cover about 3% of the total number of CpG sites that are found within the human genome. Moreover, non-CpG sites, including CpA, CpT and CpC, as well as non-cytosine modified positions are not covered by this technology. Future epigenetic studies in sAD should therefore also be complemented with whole-genome-sequencing based approaches, which could provide a broader coverage, and thus would allow for a deeper understanding of disease-related signatures. Nevertheless, these might oppose a restriction given their high costs, *i.e.* per individual and in view of the concomitant demand for larger study populations (due to the accompanying decrease in statistical power), making currently used array-based approaches still a more preferred strategy in most cases.

As also addressed in Chapters 4 and 5, as well as strongly emphasized throughout the thesis, another major issue of EWAS is related to the cellular heterogeneity of bulk cell populations that can produce noise and bias in

experimental data when used for epigenetic profiling. While single cell analyses evidently provides a more robust alternative, another way of dealing with this issue is the application of cell deconvolution approaches for downstream bioinformatics analyses [59]. However, to date, these algorithms for the human brain are only available for the frontal cortex [12, 60], while in the EWAS of this thesis brainstem nuclei were targeted. These therefore render limited advantage for such an analysis, as cortical brain regions differ significantly in their cellular proportions when compared to the DRN and the LC. As such, during the analysis performed in Chapter 5, surrogate variables that are known to correct for unmeasured confounding factors, including cell type composition, were included in the statistical model. While currently offering a valid approach, it would require further development of bioinformatic tools and alternative algorithms that would allow one to correct for the cell type composition in different structures of the human brain, especially for more specialized nuclei such as the DRN and the LC.

While the major focus of the epigenetic studies presented in this thesis is on identifying disease-related changes at the level of DNA modifications, it is also important to consider their biological relevance and coherent limitations. In fact, transcriptional regulation is not only mediated by DNA modifications, but also by histone modifications and non-coding RNAs, as well as by underlying genetic factors [61]. Hence, significant changes at a single CpG site or even in a bigger region of a gene might not directly result in (negative) functional alterations. In addition, the tendency to reduce complex disorders such as sAD to a single genetic locus opposes another major issue, as it shifts the attention towards more specific and single molecular or cellular systems, hence one might lose sight on the importance of 'the bigger picture'. Of course, by reducing a complex disorder like sAD into smaller elements, one hopes that when such elements become elucidated, it will be easier to tie them together and understand the disease in a more holistic fashion as well. However, we should not deny that for complex disorders such as sAD, an immense interacting network of numerous distinct (epi)genetic risk-, disease-modifying- and protective loci, and associated molecular pathways and cellular mechanisms, are likely to underlie the development and course of the disease. As such, one should be cautious and careful in claiming that single genes as those identified by EWAS are explaining the disease and/or related endophenotype(s). In fact, in future endeavors, it would be more meaningful to integrate current available epigenomic data sets into more multi-omics- and even multi system-levels-orientated approaches, using both genomics, epigenomics, transcriptomics and proteomics as well as other (*e.g.* imaging, clinical phenotyping) data obtained from the same patient cohorts. Overall, this might provide clarity on the contribution of each of these omics layers and might aid in disentangling the complex pathophysiology of sAD.

In contrast to data obtained in GWAS, there is the issue of causality when dealing with results obtained through EWAS. In EWAS, it is simply impossible to conclude whether epigenetic changes associated with sAD are causally involved, whether they are a consequence of its pathogenesis or whether they represent an epiphenomenon [19]. In view of sAD, it is likely a combination of both, where some epigenetic signatures identified in EWAS could represent early causal changes and others changes that are the result of pathological processes involving *e.g.* amyloid beta or neurofibrillary tangle formation. In order to elucidate their exact contribution, it requires future longitudinal studies and approaches combining epigenetic editing systems in both *in vivo* animal models and *in vitro* culture systems, using iPSC-derived neural cells, for example.

Aside from the challenges faced in neuroepigenomic studies on sAD, the second part of the thesis, which addresses to potential of iPSC-based model systems for sAD, is also subject to several limitations. While iPSC-derived neural cells are portrayed as breakthrough models for sAD and many other brain disorders, one should not overlook the fact that this area of investigation is still very nascent and, hence, highly experimental in nature. At the moment, it therefore comes with potentially even more limitations than direct translational features, which all need to be addressed in future studies if one wishes to develop robust model systems that could directly benefit preclinical and clinical research and, eventually, healthcare. The field currently has the tendency to go ahead of itself and directly starts looking for opportunities to use these patients' iPSC-derived neural cultures for interrogating pathological mechanisms related to the disease of interest, whilst the associated models themselves still require further validation in order to do this efficiently and correctly. In fact, given their juvenile and artificial character, it will be paramount to dissect and subsequently optimize each of the contributing factors that could affect establishment of a valid *in vitro* disease model for sAD. In this context, one could think of the molecular and cellular consequences of reprogramming processes, effects of varying differentiation protocols, the use of different donor cell types, the role of different culture conditions, *e.g.* 2D or 3D models, and other potential environmental influences that have been addressed above. All of these factors might affect the derivation of clinically relevant mature/aged brains cells, as well as the manifestation of disease hallmarks *in vitro*, hence influencing potential experimental outcomes when not studied in a controlled setting. Many of these general limitations or those specifically related to the establishment of sAD models are also reviewed in Chapters 6 and 7, and will require attention in future studies to strengthen the potential of this platform.

Finally, one of the most important considerations related to iPSC-models is that the acceptable quality control criteria attributed to iPSCs and their differentiated neural progeny should be defined in more detail for both fundamental and clinical applications, as this aspect has not been extensively clarified to date. Evidently, this is also difficult to address, especially for sAD patient-derived cells, because of our limited knowledge on the molecular and cellular mechanisms underlying this complex and heterogeneous disease. After all, the aim is to produce human brain cells that mimic their *in vivo* counterparts, both in healthy and diseased conditions. However, the similarity and differences between these cells that are produced under laboratory conditions and those obtained from a human brain have not been extensively characterized yet. Thus, efforts aimed at the development of a consistent and reliable translational iPSC model with well-defined cellular and molecular characteristics is therefore a vital pre-requisite to realize and understand their full potential. Well-defined standardized reprogramming, direct differentiation and directed differentiation protocols and strategies to gauge the magnitude, as well as to improve the maturation status of the cells will be highly valuable in this respect. As such, the study presented in Chapter 8 is also subject to these aforementioned limitations. While the successful differentiation into a specific neural lineage was determined based on the expression of lineage-specific transcription factors, further functional tests and a more detailed molecular characterization remains necessary. All in all, such more detailed analyses are needed if one wishes to conclude that these derived cultures are similar to cells found in the human brain, and hence could form the basis for the establishment of a reliable sAD model system.

Future perspectives

The key findings, strengths and limitations that were addressed above provide exciting opportunities for future sAD-related studies. On the one hand, the identified dysregulated epigenetic loci and associated molecular pathways in the sAD brainstem, the presented state-of-the-art methodologies for the quantification of (specific) epigenetic marks in bulk tissues or single cells, as well as the iPSC models, provide a solid foundation for future mechanistic studies on the etiopathophysiology of the disease. On the other hand, the limitations that have been described might by themselves lead to new research ideas and the development of improved methodologies by inspiring other researchers to tackle these in future studies. Overall, the knowledge that is gained from the work collected in this thesis could provide guidance for the development of new scientific breakthroughs and inventions within the field of neuroepigenomics in sAD. While moving this field forward, it remains crucial to put both the research findings and their corresponding technological-, bioinformatics- and logistical limitations into perspective in order to assign the right credibility to the data that

are being obtained from these studies. Future efforts should therefore not only be focused on acquiring more elaborated epigenetic data sets in sAD, but should also be aimed at overcoming the remaining challenges in order to strengthen and further elucidate the exact contribution, as well as significance, of the obtained epigenetic signatures associated with this disease.

The continuous development of (genome-wide) approaches for interrogating the epigenome together with an increasing number of initiatives aimed at establishing high-quality patient tissue banks could greatly benefit future endeavors. In fact, the availability of brain tissue linked to extensive records on patient demographics would allow for larger epigenetic screenings with increasing sample sizes. Furthermore, they could provide a great source for independent replication studies. Such efforts, as well as the use of meta-analyses, will be highly beneficial for the field to confirm previous observations. Moreover, as the use of bulk tissues might dilute disease-associated effects due to tissue heterogeneity, it will be crucial to implement more single-cell(type) based approaches in future studies, allowing one to dissect cell-specific contributions of epigenetic aberrations in the sAD brain. Furthermore, the field would highly benefit from more sequencing-based techniques that provide a higher genomic coverage, which could be combined with methods for the segregation of various epigenetic marks, such as in the case of DNA methylation and hydroxymethylation by means of oxidative bisulfite sequencing. Aside from targeting brain tissues or cells, it would furthermore be highly informative to include more accessible tissues into the study designs, such as blood samples from the same patients, as this would allow for addressing potential overlap and differences in disease-associated signatures that could allow for the establishment of biomarkers and diagnostic tools. In this respect, and considering the dynamic nature of epigenetic alterations, it will be highly informative to perform well-established longitudinal studies that aid in developing a better understanding on the temporal association of these epigenetic marks with the disease.

Importantly, it shall not be overlooked that methodological advances should be accompanied by the development of robust bioinformatics tools to handle epigenetic data sets. In fact, advances in statistical approaches and mathematical modelling serve as the backbone for taking full advantage of such data. Furthermore, emphasis should be put on future efforts to integrate epigenomic data into bigger multi-omics and multi-level (systems biology/medicine) data sets, including genomics, transcriptomics, proteomics, and demographic and phenotypical data, among others, in order to dissect the functional consequences of changes in epigenetic marks that would be necessary to address the significance of these alterations in the disease causation and progression. In fact,

it is anticipated that more advanced bioinformatics tools such as machine learning, computational modeling and systems biology approaches will be highly valuable in the attempt to differentiate the molecular cascades that induce or drive sAD from those that merely represent a consequence of it or just an epiphenomenon, *e.g.* epigenetic changes that occur during normal aging. As such, the establishment of future (inter)national collaborations in which assets from these different scientific disciplines are merged, *e.g.* involving clinicians, molecular neurobiologists and data scientists, are indispensable for the field to thrive.

In light of future iPSC models, all of the aforementioned efforts will also be highly instrumental. In fact, the availability of extensive and well-established tissue banks having both brain and other bodily tissues available from the same patients would allow for more reliable and detailed comparisons of *ex vivo* neural cells with iPSC-differentiated- or direct reprogrammed neural cultures. Such efforts could not only contribute to elaborating knowledge on molecular mechanisms underlying the disease, but also aid in further validating these novel humanized *in vitro* models. By studying these cells at the multi-omics level, it should be possible to obtain fingerprints of both iPSCs and their differentiated progeny, which allows one to assess how reprogramming affects biological processes, *e.g.* aging marks and epigenetic memory of the donor cells, thus solving many of the current challenges accompanying the use of these cells. Furthermore, the availability of such matched tissues could allow for the establishment of well-characterized cell panels of both healthy and diseased individuals, which would greatly benefiting future research into sAD. The availability of such well-defined iPSC and differentiated neural cultures might furthermore provide robust studies on the causality of epigenetic marks by using e.g. epigenetic editing systems in combination with exposures to environmental insults, and *vice versa*.

To conclude, the studies compiled in this thesis offer novel insights into both the role of epigenetic mechanisms in sAD, as well as into current state-of-the-art methodologies that are being used to study this disease. Hence, the research and accompanying strengths and limitations presented here, set the pillars for future endeavors to clarify the functional consequences of the identified epigenetic signatures in the etiology and progression of sAD. Altogether, the work presented in this thesis contributes to the global apprehension of contributing and modifying factors in relation to sAD that are crucially important for other scientists in the field aiming to further advance our current understanding, as well as to combat this devastating neurodegenerative disorder.

References

1. Liu X, Jiao B, Shen L. The epigenetics of Alzheimer's disease: Factors and therapeutic implications. Frontiers in genetics. 2018;9:579.
2. Lahiri DK, Maloney B, Zawia NH. The LEARn model: an epigenetic explanation for idiopathic neurobiological diseases. Molecular psychiatry. 2009;14(11):992-1003.
3. Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, et al. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. Journal of Alzheimer's Disease. 2012;29(3):571-88.
4. Sanchez-Mut JV, Aso E, Heyn H, Matsuda T, Bock C, Ferrer I, et al. Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. Hippocampus. 2014;24(4):363-8.
5. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.
6. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17(9):1156-63.
7. Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. Genome medicine. 2016;8(1):1-14.
8. Gasparoni G, Bultmann S, Lutsik P, Kraus TF, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics & chromatin. 2018;11(1):41.
9. Smith RG, Hannon E, De Jager PL, Chibnik L, Lott SJ, Condliffe D, et al. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimer's & Dementia. 2018;14(12):1580-8.
10. Semick SA, Bharadwaj RA, Collado-Torres L, Tao R, Shin JH, Deep-Soboslay A, et al. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. Acta neuropathologica. 2019;137(4):557-69.
11. Lardenoije R, Roubroeks JA, Pishva E, Leber M, Wagner H, Iatrou A, et al. Alzheimer's disease-associated (hydroxy) methylomic changes in the brain and blood. Clinical epigenetics. 2019;11(1):164.
12. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenetics. 2019;11(1):52.
13. Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841-51.
14. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012;336(6083):934-7.
15. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929-30.
16. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic acids research. 2010;38(19):e181-e.
17. Cheng Y, Xie N, Jin P, Wang T. DNA methylation and hydroxymethylation in stem cells. Cell biochemistry and function. 2015;33(4):161-73.
18. Song C-X, He C. Potential functional roles of DNA demethylation intermediates. Trends in biochemical sciences. 2013;38(10):480-4.

19. van den Hove DL, Riemens RJ, Koulousakis P, Pishva E. Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges. Brain Pathology. 2020.
20. Hajj NE, Kuhtz J, Haaf T. Limiting Dilution Bisulfite Pyrosequencing(R): A Method for Methylation Analysis of Individual DNA Molecules in a Single or a Few Cells. Methods Mol Biol. 2015;1315:221-39.
21. El Hajj N, Trapphoff T, Linke M, May A, Hansmann T, Kuhtz J, et al. Limiting dilution bisulfite (pyro)sequencing reveals parent-specific methylation patterns in single early mouse embryos and bovine oocytes. Epigenetics. 2011;6(10):1176-88.
22. Mattern F, Heinzmann J, Herrmann D, Lucas-Hahn A, Haaf T, Niemann H. Gene-specific profiling of DNA methylation and mRNA expression in bovine oocytes derived from follicles of different size categories. Reprod Fertil Dev. 2017;29(10):2040-51.
23. Trapphoff T, El Hajj N, Zechner U, Haaf T, Eichenlaub-Ritter U. DNA integrity, growth pattern, spindle formation, chromosomal constitution and imprinting patterns of mouse oocytes from vitrified pre-antral follicles. Human reproduction. 2010;25(12):3025-42.
24. Heinzmann J, Hansmann T, Herrmann D, Wrenzycki C, Zechner U, Haaf T, et al. Epigenetic profile of developmentally important genes in bovine oocytes. Molecular reproduction and development. 2011;78(3):188-201.
25. Diederich M, Hansmann T, Heinzmann J, Barg-Kues B, Herrmann D, Aldag P, et al. DNA methylation and mRNA expression profiles in bovine oocytes derived from prepubertal and adult donors. Reproduction. 2012;144(3):319.
26. Simic G, Stanic G, Mladinov M, Jovanov-Milosevic N, Kostovic I, Hof PR. Does Alzheimer's disease begin in the brainstem? Neuropathology and applied neurobiology. 2009;35(6):532-54.
27. Iatrou A, Kenis G, Rutten BP, Lunnon K, van den Hove DL. Epigenetic dysregulation of brainstem nuclei in the pathogenesis of Alzheimer's disease: looking in the correct place at the right time? Cell Mol Life Sci. 2017;74(3):509-23.
28. Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. Journal of Neuropathology & Experimental Neurology. 2011;70(11):960-9.
29. Grinberg L, Rüb U, Ferretti R, Nitrini R, Farfel J, Polichiso L, et al. The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset? Neuropathology and applied neurobiology. 2009;35(4):406-16.
30. Lee JH, Ryan J, Andreescu C, Aizenstein H, Lim HK. Brainstem morphological changes in Alzheimer's disease. Neuroreport. 2015;26(7):411.
31. Ji X, Wang H, Zhu M, He Y, Zhang H, Chen X, et al. Brainstem atrophy in the early stage of Alzheimer's disease: a voxel-based morphometry study. Brain Imaging and Behavior. 2021;15(1):49-59.
32. Chapuis J, Hot D, Hansmannel F, Kerdraon O, Ferreira S, Hubans C, et al. Transcriptomic and genetic studies identify IL-33 as a candidate gene for Alzheimer's disease. Molecular psychiatry. 2009;14(11):1004-16.
33. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. Journal of Alzheimer's Disease. 2011;23(2):349-59.
34. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. Alzheimer's research & therapy. 2019;11(1):5.
35. Chang T, editor Tau Network Genes in a Genome Wide Association Study of Progressive Supranuclear Palsy. MOVEMENT DISORDERS; 2017: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.
36. Valcourt U, Alcaraz LB, Exposito J-Y, Lethias C, Bartholin L. Tenascin-X: beyond the architectural function. Cell adhesion & migration. 2015;9(1-2):154-65.

37. Sullivan SE, Young-Pearse TL. Induced pluripotent stem cells as a discovery tool for Alzheimers disease. Brain Res. 2015.
38. Yang J, Li S, He XB, Cheng C, Le W. Induced pluripotent stem cells in Alzheimer's disease: applications for disease modeling and cell-replacement therapy. Mol Neurodegener. 2016;11(1):39.
39. Mungenast AE, Siegert S, Tsai LH. Modeling Alzheimer's disease with human induced pluripotent stem (iPS) cells. Mol Cell Neurosci. 2016;73:13-31.
40. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012;482(7384):216-20.
41. Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, et al. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. Cell Stem Cell. 2013;12(4):487-96.
42. Foveau B, Correia AS, Hébert SS, Rainone S, Potvin O, Kergoat M-J, et al. Stem cell-derived neurons as cellular models of sporadic Alzheimer's disease. Journal of Alzheimer's Disease. 2019;67(3):893-910.
43. Duan L, Bhattacharyya BJ, Belmadani A, Pan L, Miller RJ, Kessler JA. Stem cell derived basal forebrain cholinergic neurons from Alzheimer's disease patients are more susceptible to cell death. Mol Neurodegener. 2014;9:3.
44. Ochalek A, Mihalik B, Avci HX, Chandrasekaran A, Téglási A, Bock I, et al. Neurons derived from sporadic Alzheimer's disease iPSCs reveal elevated TAU hyperphosphorylation, increased amyloid levels, and GSK3B activation. Alzheimer's research & therapy. 2017;9(1):1-19.
45. Feng JA, Fouse S, Fan GP. Epigenetic regulation of neural gene expression and neuronal function. Pediatr Res. 2007;61(5):58r-63r.
46. Fitzsimons CP, van Bodegraven E, Schouten M, Lardenoije R, Kompotis K, Kenis G, et al. Epigenetic regulation of adult neural stem cells: implications for Alzheimer's disease. Mol Neurodegener. 2014;9:25.
47. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009;27(3):275-80.
48. Fernandez-Jimenez N, Allard C, Bouchard L, Perron P, Bustamante M, Bilbao JR, et al. Comparison of Illumina 450K and EPIC arrays in placental DNA methylation. Epigenetics. 2019;14(12):1177-82.
49. Wojda U, Kuznicki J. Alzheimer's disease modeling: ups, downs, and perspectives for human induced pluripotent stem cells. J Alzheimers Dis. 2013;34(3):563-88.
50. Han F, Liu C, Huang J, Chen J, Wei C, Geng X, et al. The application of patient-derived induced pluripotent stem cells for modeling and treatment of Alzheimer's disease. Brain Science Advances. 2019;5(1):21-40.
51. Haston KM, Finkbeiner S. Clinical Trials in a Dish: The Potential of Pluripotent Stem Cells to Develop Therapies for Neurodegenerative Diseases. Annu Rev Pharmacol Toxicol. 2015.
52. Riemens R, Kenis G, van den Beucken T. Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease. Neurobiology of learning and memory. 2020:107318.
53. Ooi L, Sidhu K, Poljak A, Sutherland G, O'Connor MD, Sachdev P, et al. Induced pluripotent stem cells as tools for disease modelling and drug discovery in Alzheimer's disease. J Neural Transm (Vienna). 2013;120(1):103-11.
54. Mooijman D, Dey SS, Boisset J-C, Crosetto N, Van Oudenaarden A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. Nature biotechnology. 2016;34(8):852-6.

55. Qu J, Zhou M, Song Q, Hong EE, Smith AD. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation. Bioinformatics. 2013;29(20):2645-6.
56. Kiihl SF, Martinez-Garrido MJ, Domingo-Relloso A, Bermudez J, Tellez-Plaza M. MLML2R: an R package for maximum likelihood estimation of DNA methylation and hydroxymethylation proportions. Statistical applications in genetics and molecular biology. 2019;18(1).
57. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC genomics. 2019;20(1):1-15.
58. Smith R, Pishva E, Shireby G, Smith AR, Roubroeks JA, Hannon E, et al. Meta-analysis of epigenome-wide association studies in Alzheime's disease highlights 220 differentially methylated loci across cortex. BioRxiv. 2020.
59. Titus AJ, Gallimore RM, Salas LA, Christensen BC. Cell-type deconvolution from DNA methylation: a review of recent applications. Human molecular genetics. 2017;26(R2):R216-R24.
60. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics. 2013;8(3):290-302.
61. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131:21-64.

Chapter 10

Summary

Chapter 1 introduced the reader to the field of neuroepigenomics in sporadic Alzheimer's disease (sAD), as well as to the use of induced pluripotent stem cell (iPSC)-based models for studying this neurodegenerative disorder. After this brief introduction, an overview of the general structure of the thesis was presented, which could be divide into two main parts focused on each of the aforementioned research lines was presented. In more detail, the studies presented in the first part of this thesis were centered on the role of epigenetic dysregulation in the etiopathophysiology of sAD. This first part, which includes **Chapters 2-5**, consisted of one perspective paper, two methodological research papers and one original research article**.** The work in the second part of this thesis, focused on the application and establishment of iPSC-based models for sAD, *e.g.* in view of mechanistic studies into epigenetic dysregulation. The second part, consisting of **Chapters 6-8**, presented two in-depth review articles and one exploratory pilot study.

In **Chapter 2**, a perspective paper on epigenome-wide association studies (EWAS), or more specifically methylome-wide association studies (MWAS), in sAD, was presented. Epigenetic mechanisms, which mediate the interaction between the genome and the environment, are thought to provide a mechanistic explanation in view of the etiopathogenesis of sAD. In this paper, all studies that have been performed to date targeting various (cortical) regions of the brain, as well as peripheral blood samples, derived from patients and matched non-demented controls, were briefly reviewed. Furthermore, relevant caveats in relation to these studies that challenge the interpretation of the experimental outcomes, including genomic coverage, statistical power, specificity of the epigenetic marks assessed, cell-type specificity and composition, causal inference and multi-omics, were discussed. Finally, an outlook on possible solutions that might overcome these challenges, including methodological developments and advances in related data science disciplines, was provided. Overall, what was concluded from this chapter is that although MWAS are highly relevant in the context of sAD research, it remains vital to address the aforementioned caveats in future studies in order to produce more meaningful data, which could significantly increase our understanding about the processes underlying the disease.

In view of these challenges, **Chapter 3** described a standardized protocol for oxidative bisulfite pyrosequencing that allows for an accurate discrimination between DNA methylation and hydroxymethylation, *i.e.* between 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) bases, in the context of cytosine-phosphate-guanine (CpG) sites. The pyrosequencing protocol described in this chapter provides a workable solution for the issue of

modification-specificity as faced in neuroepigenomic studies of sAD. The approach relies on a highly selective chemical oxidation using $KRuO_4$ that is applied prior to the bisulfite treatment, polymerase chain reaction (PCR) and pyrosequencing steps. Moreover, a novel spike-in DNA standard set that allows the user to accurately determine the oxidative bisulfite conversion efficiency was developed for specifically this approach. As a proof-of-principle, the protocol was conducted on both post-mortem brain tissue and cultured iPSCs, demonstrating that 5-mC and 5-hmC, as well as unmodified cytosine (5-uC) bases, could be detected in *OXT* and *DNAJB13*. Furthermore, it was confirmed that the novel (5-hmC) spike-in control could be used as an internal pyrosequencing control that does not interfere with the analysis of the accompanying sample, and *vice versa*. Such innovative approaches as presented in this chapter are crucial for moving the field of neuroepigenomics in sAD forward, as they aid in increasing the validity of the data.

Chapter 4 introduced a novel approach for DNA methylation profiling in small pools of 50 neurons, which relied on a combination of limiting dilution bisulfite pyrosequencing (LDBSP) and laser capture microdissection (LCM). This method allows one to determine CpG site methylation rates on single alleles in a multi-targeted and cell type-specific manner, hence providing a solution to the issue of cellular heterogeneity that is encountered in neuroepigenomic studies of sAD. The general working procedure of LDBSP on cells isolated from post-mortem brain tissue using LCM was described in this chapter. Furthermore, as proof-of-principle, a targeted methylation analysis of *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* and *OXT* was performed. Interestingly, LDBSP on these pools of neurons often rendered downstream reactions with more than one DNA molecule, a scenario that rarely occurs when conducting LDBSP on a one-cell-sample or just a few cells. As such, an adapted data analysis pipeline for LDBSP was developed, allowing one to include and correct CpG methylation rates derived from multi-allele reactions. Overall, the method described in this chapter provides the user with a more accurate estimation of the DNA methylation status of each target gene in the analyzed cell pools, thereby adding further validity to the data. In addition, it was demonstrated that the efficiency of LDBSP on neurons isolated with LCM is similar to the efficiency achieved in previously published studies using this technique on other isolated cell types. It is anticipated that single cell(type) approaches as described in this chapter will be increasingly valuable for future neuroepigenomic studies in sAD.

Chapter 5 represented the first large-scale epigenetic analysis in the brainstem of sAD to date, targeting the dorsal raphe nuclei (DRN) and locus coeruleus (LC), thereby complementing the body of EWAS reviewed in **Chapter 2**. Differentially

modified positions and regions in bulk tissues obtained from both brainstem nuclei were identified at the level of DNA methylation, hydroxymethylation and unmodified cytosines by using the Illumina Infinium MethylationEPIC BeadChip array in combination with an oxidative bisulfite treatment of the isolated target DNA (similar to what was described in **Chapter 3**). Aside from a strong overlapping dysregulation in the Tenascin XB gene (*TNXB*) in both the DRN and LC, common and novel epigenetic signatures compared to previous studies targeting the sAD brain were discovered, all of which may play a pivotal role in the etiopathogenesis of sAD. A subsequent bisulfite pyrosequencing analysis in the same patient cohort confirmed the observed dysregulation of *TNXB* in both of the brainstem regions assessed. As a follow up on these discovery findings, a targeted analysis of *TNXB* was performed in an independent patient cohort, assessing methylomic signatures in single serotonergic neurons and non-serotonergic cells isolated from the DRN by means of LCM. This study therefore represented the first cell subtype-specific validation analysis performed in the brainstem to date, relying on the LCM-LDBSP approach that was described in **Chapter 4**. Strikingly, when comparing the bisulfite methylation levels of *TNXB* between sAD patients and controls in serotonergic and non-serotonergic neurons, a significant interaction between cell-type and experimental condition was identified. The sAD-associated methylation profiles were opposite in the serotonergic neurons and non-serotonergic cells, the latter of which resembled the EWAS data. As such, this study demonstrated for the first time that epigenetic signatures within the DRN can strongly dependent on both the disease phenotype and the cell type analyzed. These findings therefore emphasize the importance of future cell type-specific neuroepigenomic studies in sAD, which was already briefly addressed in **Chapters 2** and **4**.

Chapter 6 presented a review paper on the potential value of iPSCs for sAD research. Since their discovery, these cells have been offering a promising avenue to fill the translational gap between pre-clinical and clinical sAD research, by allowing the establishment of patient-specific *in vitro* disease models that can be applied for fundamental research and drug discovery. It is anticipated that efforts in this field will deepen our knowledge on various underlying disease mechanisms and aid in the establishment of therapeutic interventions. All pioneering studies utilizing iPSCs from sAD patients were reviewed in this chapter, demonstrating that the iPSC-derived neural cells recapitulate neuropathological processes of the disease, although with quite a high degree of variability in terms of their presence and severity. Therefore, sources of variability related to the model in addition to those that might be explained by the heterogeneous nature of sAD were critically assessed. What followed from this chapter is that developing iPSC models for sAD remains a challenging endeavor

due to the multifactorial nature of the disease, the nearly life-long disease progression and the high degree of inter-individual heterogeneity that might be reflected in cells obtained from patients. Aside from the variability associated with the disease, several methodological factors, such as those related to iPSC generation and (neural) differentiation, further impact on this degree of variation. Thus, future efforts aimed at extensively characterizing the iPSC-derived models and sources of variability in fine detail, are necessary if one wishes to establish robust disease models for the study of sAD. Nonetheless, whilst taking into account these considerations, the developing iPSC models clearly provide an exciting avenue in sAD research.

Within the framework of developing iPSC-based models, **Chapter 7** presented an in-depth review article on directed- and direct neural differentiation protocols starting from stem cells and somatic cells, respectively. As addressed in this review, insights from basic research and developmental biology have guided the design of current protocols and numerous approaches for the derivation of regional-specific glutamatergic, dopaminergic, GABAergic, serotonergic, and cholinergic/motor neurons have become available. Approaches for the derivation of each of these disease-relevant neural subtypes were summarized, relying on either chemically defined systems using patterning factors, transcription factor-mediated reprogramming and epigenetic-based strategies. This review furthermore highlighted that although distinct approaches have shown to be successful in directing neuronal cell fate *in vitro*, their refinement and optimization, as well as the search for alternative strategies, remains necessary to help realize the full potential of the eventually derived neuronal populations. Furthermore, existing protocols are still limited in the number of neuronal subtypes whose induction is fully established, and different cultivation protocols for each subtype exist. As such, future detailed characterization of the cellular and molecular characteristics involved in guiding stem cell differentiation and somatic cell reprogramming along the neural lineage is expected to contribute to the development of highly robust protocols. Evidently, these efforts could aid in the development of highly meaningful *in vitro* disease models for disorders such as sAD.

Finally, **Chapter 8** comprised an exploratory pilot study, in which the knowledge of **Chapters 6** and **7** was combined in an effort to establish and characterize a cortical forebrain differentiation protocol from iPSCs. The overarching aim of this chapter was to use this differentiation protocol for the establishment of an *in vitro* model relevant for sAD. First, the stemness of the iPSC-line, which was used for neural differentiation, was confirmed based on the expression of pluripotency markers such as OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 and TRA-1-81.

Next, by the sequential application of neural patterning factors, iPSCs were differentiated to neural progenitor cells (NPCs) and subsequently to cortical neurons and glia. At each stage of the differentiation process, the expression of key-lineage transcription factors was assessed, which confirmed that the cells progressed towards a cortical forebrain fate. Successful derivation of NPCs was confirmed by the expression of NESTIN, SOX2 and PAX6. Quantification of these markers furthermore revealed that the iPSCs differentiated towards relative homogeneous population of NPCs (>82%). Immunofluorescent examination following differentiation of the NPCs verified the presence of the neuronal markers such as MAP2AB, DCX and TUJ1, and the astrocytic marker GFAP. Additionally, the presence of cortical forebrain markers, including CTIP2 and FOXG1, were identified in the differentiated cultures. Overall, these preliminary findings confirmed that the protocol described in this chapter was successful in differentiating the iPSCs into a mixed population of cortical forebrain neurons and glia. Although further functional characterization of these neural cells remains necessary, it was postulated that the protocol applied in this chapter can be used for the establishment of a cortical *in vitro* disease model system for sAD. In order to further explore this potential, expression of TNXB and the oxytocin receptor (OXTR) in the iPSC-derived cortical cultures was assessed. As previously demonstrated in **Chapter 4** of this thesis, epigenetic and genetic dysregulation of *TNXB* have been associated with AD pathophysiology in multiple regions of the sAD brain. In addition, epigenetic deregulation of the oxytocin gene (*OXT*) and alterations in OXT signaling have previously also been implicated in sAD or associated disease phenotypes. Interestingly, expression of both TNXB and OXTR could be detected in the differentiated neural cells already at 14 days and maintained up until 56 days of NPC-differentiation, suggesting that these cortical forebrain cells might represent an appealing model system to study mechanistic processes related to *TNXB* and *OXT* signaling in the context of sAD. All in all, efforts as attained in this chapter are imperative if one wishes to establish disease-relevant iPSC-based models.

Chapter 11

Samenvatting

Hoofdstuk 1 introduceerde de lezer op het gebied van neuroepigenetica in relatie tot de sporadische ziekte van Alzheimer (sAD), evenals op het gebruik van geïnduceerde pluripotente stamcel (iPSC)-gebaseerde modellen voor het bestuderen van deze neurodegeneratieve aandoening. Na deze korte inleiding werd een overzicht gepresenteerd van de algemene structuur van het proefschrift, dat zou kunnen worden onderverdeeld in twee hoofddelen gericht op elk van de bovengenoemde onderzoekslijnen. In meer detail, de studies gepresenteerd in het eerste deel van dit proefschrift zijn gericht op de rol van epigenetische ontregelingen in de etiopathofysiologie van sAD. Dit eerste deel, dat de **Hoofdstukken 2-5** omvat, bestond uit een perspectief artikel, twee methodologische onderzoekartikelen en een origineel onderzoekartikel. Het werk in het tweede deel van dit proefschrift was gericht op de toepassing en totstandkoming van iPSC-gebaseerde modellen voor sAD, bijvoorbeeld met het oog op mechanistische studies naar epigenetische ontregeling. Het tweede deel, bestaande uit de **Hoofdstukken 6-8**, bevatte twee diepgaande overzichtsartikelen en een verkennende pilotstudie.

In **Hoofdstuk 2**, wordt een perspectief artikel over epigenoom-brede associatiestudies (EWAS), of meer specifiek methylome-brede associatiestudies (MWAS), in sAD gepresenteerd. Epigenetische mechanismen, die de interactie tussen het genoom en de omgeving mediëren, worden verondersteld een mechanistische verklaring te bieden met het oog op de etiopathogenese van sAD. In dit artikel werden alle onderzoeken die tot nu toe zijn uitgevoerd, gericht op verschillende (corticale) hersengebieden, evenals perifere bloedmonsters, afkomstig van patiënten en overeenkomende niet-demente controlepersonen, kort besproken. Verder werden relevante kanttekeningen met betrekking tot deze onderzoeken die de interpretatie van de onderzoeksresultaten beïnvloeden besproken, waaronder genomische dekking, statistische bewijskracht, specificiteit van de geëvalueerde epigenetische kenmerken, celtypespecificiteit en samenstelling, causale inferentie en multi-omica. Ten slotte werd een vooruitblik gegeven op mogelijke oplossingen om deze uitdagingen het hoofd te bieden, bestaande uit methodologische ontwikkelingen en vorderingen in gerelateerde datawetenschap-disciplines. In het algemeen werd uit dit hoofdstuk geconcludeerd dat, hoewel MWAS zeer relevant zijn in de context van sAD-onderzoek, het essentieel blijft om de bovengenoemde kanttekeningen in toekomstige studies aan te pakken om meer zinvolle gegevens te produceren, die het begrip van de processen die ten grondslag liggen aan de ziekte aanzienlijk zouden kunnen vergroten.

In relatie tot bovengenoemde uitdagingen, beschreef **Hoofdstuk 3** een gestandaardiseerd protocol voor oxidatieve bisulfiet pyrosequentie bepaling

waarmee een nauwkeurig onderscheid kan worden gemaakt tussen DNA-methylatie en hydroxymethylatie, dat wil zeggen tussen 5-methylcytosine (5-mC) en 5-hydroxymethylcytosine (5-hmC) basen, in de context van cytosine-fosfaat-guanine (CpG) posities. Het pyrosequentie bepaling protocol dat in dit hoofdstuk wordt beschreven, biedt een werkbare oplossing voor het probleem van modificatiespecificiteit waarmee neuroepigenetische onderzoeken naar sAD worden geconfronteerd. De aanpak is gebaseerd op een zeer selectieve chemische oxidatie met behulp van $KRuO_4$ die wordt toegepast voorafgaand aan de bisulfiet behandeling, polymerasekettingreactie (PCR) en pyrosequentie bepaling stappen. Bovendien werd specifiek voor deze methode een nieuwe spike-in DNA-standaard set ontwikkeld waarmee de gebruiker nauwkeurig de oxidatieve bisulfiet conversie-efficiëntie kan bepalen. Om het principe aan te tonen werd het protocol uitgevoerd op zowel post-mortem hersenweefsel als iPSCs, waarmee werd aangetoond dat 5-mC en 5-hmC, evenals niet gemodificeerde cytosine (5-uC) basen, kunnen worden gedetecteerd in *OXT* en *DNAJB13*. Bovendien werd bevestigd dat de nieuwe (5-hmC) spike-in controle zou kunnen worden gebruikt als een interne pyrosequentie bepaling-controle die de analyse van het doelmonster niet verstoort, en andersom. Dergelijke innovatieve benaderingen zoals gepresenteerd in dit hoofdstuk zijn cruciaal om het veld van neuroepigenetica in sAD vooruit te helpen, omdat deze de validiteit van de onderzoeksresultaten vergroten.

Hoofdstuk 4 introduceerde een nieuwe methode voor DNA-methylatie profilering in kleine clusters van 50 neuronen, die gebaseerd was op een combinatie van limiting dilution bisulfite pyrosequencing (LDBSP) en laser capture microdissectie (LCM). Met deze methode kan men de methylatie graad van CpG posities op afzonderlijke allelen en op een meervoudig-gerichte en celtype-specifieke manier bepalen, waardoor er een oplossing wordt geboden voor het probleem van cellulaire heterogeniteit dat wordt aangetroffen in neuroepigenetisch onderzoek naar sAD. De algemene werkwijze van LDBSP op cellen geïsoleerd uit post-mortem hersenweefsel met behulp van LCM werd in dit hoofdstuk beschreven. Bovendien werd om het principe aan te tonen een gerichte methylatie analyse van *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* en *OXT* uitgevoerd. Interessant is dat LDBSP op deze clusters van neuronen vaak afgeleide reacties vertoonde met meer dan één DNA-molecuul, een scenario dat zelden voorkomt bij het uitvoeren van LDBSP op een eencellig monster of slechts een paar cellen. Als zodanig werd een aangepaste methode voor de analyse van de onderzoeksresultaten voor LDBSP ontwikkeld, die het mogelijk maakt om de methylatie graad van CpG posities die zijn verkregen van multi-allelreacties op te nemen en te corrigeren. Over het algemeen biedt de methode die in dit hoofdstuk wordt beschreven de gebruiker een nauwkeurigere schatting van de DNA-

methylatie status van elk doel gen in de geanalyseerde cel clusters, waardoor de validiteit van de onderzoeksresultaten toenemen. Bovendien werd aangetoond dat de efficiëntie van LDBSP op neuronen geïsoleerd met LCM vergelijkbaar is met de efficiëntie die werd bereikt in eerder gepubliceerde onderzoeken met deze techniek op andere geïsoleerde celtypen. Verwacht wordt dat analyses van een individuele cel(type), zoals beschreven in dit hoofdstuk, steeds waardevoller zullen worden voor toekomstige neuroepigenetische studies naar sAD.

Hoofdstuk 5 vertegenwoordigde de eerste grootschalige epigenetische analyse in de hersenstam van sAD tot nu toe, gericht op de dorsale raphe nuclei (DRN) en locus coeruleus (LC), en vormt daarmee een aanvulling op het geheel van EWAS besproken in **Hoofdstuk 2**. Differentiaal gemodificeerde posities en regio's in bulkweefsels verkregen uit beide hersenstamkernen werden geïdentificeerd op het niveau van DNA-methylatie, hydroxymethylatie en niet gemodificeerde cytosines door gebruik te maken van de Illumina Infinium MethylationEPIC BeadChip-array in combinatie met een oxidatieve bisulfiet behandeling van het geïsoleerde doel-DNA (vergelijkbaar met wat werd beschreven in **Hoofdstuk 3**). Naast een sterke overlappende ontregeling in het Tenascin XB gen (*TNXB*) in zowel de DRN als de LC, werden gemeenschappelijke en nieuwe epigenetische patronen ontdekt in vergelijking met eerdere onderzoeken gericht op de sAD-hersenen, die allemaal een cruciale rol kunnen spelen in de etiopathogenese van sAD. Een daaropvolgende analyse gebruikmakend van een bisulfiet pyrosequentie bepaling in hetzelfde patiënten cohort bevestigde de waargenomen ontregeling van *TNXB* in beide onderzochte hersenstamgebieden. Als vervolg op deze bevindingen werd een gerichte analyse van *TNXB* uitgevoerd in een onafhankelijk patiënten cohort, waarbij methylatie patronen werden onderzocht in enkele serotonerge neuronen en niet-serotonerge cellen geïsoleerd uit de DRN door middel van LCM. Deze studie vertegenwoordigde daarom de eerste cel subtype-specifieke validatie analyse die tot nu toe in de hersenstam is uitgevoerd, gebaseerd op de LCM-LDBSP-methode die werd beschreven in **Hoofdstuk 4**. Opvallend is dat bij het vergelijken van de bisulfiet methylatie niveaus in *TNXB* tussen sAD-patiënten en controles in serotonerge en niet-serotonerge neuronen, een significante interactie tussen celtype en experimentele conditie werd geïdentificeerd. De sAD-geassocieerde methylatieprofielen waren tegengesteld in de serotonerge neuronen en niet-serotonerge cellen, waarvan de laatste vergelijkbaar waren met de EWAS-onderzoeksresultaten. Als zodanig toonde deze studie voor het eerst aan dat epigenetische patronen binnen de DRN sterk kunnen afhangen van zowel het ziektebeeld als het geanalyseerde celtype. Deze bevindingen benadrukken daarom het belang van toekomstige celtype-specifieke neuroepigenetische studies naar sAD, wat al kort aan de orde kwam in de **Hoofdstukken 2** en **4**.

Hoofdstuk 6 presenteerde een overzichtsartikel over de potentiële waarde van iPSCs voor sAD-onderzoek. Sinds hun ontdekking bieden deze cellen een veelbelovende manier om de translationele kloof tussen preklinisch en klinisch sAD-onderzoek te dichten, door het opzetten van patiënt-specifieke *in vitro* ziektemodellen die kunnen worden toegepast voor fundamenteel onderzoek en ontdekking van geneesmiddelen. Verwacht wordt dat inspanningen op dit gebied onze kennis over verschillende onderliggende ziektemechanismen zullen verdiepen en zullen helpen bij het opzetten van therapeutische interventies. Alle baanbrekende onderzoeken met iPSCs van sAD-patiënten werden in dit hoofdstuk besproken, welke hebben aangetoond dat de iPSC-afgeleide neurale cellen de neuropathologische processen van de ziekte recapituleren, hoewel met een vrij hoge mate van variabiliteit in termen van hun aanwezigheid en ernst. Derhalve werden bronnen van variabiliteit gerelateerd aan het model, naast degene die zouden kunnen worden verklaard door de heterogene aard van sAD, kritisch belicht. Voortvloeiend uit dit hoofdstuk, is dat het ontwikkelen van iPSC-modellen voor sAD een uitdagende inspanning blijft vanwege de multifactoriële aard van de ziekte, de bijna levenslange ziekteprogressie en de hoge mate van interindividuele heterogeniteit die kan worden weerspiegeld in cellen die zijn verkregen van patiënten. Afgezien van de variabiliteit die met de ziekte gepaard gaat, hebben verschillende methodologische factoren, zoals die gerelateerd aan iPSC-generatie en (neurale) differentiatie, een verdere invloed op deze mate van variatie. Toekomstige inspanningen gericht op het uitgebreid karakteriseren van de iPSC-afgeleide modellen en bronnen van variabiliteit in fijn detail, zijn dus noodzakelijk als men robuuste ziektemodellen wil opzetten voor de studie van sAD. Ondanks deze overwegingen bieden de zich ontwikkelende iPSC-modellen duidelijk een opwindende weg in sAD onderzoek.

In het kader van het ontwikkelen van iPSC-gebaseerde modellen, presenteerde **Hoofdstuk 7** een diepgaand overzichtsartikel over aangestuurde en directe neurale differentiatieprotocollen, uitgaande van respectievelijk stamcellen en somatische cellen. Zoals in dit overzicht wordt besproken, hebben inzichten uit fundamenteel onderzoek en ontwikkelingsbiologie het ontwerp van huidige protocollen geleid en zijn er talloze methoden beschikbaar voor de afleiding van regionaal-specifieke glutamaterge, dopaminerge, GABAergische, serotonerge en cholinerge/motor neuronen. Methoden voor de afleiding van elk van deze ziekte gerelateerde neurale subtypes werden samengevat, gebaseerd op ofwel chemisch gedefinieerde systemen met behulp van patroonfactoren, op transcriptiefactoren-gemedieerde herprogrammering en op epigenetische-gebaseerde strategieën. Dit review benadrukte verder dat hoewel verschillende methoden succesvol zijn gebleken bij het aansturen van de lotsbestemming van

neuronale cellen *in vitro*, hun verfijning en optimalisatie, evenals het zoeken naar alternatieve strategieën, noodzakelijk blijven om het volledige potentieel van de uiteindelijk afgeleide neuronale populaties te helpen realiseren. Bovendien zijn de bestaande protocollen nog steeds beperkt in het aantal neuronale subtypen waarvan de inductie volledig is vastgesteld, en er bestaan verschillende kweekprotocollen voor elk subtype. Als zodanig wordt verwacht dat toekomstige gedetailleerde karakterisering van de cellulaire en moleculaire kenmerken die betrokken zijn bij het begeleiden van stamceldifferentiatie en somatische cel herprogrammering langs de neurale lijn, zal bijdragen aan de ontwikkeling van zeer robuuste protocollen. Het is evident dat deze inspanningen kunnen helpen bij de ontwikkeling van zeer zinvolle *in vitro* ziektemodellen voor aandoeningen zoals sAD.

Ten slotte omvatte **Hoofdstuk 8** een verkennende pilotstudie, waarin de kennis van **Hoofdstuk 6** en **7** werd gecombineerd in een poging om een corticale voorhersendifferentiatieprotocol vanuit iPSCs op te zetten en te karakteriseren. Het overkoepelende doel van dit hoofdstuk was om dit differentiatieprotocol te gebruiken voor het opzetten van een *in vitro* model dat relevant is voor sAD. Ten eerste werd de stamcel karakteristieken van de iPSC-lijn, die werd gebruikt voor neurale differentiatie, bevestigd op basis van de expressie van pluripotente markers zoals OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 en TRA-1-81. Vervolgens werden iPSCs door de opeenvolgende toepassing van neurale patroonfactoren gedifferentieerd naar neurale voorlopercellen (NPCs) en vervolgens naar corticale neuronen en glia. In elke fase van het differentiatieproces werd de expressie van belangrijke afstamming transcriptiefactoren beoordeeld, wat bevestigde dat de cellen vorderde naar een corticale voorhersenen lotsbestemming. Succesvolle afleiding van NPCs werd bevestigd door de expressie van NESTIN, SOX2 en PAX6. Kwantificering van deze markers onthulde verder dat de iPSCs differentieerde naar een relatief homogene populatie van NPCs (> 82%). Immunofluorescentie bepaling na differentiatie van de NPCs bevestigde de aanwezigheid van de neuronale markers zoals MAP2AB, DCX en TUJ1, en de astrocytische marker GFAP. Bovendien werd de aanwezigheid van markers van de corticale voorhersenen, waaronder CTIP2 en FOXG1, geïdentificeerd in de gedifferentieerde culturen. Over het algemeen bevestigden deze voorlopige bevindingen dat het protocol dat in dit hoofdstuk wordt beschreven, succesvol was in het differentiëren van de iPSCs in een gemengde populatie van corticale frontale neuronen en glia. Hoewel verdere functionele karakterisering van deze neurale cellen noodzakelijk blijft, werd gepostuleerd dat het protocol dat in dit hoofdstuk wordt toegepast gebruikt kan worden voor het opzetten van een corticaal *in vitro* ziektemodelsysteem voor sAD. Om dit potentieel verder te onderzoeken, werd

de expressie van TNXB en de oxytocinereceptor (OXTR) in de iPSC-afgeleide corticale culturen vastgesteld. Zoals eerder aangetoond in **Hoofdstuk 4** van dit proefschrift, zijn epigenetische en genetische ontregeling van TNXB geassocieerd met AD pathofysiologie in meerdere regio's van de sAD hersenen. Bovendien was epigenetische dysregulatie van het oxytocine-gen (*OXT*) en veranderingen in OXT-signalering eerder ook betrokken bij sAD of geassocieerde ziektefenotypes. Interessant is dat expressie van zowel TNXB als OXTR al na 14 dagen in de gedifferentieerde neurale cellen kon worden gedetecteerd en werd gehandhaafd tot 56 dagen na NPC-differentiatie, wat suggereert dat deze corticale voorhersenencellen een aantrekkelijk modelsysteem zouden kunnen zijn om mechanistische processen gerelateerd aan TNXB en OXT-signalering te bestuderen in de context van sAD. Al met al zijn de inspanningen die in dit hoofdstuk werden verricht noodzakelijk als men ziektegerelateerde iPSC-gebaseerde modellen wil ontwikkelen.

Chapter 12

Zusammenfassung

Kapitel 1 führte den Leser in das Gebiet der Neuroepigenomik bei sporadischer Alzheimer-Krankheit (sAD) sowie in die Verwendung von auf induzierten pluripotenten Stammzellen (iPSC) basierenden Modellen zur Untersuchung dieser neurodegenerativen Störung ein. Nach dieser kurzen Einführung wurde ein Überblick über die allgemeine Struktur der Arbeit gegeben, der in zwei Hauptteile unterteilt werden konnte, die sich auf jede der oben genannten Forschungslinien konzentrierten. Ausführlicher, konzentrierten sich die im ersten Teil dieser Arbeit vorgestellten Studien auf die Rolle der epigenetischen Dysregulation in der Ätiopathophysiologie von sAD. Dieser erste Teil, der die **Kapitel 2** bis **5** enthält, bestand aus einem Perspektivpapier, zwei methodischen Forschungspapieren und einem Original-Forschungsartikel. Die Arbeit im zweiten Teil dieser Arbeit konzentrierte sich auf die Anwendung und Etablierung von iPSC-basierten Modellen für sAD, zum Beispiel im Hinblick auf mechanistische Studien zur epigenetischen Dysregulation. Der zweite Teil, bestehend aus den **Kapiteln 6-8**, enthielt zwei ausführliche Übersichtsartikel und eine explorative Pilotstudie.

In **Kapitel 2**, wurde ein Perspektivpapier zu epigenomweiten Assoziationsstudien (EWAS) oder genauer zu methylomweiten Assoziationsstudien (MWAS) in sAD vorgestellt. Es wird angenommen, dass epigenetische Mechanismen, die die Wechselwirkung zwischen Genom und Umwelt vermitteln, eine mechanistische Erklärung im Hinblick auf die Ätiopathogenese von sAD liefern. In diesem Artikel wurden alle bisher durchgeführten Studien, die auf verschiedene (kortikale) Regionen des Gehirns abzielten, sowie periphere Blutproben, die von Patienten stammen und mit nicht dementen Kontrollen übereinstimmen, kurz besprochen. Darüber hinaus wurden relevante Vorbehalte in Bezug auf diese Studien diskutiert, die die Interpretation der experimentellen Ergebnisse in Frage stellen, einschließlich der genomischen Abdeckung, der statistischen Aussagekraft, der Spezifität der bewerteten epigenetischen Markierungen, der Spezifität und Zusammensetzung des Zelltyps, der kausalen Inferenz und der Multi-Omics. Schließlich wurde ein Ausblick auf mögliche Lösungen gegeben, die diese Herausforderungen bewältigen könnten, einschließlich methodischer Entwicklungen und Fortschritte in verwandten datenwissenschaftlichen Disziplinen. Insgesamt wurde aus diesem Kapitel der Schluss gezogen, dass MWAS zwar im Kontext der sAD-Forschung von hoher Relevanz sind, es jedoch weiterhin wichtig ist, die oben genannten Vorbehalte in zukünftigen Studien zu berücksichtigen, um aussagekräftigere Daten zu erhalten, die unser Verständnis der Prozesse die der Krankheit zugrunde liegen erheblich verbessern könnten.

In Anbetracht dieser Herausforderungen wurde in **Kapitel 3** ein standardisiertes Protokoll für die oxidative Bisulfit Pyrosequenzierung beschrieben, das eine

genaue Unterscheidung zwischen DNA-Methylierung und -Hydroxymethylierung ermöglicht, das heißt zwischen 5-Methylcytosin (5-mC) und 5-Hydroxymethylcytosin (5-hmC)-Basen in der Kontext von Cytosin-Phosphat-Guanin (CpG)-Stellen. Das in diesem Kapitel beschriebene Pyrosequenzierungsprotokoll bietet eine praktikable Lösung für das Problem der Modifikationsspezifität, mit dem neuroepigenomische Studien von sAD konfrontiert sind. Der Ansatz beruht auf einer hochselektiven chemischen Oxidation unter Verwendung von KRuO4, die vor der Bisulfitbehandlung, der Polymerasekettenreaktion (PCR) und den Pyrosequenzierungs schritten angewendet wird. Darüber hinaus wurde für diese Methode ein neuartiger Spike-In-DNA-Standardsatz entwickelt, mit dem der Benutzer die Effizienz der oxidativen Bisulfitumwandlung genau bestimmen kann. Als Beweis des Prinzips wurde das Protokoll sowohl an post mortem Hirngewebe als auch an iPSCs durchgeführt, was zeigte, dass 5-mC und 5-hmC sowie unmodifizierte Cytosin (5-uC)-Basen in *OXT* und *DNAJB13* nachgewiesen werden konnte. Darüber hinaus wurde bestätigt, dass die neuartige (5-hmC) Spike-In-Kontrolle als interne Pyrosequenzierungskontrolle verwendet werden kann, die die Analyse der Ziel Probe nicht beeinträchtigt, und umgekehrt. Solche innovativen Ansätze, wie sie in diesem Kapitel vorgestellt werden, sind entscheidend, um das Gebiet der Neuroepigenomik in sAD voranzubringen, da sie dazu beitragen die Gültigkeit der Daten zu erhöhen.

In **Kapitel 4** wurde eine neuartigere Methode für die Erstellung von DNA-Methylierungsprofilen in kleinen Kolonien von 50 Neuronen vorgestellt, der auf einer Kombination aus Grenzverdünnungs-Bisulfit-Pyrosequenzierung (LDBSP) und Laser-Capture-Mikrodissektion (LCM) beruhte. Diese Methode ermöglicht es, die Methylierungsraten der CpG-Stelle auf einzelnen Allelen auf eine vielfach zielgerichtete und zelltypspezifische Weise zu bestimmen, wodurch eine Lösung für das Problem der zellulären Heterogenität bereitgestellt wird das in neuroepigenomischen Studien von sAD auftritt. In diesem Kapitel wurde das allgemeine Arbeitsverfahren von LDBSP an Zellen beschrieben, die mit LCM aus post mortem Hirngewebe isoliert wurden. Darüber hinaus wurde als Beweis des Prinzips eine gezielte Methylierungsanalyse der *DNAJB13*, *PGLYRP1*, *RHBDF2*, *C3*, *LMX1B* und *OXT* durchgeführt. Interessanterweise führte LDBSP in diesen Neuronekolonien häufig zu nachgeschalteten Reaktionen mit mehr als einem DNA-Molekül, ein Szenario das selten auftritt wenn LDBSP an einer Einzellprobe oder nur wenigen Zellen durchgeführt wird. Als solches wurde eine angepasste Datenanalyse-Pipeline für LDBSP entwickelt, die es ermöglicht CpG-Methylierungsraten die aus Multi-Allel-Reaktionen abgeleitet wurden einzuschließen und zu korrigieren. Insgesamt bietet die in diesem Kapitel beschriebene Methode dem Benutzer eine genauere Schätzung des DNA-

Methylierungsstatus jedes Zielgens in den analysierten Zellpools, wodurch die Daten weiter validiert werden. Darüber hinaus wurde gezeigt, dass die Effizienz von LDBSP bei mit LCM isolierten Neuronen ähnlich der Effizienz ist, die in zuvor veröffentlichten Studien unter Verwendung dieser Technik bei anderen isolierten Zelltypen erzielt wurde. Es wird erwartet, dass Einzelzell(typ)ansätze, wie sie in diesem Kapitel beschrieben werden, für zukünftige neuroepigenomische Studien in sAD zunehmend wertvoll sein werden.

Kapitel 5 stellte die erste groß angelegte epigenetische Analyse im Hirnstamm von sAD dar, die auf die dorsalen Raphekerne (DRN) und den Locus coeruleus (LC) abzielte und damit die in **Kapitel 2** besprochenen EWAS-Literaturergänzte. Aus beiden Hirnstammkernen erhaltene Massengewebe wurden Differential modifizierte Positionen und Regionen identifiziert auf der Ebene der DNA-Methylierung, -Hydroxymethylierung und unmodifizierten Cytosine unter Verwendung des Illumina Infinium MethylationEPIC BeadChip-Arrays in Kombination mit einer oxidativen Bisulfit-Behandlung der isolierten Ziel-DNA (ähnlich wie in **Kapitel 3** beschrieben). Abgesehen von einer stark überlappenden Dysregulation im Tenascin XB-Gen (*TNXB*), sowohl im DRN als auch im LC, wurden gemeinsame und neuartige epigenetische Signaturen im Vergleich zu früheren Studien zum sAD-Gehirn entdeckt, die alle eine entscheidende Rolle bei der Ätiopathogenese von sAD spielen können. Eine anschließende Bisulfit-Pyrosequenzierungsanalyse in derselben Patientenkohorte bestätigte die beobachtete Dysregulation von *TNXB* in beiden untersuchten Hirnstammregionen. Im Anschluss an diese Entdeckungsergebnisse wurde eine gezielte Analyse von TNXB in einer unabhängigen Patientenkohorte durchgeführt, wobei die methylomischen Signaturen in einzelnen serotonergen Neuronen und nicht serotonergen Zellen bewertet wurden, die mittels LCM aus dem DRN isoliert wurden. Diese Studie stellte daher die erste zellsubtypspezifische Validierungsanalyse dar, die bisher im Hirnstamm durchgeführt wurde und sich auf den in **Kapitel 4** beschriebenen LCM-LDBSP-Ansatz stützte. Bemerkenswerterweise beim Vergleich der Bisulfitmethylierungsniveaus von *TNXB* zwischen sAD-Patienten und Kontrollen bei Serotonergen und nicht serotonergen Neuronen wurde eine signifikante Wechselwirkung zwischen Zelltyp und experimentellem Zustand identifiziert. Die sAD-assoziierten Methylierungsprofile waren in den serotonergen Neuronen und nicht serotonergen Zellen entgegengesetzt, wobei letztere den EWAS-Daten ähnelten. Als solche zeigte diese Studie zum ersten Mal, dass epigenetische Signaturen innerhalb des DRN sowohl vom Krankheitsbild als auch vom analysierten Zelltyp stark abhängen können. Diese Ergebnisse unterstreichen daher die Bedeutung zukünftiger zelltypspezifischer neuroepigenomischer Studien bei sAD, auf die bereits in den **Kapiteln 2** und **4** kurz eingegangen wurde.

In **Kapitel 6** wurde ein Übersichtsartikel über den potenziellen Wert von iPSCs für die sAD-Forschung vorgestellt. Seit ihrer Entdeckung bieten diese Zellen einen vielversprechenden Weg, um die Translationslücke zwischen präklinischer und klinischer sAD-Forschung zu schließen, indem sie die Erstellung patientenspezifischer In-vitro-Krankheitsmodelle ermöglichen, die für die Grundlagenforschung und die Wirkstoffentdeckung angewendet werden können. Es wird erwartet, dass die Bemühungen auf diesem Gebiet unser Wissen über verschiedene zugrundeliegende Krankheitsmechanismen vertiefen und bei der Etablierung therapeutischer Interventionen helfen werden. Alle wegweisenden Studien unter Verwendung von iPSCs von sAD-Patienten wurden in diesem Kapitel überprüft, die zeigen, dass die von iPSC abgeleiteten Nervenzellen neuropathologische Prozesse der Krankheit rekapitulieren, wenn auch mit einem recht hohen Grad an Variabilität hinsichtlich ihres Vorhandenseins und ihrer Schwere. Daher wurden Variabilitätsquellen in Bezug auf das Modell zusätzlich zu denen, die durch die heterogene Natur von sAD erklärt werden könnten, kritisch diskutiert. Aus diesem Kapitel geht hervor, dass die Entwicklung von iPSC-Modellen für sAD ein herausforderndes Unterfangen bleibt aufgrund der multifaktoriellen Natur der Krankheit, des nahezu lebenslangen Fortschreitens der Krankheit und des hohen Grads an interindividueller Heterogenität, die sich in Zellen von Patienten widerspiegeln könnte. Abgesehen von der mit der Krankheit verbundenen Variabilität wirken sich verschiedene methodische Faktoren, wie die mit der iPSC-Erzeugung und der (neuronalen) Differenzierung verbundenen, weiter auf diesen Variationsgrad aus. Daher sind zukünftige Anstrengungen erforderlich, um die von iPSC abgeleiteten Modelle und Variabilitätsquellen detailliert zu charakterisieren, wenn robuste Krankheitsmodelle für die Untersuchung von sAD erstellt werden sollen. Unter Berücksichtigung dieser Überlegungen bieten die sich entwickelnden iPSC-Modelle jedoch eindeutig einen spannenden Ansatzin der sAD-Forschung.

Im Verhältnis zu der Entwicklung von iPSC-basierten Modellen wurde in **Kapitel 7** ein ausführlicher Übersichtsartikel zu Protokollen zur gerichteten und direkten neuronalen Differenzierung vorgestellt ausgehend von beziehungsweise Stammzellen und somatischen Zellen. Wie in dieser Rezension diskutiers, haben Erkenntnisse aus der Grundlagenforschung und der Entwicklungsbiologie das Design aktueller Protokolle geleitet, und es sind zahlreiche Ansätze zur Ableitung regionalspezifischer glutamaterger, dopaminerger, GABAerger, serotonerger und cholinerger/motor Neuronen verfügbar geworden. Ansätze zur Ableitung jedes dieser krankheitsrelevanten neuronalen Subtypen wurden zusammengefasst, wobei entweder chemisch definierte Systeme unter Verwendung von Musterfaktoren, Transkriptionsfaktor-vermittelte Reprogrammierung und epigenetische Strategien zugrunde gelegt wurden. In

dieser Übersicht wurde ferner hervorgehoben, dass sich zwar unterschiedliche Ansätze als erfolgreich erwiesen haben, um das Schicksal neuronaler Zellen *in vitro* zu steuern, ihre Verfeinerung und Optimierung sowie die Suche nach alternativen Strategien jedoch weiterhin erforderlich sind, um das volle Potenzial der letztendlich abgeleiteten neuronalen Populationen auszuschöpfen. Darüber hinaus sind bestehende Protokolle in der Anzahl der neuronalen Subtypen, deren Induktion vollständig etabliert ist, immer noch begrenzt, und es existieren unterschiedliche Kultivierungsprotokolle für jeden Subtyp. Daher wird erwartet, dass die zukünftige detaillierte Charakterisierung der zellulären und molekularen Eigenschaften, die bei der Steuerung der Stammzelldifferenzierung und der Reprogrammierung somatischer Zellen entlang der neuralen Linie eine Rolle spielen, zur Entwicklung hoch robuster Protokolle beiträgt. Offensichtlich könnten diese Bemühungen zur Entwicklung hoch aussagekräftiger *In-vitro*-Krankheitsmodelle für Erkrankungen wie sAD beitragen.

Schließlich umfasste **Kapitel 8** eine explorative Pilotstudie, in der die Kenntnisse der **Kapitel 6** und **7** kombiniert wurden, um ein Differenzierungsprotokoll für das kortikale Vorderhirn von iPSCs zu etablieren und zu charakterisieren. Das übergeordnete Ziel dieses Kapitels war es, dieses Differenzierungsprotokoll zur Erstellung eines für sAD relevanten In-vitro-Modells zu verwenden. Zunächst wurde der Stamm der iPSC-Linie, die zur neuronalen Differenzierung verwendet wurde, anhand der Expression von Pluripotenzmarkern wie OCT4, NANOG, SSEA3, SSEA4, TRA-1-60 und TRA-1-81 bestätigt. Als nächstes wurden durch die sequentielle Anwendung neuronaler Strukturierungsfaktoren iPSCs zu neuralen Vorläuferzellen (NPCs) und anschließend zu kortikalen Neuronen und Glia differenziert. In jeder Phase des Differenzierungsprozesses wurde die Expression von Transkriptionsfaktoren der Schlüssellinie bewertet, was bestätigte, dass die Zellen in Richtung eines Schicksals des kortikalen Vorderhirns voranschritten. Die erfolgreiche Ableitung von NPCs wurde durch die Expression von NESTIN, SOX2 und PAX6 bestätigt. Die Quantifizierung dieser Marker ergab weiterhin, dass sich die iPSCs in Richtung einer relativ homogenen Population von NPCs (>82%) differenzierten. Die Immunfluoreszenzuntersuchung nach Differenzierung der NPCs bestätigte das Vorhandensein der neuronalen Marker wie MAP2AB, DCX und TUJ1 sowie des Astrozytenmarkers GFAP. Zusätzlich wurde das Vorhandensein von kortikalen Vorderhirnmarkern, einschließlich CTIP2 und FOXG1, in den differenzierten Kulturen identifiziert. Insgesamt bestätigten diese vorläufigen Ergebnisse, dass das in diesem Kapitel beschriebene Protokoll die iPSCs erfolgreich in eine gemischte Population von kortikalen Vorderhirnneuronen und Glia differenzierte. Obwohl eine weitere funktionelle Charakterisierung dieser neuralen Zellen weiterhin erforderlich ist, wurde postuliert, dass das in diesem Kapitel

angewandte Protokoll zur Einrichtung eines kortikalen *In-vitro*-Krankheitsmodellsystems für sAD verwendet werden kann. Um dieses Potenzial weiter zu untersuchen, wurde die Expression von TNXB und des Oxytocinrezeptors (OXTR) in den von iPSC abgeleiteten kortikalen Kulturen bewertet. Wie bereits in **Kapitel 4** dieser Arbeit gezeigt, wurde die epigenetische und genetische Dysregulation von TNXB in mehreren Regionen des sAD-Gehirns mit der AD-Pathophysiologie in Verbindung gebracht. Darüber hinaus wurden die epigenetische Deregulierung des Oxytocin-Gens (*OXT*) und Veränderungen der OXT-Signalübertragung zuvor auch mit sAD- oder assoziierten Krankheitsphänotypen in Verbindung gebracht. Interessanterweise konnte die Expression von TNXB und OXTR bereits nach 14 Tagen in den differenzierten Nervenzellen nachgewiesen und bis zu 56 Tage nach NPC-Differenzierung aufrechterhalten werden, was darauf hindeutet, dass diese kortikalen Vorderhirnzellen ein ansprechendes Modellsystem zur Untersuchung mechanistischer Prozesse im Zusammenhang mit TNXB und OXT-Signalisierung im Kontext von sAD darstellen könnten. Alles in allem sind die in diesem Kapitel unternommenen Anstrengungen unerlässlich, wenn krankheitsrelevante iPSC-basierte Modelle erstellt werden sollen.

Chapter 13

Impact paragraph

In this section, the scientific and societal impact of the research described in this thesis will be discussed.

Scientific impact

In spite of the enormous intellectual and financial investments into sporadic Alzheimer's disease (sAD) research over the last century, there is still considerable debate regarding the underlying causes of the disease and the precise mechanisms behind disease progression. Curing sAD when the disease is already in a more advanced stage is unlikely, as the damage done to the patient's brain seems irreversible at this point. Moreover, current diagnostic tools fall short in detecting the development of sAD in a timely manner that could improve treatment outcome. While early detection and intervention appear to be key in treating sAD, at this moment, even when detected at an early stage, available treatment options solely provide symptomatic relief rather than disease-modifying benefits; hence, gradual development of the full-blown disease phenotype is still inevitable. Thus, there is an unmet scientific need to develop a better understanding of early sAD-associated molecular mechanisms, which in turn could aid in the establishment of therapeutic interventions and diagnostic alternatives.

In recent years, the role of epigenetic mechanisms in sAD has received increasing attention, as they are thought to provide a mechanistic link between environmental exposures and both the development and course of the disease. In fact, the identification of sAD-associated epigenetic profiles is anticipated to provide us with novel cues towards affected molecular mechanisms and interacting environmental factors that could explain the complex underpinnings of the disease. Within this context, efforts aimed at interrogating brain regions with early manifestation of disease processes are of particular importance, as it is thought that the observed molecular changes in these brain areas are more relevant for sAD's etiology and pathophysiology. The scientific interest in epigenetic mechanisms is furthermore instigated by their clinical potential, as they could serve as biomarkers (when targeting *e.g.* blood), allowing earlier detection of sAD and/or provide opportunities for more detailed disease stratification. Their reversible nature furthermore makes them a realistic target for future preventive treatment strategies and pharmacological interventions.

Despite the growing number of epigenetic studies in sAD (reviewed in **Chapter 2**), the field of neuroepigenomics remains a relatively nascent area of investigation. Recent advances in microarray and sequencing technologies show that both whole genome-scale and targeted studies on the epigenome across much larger sample collections are now conceivable. It is important to

acknowledge, however, that a plain exhaustive search for contributing factors, as has been employed in genetic studies of sAD, is unlikely to be fruitful for neuroepigenomic research. In reality, studies aiming to identify epigenetic mechanisms in complex diseases such as sAD need to consider several important issues related to study design, methodological limitations, tissue/cell type-specificity of epigenetic marks, and inference on causality, among others (reviewed in **Chapter 2**). As such, a more systematic approach towards understanding the role of epigenetic mechanisms in sAD is essential. While pushing this field forward, it will be crucial to remain cautious and scrutinize each of these issues thoroughly; next to extensively profiling brain tissues and/or cells derived from independent patient cohorts – if one wishes to provide conclusive insights into the underlying mechanisms of the disease.

The research described in the first part of this thesis builds upon this notion and highlights the scientific impact of the presented studies. More specifically, **Chapter 3** describes an approach for targeted epigenetic profiling allowing one to distinguish between different epigenetic marks, *i.e.* DNA methylation and hydroxymethylation. Many of the existing methods used to interrogate the epigenome in sAD have been unable to specifically discriminate between these two closely related, but functionally distinct, epigenetic marks. As such, previous data obtained from these studies could be confounded and epigenetic effects related to sAD could have been overlooked and/or over- or underestimated. Thus, the work presented in this chapter represents an elegant solution towards overcoming methodological limitations as currently encountered in neuroepigenetic studies. The incorporation of such standardized protocols is therefore crucial for future neuroepigenomic research, even outside the field of sAD, as they will attribute further meaning and validity to the experimental outcomes.

Chapter 4 describes how a second challenge in the field, *i.e.* that of cellular heterogeneity, is being tackled, hence contributing to directing future neuroepigenomic research into more fruitful avenues. It is well-known that epigenetic profiles vary substantially between different cell types, even in the healthy human brain. Hence, a potential problem arises when heterogeneous bulk tissue samples, such as those from the brain, are being used for epigenomic profiling. Epigenetic marks in one cell type may oppose or dilute those in another, potentially obscuring important cell type-specific changes when analyzed all together. This issue is intensified when heterogeneous samples from healthy individuals are compared to those derived from sAD patients at different stages of the disease, which is common practice. Aside from sampling-induced variation between the tissues, differences in cellular composition *e.g.* as a result from

neurodegeneration and increased immune activation as observed in sAD can significantly affect the cellular proportions of different brain samples. Overall, such variation can tremendously affect epigenetic data, resulting in experimental differences that in reality are not attributable to the disease, or, alternatively, masking actual differences. A practical solution to this issue, however, would be to profile individual cells (or cell types) isolated from these heterogeneous tissue samples. For this reason, **Chapter 4** describes a novel approach relying on a combination of laser capture microdissection (LCM) and limiting dilution bisulfite pyrosequencing (LDBSP) that allows for targeted methylation profiling in individually isolated cells. This novel approach or similar strategies using other isolation techniques in combination with LDBSP will be increasingly valuable for future neuroepigenomic studies, even outside the scope of sAD.

Chapter 5 represents the first large-scale epigenetic analysis performed in the brainstem of sAD to date, targeting both the dorsal raphe nuclei (DRN) and locus coeruleus (LC). Here, the state-of-the-art techniques described in the previous chapters are being combined in an effort to obtain more in-depth knowledge on the exact contribution of epigenetic mechanisms in these brain regions affected early in sAD. The scientific impact of this work is therefore threefold. First, this study is among the first of its kind to examine different epigenetic marks in the sAD brain simultaneously, including both DNA methylation and hydroxymethylation. Second, in contrast with previous studies that targeted (primarily cortical) brain regions affected in more advanced stages of sAD, the work in this chapter aimed at identifying potential disease-specific epigenetic marks in the brainstem, which are indicative of the more incipient stages of sAD. Third, the bulk tissue analysis in the DRN described here was complemented by a targeted cell-specific validation study were individual cells isolated from this brainstem region were analyzed. While replication of the obtained findings in independent patient cohorts remains necessary, the identified genes presented here could serve as pillars for future mechanistic studies into sAD. Moreover, this is the first study to date demonstrating that epigenetic signatures within the DRN are strongly dependent upon the cell type analyzed. Overall, the conclusion we can draw from this work is that future studies will need to implement single cell(type) analyses next to targeting heterogeneous bulk tissue samples.

In the second part of the thesis, starting from **Chapter 6**, the focus moves away from epigenetic profiling and, instead, insights into the potential of human induced pluripotent stem cell (iPSC)-based models for sAD research are being offered. An overview of the studies to date, exploring the use of iPSC-based models for sAD research, is presented, which includes but is not limited to neuroepigenomic studies. Furthermore, opportunities, challenges and considerations related to

their use are being addressed. Despite the fact that further detailed characterization and validation of iPSC-based models remains necessary, overall, they are projected to significantly advance our current understanding of many disease processes and to revolutionize approaches for the identification of therapeutics for sAD. In fact, the recent availability of iPSC-based models, cellular reprogramming techniques and directed neural differentiation protocols, which are reviewed in **Chapter 7**, are anticipated to overcome the persistent translational gap between preclinical and clinical studies into sAD, emphasizing their scientific impact. Furthermore, it is anticipated that the establishment of robust iPSC-based *in vitro* models will contribute to reducing the need for animal experimentation.

Chapter 8 describes an exploratory approach on the characterization of a protocol to differentiate iPSCs cortical forebrain cells, with the ultimate aim to establish an sAD-relevant *in vitro* model. Even though the findings should be regarded as preliminary and the iPSC-derived neural cells are in need of further characterization and validation, the availability of such disease-relevant cultures offers appealing opportunities for future sAD studies. In this regard, iPSC-derived neural cells can be applied for modelling complex gene-environment interactions, by *e.g.* exposing cells derived from healthy individuals and sAD patients with different genetic backgrounds to environmental insults. This would allow more detailed studies on the cellular and molecular responses to these insults, even at an epigenetic level, for example. Furthermore, within this framework, the use of epigenome editing tools in these cultures could broaden our understanding on their causal involvement into sAD. All in all, also the iPSC 'adds' to neuroepigenomic research into sAD.

Anticipated societal impact

At this stage, it is important to acknowledge that the research described in this thesis is very fundamental and exploratory in nature. Therefore, the studies presented here will likely not have a direct impact on society in the short run, but mainly offer novel insights into methodological advances and the disease itself that are important for other scientists in the field. In spite of that, and while replication of the epigenetic analysis performed in the brainstem (**Chapter 5**) is crucial, the identified genes could be further investigated for their mechanistic and functional roles in sAD, and, as such, might turn out to be targets for treatment strategies in future studies. All in all, the work presented in this thesis, ranging from systematic studies on the contribution of epigenetic mechanisms in sAD (**Chapters 2**, **3**, **4** and **5**) to the development of clinically relevant human disease models (**Chapters 6**, **7** and **8**), serve as pillars that could maneuver future studies towards novel directions, aiding in the generation of new

knowledge, and eventually, new applications that may benefit the clinic and, as such, society. In essence, expanding our current understanding on factors contributing to disease causation and progression are crucial if one wish to develop a cure for sAD. As such, it will furthermore be crucial to embed the generated data from this thesis within larger efforts that aim to combine multiple molecular layers, including genomics, epigenomics, transcriptomics, proteomics, metabolomics and other data modalities, in order to deepen our understanding of the mechanisms underlying sAD. It is anticipated that such endeavors will allow for better knowledge-driven development of therapeutics that could directly impact on society.

Having that said, it is important to recognize that advances in the management of other common diseases and the improvement in general health, have resulted in an increasingly elderly population so that the prevalence of sAD is expected to increase significantly in the upcoming years. Given the lack of treatments and (early) diagnostics for sAD, the disease is currently acknowledged as one of the most costly to society, with an immense socioeconomic burden. In fact, patients require the support of multiple stakeholders across the healthcare system and social care sectors, as well as from their direct family members and friends, hence affecting not only the patient, but society as a whole. This also means that the cost of care is not only captured by the healthcare systems but also in the personal sector, impacting the quality of life of patients, caregivers and patients' relatives. I therefore strongly believe that even the smallest achievements in fundamental research contribute 'to the greater good' and thus, the work presented in this thesis might hopefully contribute, in the long-term, to alleviating this social and economic burden. Of course, this will not happen in a matter of days, but the pillars that are set by fundamental research, like described in the present thesis, might inspire other researchers and, as such, contribute to the development of novel ideas that could eventually lead to revolutionary scientific breakthroughs.

Chapter 14

Curriculum vitae

Chapter 15

List of publications

Published

Research articles

- De Jesus DF, Orime K, Kaminska D, Kimura T, Basile G, Wang CH, Haertle L, **Riemens RJM**, Brown NK, Hu J, Männistö V, Silva AM, Dirice E, Tseng Y, Haaf T, Pihlajamaki J, Kulkarni RN. *Parental metabolic syndrome epigenetically reprograms offspring hepatic lipid metabolism in mice.* Journal of Clinical Investigation. 2020 Apr 6. pii: 127502.

- Haertle L, Müller T, Lardenoije R, Maierhofer A, Dittrich M, **Riemens RJM**, Stora S, Roche M, Leber M, Riedel-Heller S, Wagner M, Scherer M, Ravel A, Mircher C, Cieuta-Walti C, Durand S, van de Hove DLA, Hoffmann P, Ramirez A, Haaf T, El Hajj N, Mégarbané A. *Methylomic profiling in trisomy 21 identifies cognition- and Alzheimer's disease-related dysregulation.* Clinical Epigenetics. 2019 Dec 16;11(1):195.

- Lippert J, Appenzeller S, Liang R, Sbiera S, Kircher S, Altier B, Nanda I, Weigand I, Gehrig A, Steinhauer S, **Riemens RJM**, Muller CR, Kroiss M, Rost S, Rosenwald A, Fassnacht M, Ronchi CL. *Targeted molecular analysis in adrenocortical carcinomas: a way towards improved personalized prognostication.* Journal of Clinical Endocrinology & Metabolism. 2018 Dec 1;103(12):4511-4523.

Review articles

- **Riemens RJM,** Kenis G, van den Beucken T. *Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease.* Neurobiology of Learning and Memory. 2020 Nov;175:107318.

- **Riemens RJM**, van den Hove DLA, Esteller M, Delgado-Morales R. *Directing neuronal cell fate in vitro: Achievements and challenges.* Progress in Neurobiology. 2018 Apr 10. pii: S0301-0082(17)30182-X.

Perspective articles

- **Riemens RJM***, van den Hove DLA*, Koulousakis P*, Pishva E. *Epigenome-wide association studies in Alzheimer's Disease; Achievements and Challenges.* Brain Pathology. 2020 Sep;30(5):978-983.

Book chapters

- **Riemens RJM**, Soares ES, Esteller M, Delgado-Morales R. *Stem Cell Technology for (Epi)genetic Brain Disorders.* Advances in Experimental Medicine and Biology. 2017;978:443-475.

Abstract publications

- **Riemens RJM**, EPI-AD Consortium. *Alzheimer's disease-associated (hydroxy)-methylomic changes in the brain and blood: Genetics/omics and systems biology*. Alzheimer's & Dementia. Alzheimer's Association International Conference. 2020. 16, e042083.

Submitted articles

- Tiane A, Schepers M, **Riemens RJM**, Rombaut B, Vandormael P, Somers V, Prickaerts J, Hellings N, van den Hove DLA*, Vanmierlo T*. *DNA methylation regulates the expression of the negative transcriptional regulators ID2 and ID4 during OPC differentiation*. Cellular and Molecular Life Sciences. Submitted.

Articles in preparation

- **Riemens RJM**, Kenis G, Mastroeni D, Haaf T, van den Hove DLA. *Targeted detection of unmodified cytosine, 5-methylcytosine and 5-hydroxymethylcytosine levels at single CpG sites by oxidative bisulfite pyrosequencing*. In preparation.

- **Riemens RJM**, Kenis G, Nolz J, Susano Chaves SC, Durous D, Pishva E, Mastroeni D, van Steen K, Haaf T, van den Hove DLA. *Targeted methylation profiling of single laser-capture microdissected post-mortem brain cells by adapted limiting dilution bisulfite pyrosequencing (LDBSP)*. In preparation.

- **Riemens RJM**, Pishva E, Iatrou A, Roubroeks J, Nolz J, Lardenoije R, Ali M, Del Sol A, Delgado-Morales R, Esteller M, Kenis G, Rutten B, Lesch KP, Mill J, Mastroeni D, Ramirez A, Haaf T, Lunnon K, van den Hove DLA. *Brain-region- and cell type-specific epigenetic profiling strongly implicates a role for dysregulation of TNXB and other loci in the brainstem in Alzheimer's disease*. In preparation.

- **Riemens RJM**, Basil K, Akbulut C, Schurgers LJ, van den Hove DLA, Kleinjans JC, van den Beucken T*, Kenis G*. *Establishment and characterization of a human neuronal in vitro model system for Alzheimer's disease using induced pluripotent stem cells: An exploratory approach*. In preparation.

Chapter 16

Statement of individual author contributions

Dissertation Based on Several Published Manuscripts

Statement of individual author contributions and of legal second publication rights

Publication 1 (Chapter 2):

Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges

Renzo J.M. Riemens[1,3]*, Daniël L.A. van den Hove[1,2]*, Philippos Koulousakis[1]*, Ehsan Pishva[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Division of Molecular Psychiatry, Laboratory of Translational Neuroscience, Center of Mental Health, Department of Psychiatry, University of Würzburg, Würzburg, Germany.
[3]Institute of Human Genetics, Julius Maximilians University, Würzburg, Germany.
[4]College of Medicine and Health, University of Exeter Medical School, Exeter University, Exeter, UK.

*Authors contributed equally to this work.

Brain Pathology. 2020 Sep;30(5):978-983.
Doi: 10.1111/bpa.12880.

Participated in	**Author Initials** Responsibility decreasing from left to right				
Study Design Methods Development	N.A.				
Data Collection	N.A.				
Data Analysis and Interpretation	N.A.				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR, DvdH, PK, EP				

Publication 2 (Chapter 3):

Targeted detection of unmodified cytosine, 5-methylcytosine and 5-hydroxymethylcytosine levels at single CpG sites by oxidative bisulfite pyrosequencing

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Diego Mastroeni[3], Thomas Haaf[2], Daniël L.A. van den Hove[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]L.J. Roberts Alzheimer's Disease Center, Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

To be submitted

Participated in	**Author Initials** Responsibility decreasing from left to right				
Study Design Methods Development	RJMR	DvdH	GK	TH	DM
Data Collection	RJMR				
Data Analysis and Interpretation	RJMR	DvdH	GK	TH	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR	DvdH	GK	TH	

Publication 3 (Chapter 4):

Targeted methylation profiling of single laser-capture microdissected post-mortem brain cells by adapted limiting dilution bisulfite pyrosequencing (LDBSP)

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Jennifer Nolz[3], Sonia C. Susano Chaves[1], Diane Duroux[4], Ehsan Pishva[1], Diego Mastroeni[3], Kristel van Steen[4,5], Thomas Haaf[2], Daniël L.A. van den Hove[1,6]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[4]WELBIO, GIGA-R Medical Genomics - BIO3, University of Liège, Liège, Belgium; Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[5]Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[6]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

To be submitted

Participated in	**Author Initials** Responsibility decreasing from left to right				
Study Design Methods Development	TH, DvdH	RJMR	GK	JN, DM	EP
Data Collection	RJMR				
Data Analysis and Interpretation	RJMR	DvdH	GK, SCS	KvS, DD	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR	DvdH	GK	TH	Remaining authors

Publication 4 (Chapter 5):

Brain-region- and cell type-specific epigenetic profiling strongly implicates a role for dysregulation of *TNXB* and other loci in the brainstem in Alzheimer's disease

Renzo J.M. Riemens[1,2]*, Ehsan Pishva[1,3]*, Artemis Iatrou[4], Janou Roubroeks[3], Jennifer Nolz[5], Roy Lardenoije[1,6], Muhammad Ali[1,7], Antonio Del Sol[7], Raul Delgado-Morales[1,8], Manel Esteller Badosa[9], Gunter Kenis[1], Bart Rutten[1], Klaus-Peter Lesch[10], Jonathan Mill[3], Diego Mastroeni[1,5,11], Alfredo Ramirez[12], Thomas Haaf T[2], Katie Lunnon[3]†, Daniël L.A. van den Hove[1,10]†

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]University of Exeter Medical School, University of Exeter, Exeter, United Kingdom.
[4]Rush Alzheimer's Neurodisease Center, Rush University Medical Center, Chicago, Illinois, United States of America.
[5]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[6]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany.
[7]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg.
[8]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute - IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[9]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain; Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Madrid, Spain; Institucio Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain.
[10]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[11]L.J. Roberts Center for Alzheimer's Research Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[12]Division for Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Medical Faculty, University of Cologne, Cologne, Germany; Department for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, Bonn, Germany.

* Authors contributed equally to this work
† Authors contributed equally to this work

To be submitted

Participated in	**Author Initials** Responsibility decreasing from left to right			
Study Design Methods Development	DvdH, KL	TH	RJMR, EP	Remaining authors

Data Collection	RJMR, EP			
Data Analysis and Interpretation	RJMR, EP	DvdH	KL	Remaining authors
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR, EP	DvdH	KL	Remaining authors

Publication 5 (Chapter 6):

Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Twan van den Beucken[3]

[1]Department of Psychiatry & Neuropsychology, Graduate School MHeNS (School for Mental Health and Neuroscience), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilian University, Wuerzburg, Germany.
[3]Department of Toxicogenomics, Graduate School GROW (Research School for Oncology and Developmental Biology), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.

Neurobiology of Learning and Memory. 2020 Sep 22;107318.
Doi: 10.1016/j.nlm.2020.107318.

Participated in	**Author Initials** Responsibility decreasing from left to right	
Study Design Methods Development	N.A.	
Data Collection	N.A.	
Data Analysis and Interpretation	N.A.	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR, TvdB	GK

Publication 6 (Chapter 7):

Directing neuronal cell fate in vitro: achievements and challenges

Renzo J.M. Riemens[1,2,3], Daniël L.A. van den Hove[3,4] Manel Esteller[1,5,6], Raul Delgado-Morales[1,3]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute – IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain.
[6]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

Prog Neurobiol. 2018 Sep;168:42-68.
Doi: 10.1016/j.pneurobio.2018.04.003.

Participated in	**Author Initials** Responsibility decreasing from left to right			
Study Design Methods Development	N.A.			
Data Collection	N.A.			
Data Analysis and Interpretation	N.A.			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR	RDM	DvdH	ME

Publication 7 (Chapter 8):

Establishment and characterization of a human neuronal *in vitro* model system for Alzheimer's disease using induced pluripotent stem cells: An exploratory approach

Renzo J.M. Riemens[1,2], Katherine Basil[1], Cengiz Akbulut[3], Leon J. Schurgers[3], Daniël L.A. van den Hove[1,4], Jos C. Kleinjans[5], Twan van den Beucken[5†], Gunter Kenis[1†]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Biochemistry, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, the Netherlands.

† Authors contributed equally to this work

To be submitted

Participated in	**Author Initials** Responsibility decreasing from left to right			
Study Design Methods Development	RJMR, KB, CA	GK	TvdB	Remaining authors
Data Collection	RJMR			
Data Analysis and Interpretation	RJMR	GK	TvdB	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	RJMR	GK	TvdB	Remaining authors

The doctoral researcher confirms that she/he has obtained permission from both the publishers and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above-mentioned assessment.

_____	_____	_____
Doctoral Researcher's Name	Place, Date	Signature
_____	_____	_____
Primary Supervisor's Name	Place, Date	Signature

Dissertation Based on Several Published Manuscripts

Statement of individual author contributions to figures/tables/chapters included in the manuscripts

Publication 1 (Chapter 2):

Epigenome-wide association studies in Alzheimer's disease; Achievements and challenges

Renzo J.M. Riemens[1,3]*, Daniël L.A. van den Hove[1,2]*, Philippos Koulousakis[1]*, Ehsan Pishva[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Division of Molecular Psychiatry, Laboratory of Translational Neuroscience, Center of Mental Health, Department of Psychiatry, University of Würzburg, Würzburg, Germany.
[3]Institute of Human Genetics, Julius Maximilians University, Würzburg, Germany.
[4]College of Medicine and Health, University of Exeter Medical School, Exeter University, Exeter, UK.

*Authors contributed equally to this work.

Brain Pathology. 2020 Sep;30(5):978-983.
Doi: 10.1111/bpa.12880.

Table(s)	**Author Initials** Responsibility decreasing from left to right			
Table 1	PK	RJMR	DvdH	EP

Publication 2 (Chapter 3):

Targeted detection of unmodified cytosine, 5-methylcytosine and 5-hydroxymethylcytosine levels at single CpG sites by oxidative bisulfite pyrosequencing

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Diego Mastroeni[3], Thomas Haaf[2], Daniël L.A. van den Hove[1,4]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]L.J. Roberts Alzheimer's Disease Center, Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

To be submitted

Figure(s)	**Author Initials** Responsibility decreasing from left to right		
Figure 1	RJMR	DvdH	GK
Figure 2	RJMR	DvdH	GK
Figure 3	RJMR	DvdH	GK
Table(s)	**Author Initials** Responsibility decreasing from left to right		
Table 1	RJMR	DvdH	GK
Table 2	RJMR	DvdH	GK
Table 3	RJMR	DvdH	GK
Table 4	RJMR	DvdH	GK
Table 5	RJMR	DvdH	GK
Supplementary Table 1	RJMR	DvdH	GK
Supplementary Table 2	RJMR	DvdH	GK
Supplementary Table 3	RJMR	DvdH	GK

Publication 3 (Chapter 4):

Targeted methylation profiling of single laser-capture microdissected post-mortem brain cells by adapted limiting dilution bisulfite pyrosequencing (LDBSP)

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Jennifer Nolz[3], Sonia C. Susano Chaves[1], Diane Duroux[4], Ehsan Pishva[1], Diego Mastroeni[3], Kristel van Steen[4,5], Thomas Haaf[2], Daniël L.A. van den Hove[1,6]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[4]WELBIO, GIGA-R Medical Genomics - BIO3, University of Liège, Liège, Belgium; Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[5]Department of Human Genetics - Systems Medicine, University of Leuven, Leuven, Belgium.
[6]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

To be submitted

Figure(s)	**Author Initials** Responsibility decreasing from left to right		
Figure 1	RJMR	DvdH	GK
Figure 2	RJMR	DvdH	GK
Table(s)	**Author Initials** Responsibility decreasing from left to right		
Table 1	RJMR	DvdH	GK
Table 2	RJMR	DvdH	GK
Supplementary Table 1	RJMR	DvdH	GK
Supplementary Table 2	RJMR	DvdH	GK

Publication 4 (Chapter 5):

Brain-region- and cell type-specific epigenetic profiling strongly implicates a role for dysregulation of *TNXB* and other loci in the brainstem in Alzheimer's disease

Renzo J.M. Riemens[1,2]*, Ehsan Pishva[1,3]*, Artemis Iatrou[4], Janou Roubroeks[3], Jennifer Nolz[5], Roy Lardenoije[1,6], Muhammad Ali[1,7], Antonio Del Sol[7], Raul Delgado-Morales[1,8], Manel Esteller Badosa[9], Gunter Kenis[1], Bart Rutten[1], Klaus-Peter Lesch[10], Jonathan Mill[3], Diego Mastroeni[1,5,11], Alfredo Ramirez[12], Thomas Haaf T[2], Katie Lunnon[3†], Daniël L.A. van den Hove[1,10†]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]University of Exeter Medical School, University of Exeter, Exeter, United Kingdom.
[4]Rush Alzheimer's Neurodisease Center, Rush University Medical Center, Chicago, Illinois, United States of America.
[5]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, Arizona, United States of America.
[6]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany.
[7]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg.
[8]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute - IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[9]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain; Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Madrid, Spain; Institucio Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain.
[10]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[11]L.J. Roberts Center for Alzheimer's Research Banner Sun Health Research Institute, Sun City, Arizona, United States of America.
[12]Division for Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Medical Faculty, University of Cologne, Cologne, Germany; Department for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, Bonn, Germany.

* Authors contributed equally to this work
† Authors contributed equally to this work

To be submitted

Figure(s)	**Author Initials** Responsibility decreasing from left to right	
Figure 1	RJMR, EP	DvdH
Figure 2	RJMR	DvdH
Supplementary Figure 1	RJMR, EP	DvdH

Table(s)	**Author Initials** Responsibility decreasing from left to right	
Table 1	RJMR, EP	DvdH
Table 2	RJMR, EP	DvdH
Table 3	RJMR, EP	DvdH
Table 4	RJMR, EP	DvdH
Table 5	RJMR, EP	DvdH
Supplementary Table 1	RJMR, EP	DvdH
Supplementary Table 2	RJMR, EP	DvdH
Supplementary Table 3	RJMR, EP	DvdH
Supplementary Table 4	RJMR, EP	DvdH
Supplementary Table 5	RJMR, EP	DvdH
Supplementary Table 6	RJMR, EP	DvdH
Supplementary Table 7	RJMR, EP	DvdH
Supplementary Table 8	RJMR, EP	DvdH
Supplementary Table 9	RJMR, EP	DvdH
Supplementary Table 10	RJMR, EP	DvdH
Supplementary Table 11	RJMR, EP	DvdH
Supplementary Table 12	RJMR, EP	DvdH
Supplementary Table 13	RJMR	DvdH
Supplementary Table 14	RJMR	DvdH

Publication 5 (Chapter 6):

Human-induced pluripotent stem cells as a model for studying sporadic Alzheimer's disease

Renzo J.M. Riemens[1,2], Gunter Kenis[1], Twan van den Beucken[3]

[1]Department of Psychiatry & Neuropsychology, Graduate School MHeNS (School for Mental Health and Neuroscience), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilian University, Wuerzburg, Germany.
[3]Department of Toxicogenomics, Graduate School GROW (Research School for Oncology and Developmental Biology), allocated with the Faculty Health Medicine and Life Sciences of Maastricht University, Maastricht, the Netherlands.

Neurobiology of Learning and Memory. 2020 Sep 22;107318.
Doi: 10.1016/j.nlm.2020.107318.

Figure(s)	**Author Initials** Responsibility decreasing from left to right
Figure 1	TvdB
Table(s)	**Author Initials** Responsibility decreasing from left to right
Table 1	RJMR, TvdB

Publication 6 (Chapter 7):

Directing neuronal cell fate in vitro: achievements and challenges

Renzo J.M. Riemens[1,2,3], Daniël L.A. van den Hove[3,4] Manel Esteller[1,5,6], Raul Delgado-Morales[1,3]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute – IDIBELL, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain.
[6]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

Prog Neurobiol. 2018 Sep;168:42-68.
Doi: 10.1016/j.pneurobio.2018.04.003.

Figure(s)	**Author Initials** Responsibility decreasing from left to right	
Figure 1	RJMR	DM, DvdH
Figure 2	RJMR	DM, DvdH
Figure 3	RJMR	DM, DvdH
Figure 4	RJMR	DM, DvdH
Table(s)	**Author Initials** Responsibility decreasing from left to right	
Table 1	RJMR	DM, DvdH
Table 2	RJMR	DM, DvdH
Table 3	RJMR	DM, DvdH
Supplementary Table 1	RJMR	DM, DvdH
Supplementary Table 2	RJMR	DM, DvdH
Supplementary Table 3	RJMR	DM, DvdH

Publication 7 (Chapter 8):

Establishment and characterization of a human neuronal *in vitro* model system for Alzheimer's disease using induced pluripotent stem cells: An exploratory approach

Renzo J.M. Riemens[1,2], Katherine Basil[1], Cengiz Akbulut[3], Leon J. Schurgers[3], Daniël L.A. van den Hove[1,4], Jos C. Kleinjans[5], Twan van den Beucken[5†], Gunter Kenis[1†]

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, the Netherlands.
[2]Institute of Human Genetics, Julius Maximilians University, Wuerzburg, Germany.
[3]Department of Biochemistry, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, the Netherlands.
[4]Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.
[5]Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, the Netherlands.

† Authors contributed equally to this work

To be submitted

Figure(s)	**Author Initials** Responsibility decreasing from left to right	
Figure 1	RJMR	GK
Figure 2	RJMR	GK
Figure 3	RJMR	GK
Figure 4	RJMR	GK
Figure 5	RJMR	GK
Table(s)	**Author Initials** Responsibility decreasing from left to right	
Supplementary Table 1	RJMR	GK

I also confirm my primary supervisor’s acceptance.

_____	_____	_____
Doctoral Researcher’s Name	Place, Date	Signature

Chapter 17

Affidavit

Affidavit

I hereby confirm that my thesis entitled "*Neuroepigenomics in Alzheimer's disease: The single cell ADds*" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

_____ _____

Place, Date *Signature*

Beëdigde verklaring

Hierbij verklaar ik dat mijn proefschrift getiteld *"Neuroepigenomics in Alzheimer's disease: The single cell ADds"* het resultaat is van mijn eigen werk. Ik heb geen hulp of ondersteuning gekregen van commerciële consulenten. Alle toegepaste bronnen en/of materialen zijn beschreven en gespecificeerd in het proefschrift.

Verder bevestig ik dat dit proefschrift nog niet is ingediend als onderdeel van een ander examenproces, noch in identieke noch in vergelijkbare vorm.

_____ _____

Plaats, Datum *Handtekening*

Eidestattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation *"Neuroepigenomics in Alzheimer's disease: The single cell ADds"* eigneständig, das heißt insbesondere selbständig und ohne Hilfe einse kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen un Hilfsmittel verwended zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem andere Prüfungsverfahren vorgelegen hat.

_____ _____

Ort, Datum *Unterschrift*

Chapter 18

Acknowledgements

I would not be where I am today without the help of many – both during my time as a PhD student and before that. Looking back, I have been very fortunate to always be surrounded by and to come across with amazing people, who all have been playing their parts in me becoming the researcher, and perhaps even more important, the person that I am today. Therefore, for anyone, who stood next to me and/or who contributed to my work in any form or by any means (even when not explicitly mentioned in this section), I want you to know that I am extremely grateful, and I want to thank you for your support, your encouragement and for the joyful moments or pain that we shared together during these years.

Having that said, I would like to take the opportunity to thank and to express my sincerest gratitude to a number of people in particular. First and foremost, I would like to thank my promotion team, who made it possible for me to embark on this joint-PhD adventure and who gave me the liberty to develop myself as an academic researcher. It has been a real pleasure and a truly enriching experience to conduct this joint-PhD programme between Maastricht University and the Julius-Maximilians-University.

Daniël, from the very first moment we met during my Bachelor up until the very end of my PhD you have always encouraged me to get the best out of myself and to “go that extra (s)mile”! Honestly, you have been a true inspiration and my role model all along. Your “enthusiasm for neuroscience is contagious” and a strong drive for me to try to “be a carrier”. It cannot be stressed enough that you are the perfect example of a supervisor one could wish for, and that the way you treat the people under your supervision should become the golden standard for every other staff member within (and even outside of) academia. I would like to thank you for believing in me from the very beginning, for trusting in my abilities, for guiding me, for always being there when necessary, for providing a listening ear, for being honest, for being critical, for taking care, for being patient, for convincing me that “things well be all right” and for sharing your thoughts on any matter. I am more than grateful for all the time and effort you have invested in me. Doing a PhD under your wings has been truly EPIC and I am really looking forward to my time as a post-doc in your NeuroEPICgenetics group! After all these years, I think we can agree: “science is fun” indeed!

Bart, thank you for believing in me and supporting me from the start, even before I embarked on this joint-PhD adventure. My time abroad did not allow us to meet on a frequent basis during my time as PhD student, but when we did, you always had some valuable input for my research. I think it is safe to say that without your enthusiasm and faith in me I would not be where I am today, for which I want to express my sincerest gratitude.

Gunter, you are really one of the most caring, knowledgeable and intelligent people that I know. I really look up to you and I have learned so much from you over the years! Thank you for supervising me, sharing your expertise with me, for supporting me and for having faith in me. Your contributions and insights into my research have been truly invaluable. If it had not been for you, Daniël and I would still be trying to figure out how to estimate alleles in a single reaction and my PhD would have never come to an end. Our weekly meetings, which were not only used to catch up on work, but also on life, have been strongly motivating me to keep on pushing my boundaries week after week.

Professor Haaf, my sincerest appreciation for guiding me and for providing me with such freedom in your laboratory. Your dedication to science and hard-working attitude are truly inspiring. I will always remember your visits in my office that started every time with the question “Und?” and ended with the phrase “Na ja, gut!”, which reflected your generous interest in my developing projects and your cautious optimism for the next milestone(s) to be achieved. Aside from the brief updates from my side, our chats in-between these phrases were filled with in-depth discussions on both science and on how to developing a successful career, which really have been eye-opening – so thank you for that!

Professor Förster, I still vividly remember the day I came to your office to ask whether you would be willing to join my supervisor committee in Wuerzburg. After our inspiring chat on doing translational research and the role of epigenetics in human brain disorders, you answered this question with a yes without hesitation, for which I am truly grateful. I specifically want to thank you for the time that you have invested in me and for joining me in this journey as a PhD student.

Professor Lesch, It has been a great honor and pleasure to be able to work with you and your group. The experience in cultivating and working with human stem cells that I gained during the time in your laboratory has significantly contributed to my skillset, from which I still highly benefit today. It has been a truly enriching experience for which I want to express my sincerest gratitude.

A special mention and thank you also goes out to the (other) members of my assessment committee.

Jos, my whole journey as a PhD student would not have been so successful if it had not been for you and Daniël to accept me in the Fundamental Neuroscience (FN) programme. I strongly believe that being part of FN has significantly contributed to my professional and personal development, both during my time as a research master student, as well as during my PhD. I would therefore like to

thank you for encouraging me from the start and teaching me along the way. It is also a great pleasure to be able to end my PhD together with you and to have you as the chair of my assessment committee. I genuinely hope I can keep on surprising you in the future!

Tim, I have gotten to know you as an amazing scientist with a great sense of dry humor to match. Your elaborated knowledge on a wide variety of different laboratory techniques is truly inspiring, and your witty, unexpected jokes during our labmeetings are hilarious. You definitely know how to keep the perfect balance between a serious working attitude and having fun at work. Thank you for all your invaluable advice on my research and for assisting me with my Pyrosequencing experiments in Hasselt. To be able to work with you and your group in BIOMED has been very rewarding and a great pleasure!

I am equally thankful to **Prof. Dr. Paul Lucassen**, **Prof. Dr. Geraldine Zimmer-Bensch** and **Dr. Angelika Schmitt-Böhrer** for being on my assessment committee. Thank you for taking the time to read and to assess my thesis.

I would also like to express my special thanks to **Prof. Dr. Manel Esteller**. Thank you for granting me the opportunity to work in your laboratory. The expertise that I acquired during my time in IDIBELL has contributed significantly to develop myself as a researcher and allowed me to excel during my time as a PhD student, for which I am really grateful.

A warm and special thank you also goes to **Raul**. It was a great pleasure to be able to get to know you and to be able to work with you! You were not only a great supervisor during my time in IDIBELL, but also an amazing friend. Really, you are the best! I hope my questions on laboratory- and science-related matters are not haunting you in your dreams up until this day. Thank you for your patience, for guiding me, for teaching me how to become more confident with my work and for showing me how to be more independent in the laboratory. We will meet again! I am sure!

I would also like to express my sincerest gratitude to **Jos**, **Twan**, **Marcel** and everybody else within BReIN or at the Department of Toxicogenomics at Maastricht University that has (in)directly contributed to my work. Thank you for involving me and granting me the opportunity to work together with you on this project.

I am also extremely grateful to **Gabriele** and **Nicole** for organizing all the legal matters in relation to this joint-PhD degree, as well as to **Rachelle** for organizing all the administration of my PhD thesis in Maastricht. You did an amazing job!

Then I want to thank my paranymphs for all their support over the years and for their help with my thesis defense.

Dean, we have been partners in crime (both literally and as a figure of speech), from the early days at the university up until today, where we both are about to obtain our PhD degrees soon. I would like to thank you for being there as a great friend and for sharing all these memorable moments together with me during these past years, both inside and outside of academia. From brushing our teeth with coffee early in the morning, just before rush-hour was about to start to the university, to our “hotbox sessions” until late at night, where we played loud electronic music in the university library (read worked our asses off). From changing lightbulbs with lemons during our evenings going to the “hippos”, to drinking beers at the Ramblas, to performing as DJs in Wuerzburg, and so on... It has been absolutely amazing and the “schweinerei” definitely kept me going even during the toughest moments where there was little time to sleep and only time to work.

Philippos, aside from being a Greek, a German and a Dutch friend, a college and my office-mate, you have also fulfilled the role as my psychologist during the last (crucial) phase of my PhD. I would therefore like to thank you for simply being there for me, for the free therapy sessions, and for providing me with your (well-appreciated) constructive criticism. Yes, when reading this you can finally say: “Er hat schließlich seine Doktorarbeit eingereicht. Wer hätte das erwartet?”. I am very happy we will remain colleagues for a while, but be aware that everything you say can be used against you and that you still have to defend your PhD thesis one day. Of note, your amazing dance moves, your passion for dressing up as a medieval soldier and your love for watching gaming streams (even during working hours) have not gone unnoticed. Overall, it has been a blast of Oxytocin getting to know you and irrespective of the results that you will require in the remaining years of your PhD, I am convinced that our friendship has reduced my risk of developing Alzheimer’s disease in the future.

A massive thank you also goes to the Neuroepigenetics group at the Department of Psychiatry and Neuropsychology at Maastricht University. Thank you for sharing all your expertise with me, for your support in the laboratory, for your (career) advice, for the pizza meetings (read journal clubs), for the coffee-

moments, for all the enjoyable (online) gatherings outside of the laboratory and simply for the lovely chats in the hallway.

Ehsan, a special thank you to you! Thanks for guiding me and working together with from the start of my Bachelor internship up until the very end of my PhD, for your straight-to-the-point advice and for always being there in case I was in desperate need for bioinformatics expertise. You really can breakdown bioinformatics to a level where even I understand it. A wise man once told me (read Gunter): “You need to be an expert in order to be able to explain complicated things in a very simple way”. This definitely also applies to you and it is something I truly admire.

Chris, “lunch?”. As you can see, I kept my promise to give you your own paragraph in my acknowledgements section. Thank you for being such a caring, funny and lovely office-mate, as well as for being a true friend and a “Pyro Ranger”. Your personality, your sense of humor and the weird facts that you bring to the office every day, significantly ($p < 0.001$) contribute to the positive working atmosphere. Perhaps this is also one of the reasons why I do not want to leave our office yet – so please stop asking.

Katherine, aside from being a great researcher you have also proven to be an exceptional neuroethicist, something that is ever more important today and truly admirable. I want to thank you for sharing all your stem cell-expertise with me and for assisting me in setting up my iPSC experiments in Maastricht.

Clara, I am still not sure whether I should talk to you in Dutch or in English or maybe use a combination of both – if that ever makes sense. I want to thank you for sharing all the ins and outs of writing a PhD thesis with me. This has been extremely helpful. I wish you all the best of luck during your time as a post-doc in Boston, I am sure you will do an amazing job!

Roy, **Janou**, **Artemis** and **Nicole**, we met each other for the first time even before my PhD and our paths have crossed multiple times over the years; during research electives, during (online) meetings, at conferences, in the office, in the laboratory or during the EPI-AD meetings for example. I would like to thank you all for taking me by the hand and showing me around in any of these occasions!

Last, but not least: **Lars**, **Rick**, **Ali**, **David**, **Manon**, **Martin**, **Alix**, **Ghazi** and **Laurence**, you have been such amazing colleagues! I would like to thank you for all the useful comments and advice on my research, presentations, laboratory

issues, talks and so on. The brainstorms we had during our labmeetings were always very fruitful and rewarding. I has been a pleasure to work with you.

A special mention and thank you also goes to the technician team at the Department of Psychiatry and Neuropsychology. **Hellen**, **Denise**, **Sandra**, **Barbie**, **Wouter** and **Theodora**, you were always readily available to share your expertise, to provide me with materials or equipment, and to assist me in my experiments whenever needed. Your efforts have really helped me to develop my skillset over the years, so many thanks for that!

I also would like to thank all my other colleagues, PhD students, post-docs and senior scientists (past and present) at the Department of Psychiatry and Neuropsychology for the nice working atmosphere and for making my time as a PhD student in Maastricht memorable: **Ellis**, **Glenn**, **Christian**, **Roman**, **Jeroen**, **Shengua**, **Tanya**, **Faisal**, **Caterina**, **Jackson**, **An, Qian**, **Aryo**, **Mathijs**, **Sarah**, **Rose**, **Perla**, **Sylvana**, **Christian**, **Jana**, **Daan**, **Roel**, **Marina**, **Maarten**, **Ralph**, **Nikita**, **Nynke**, **Faris**, **Martijn**, **Thomas**, **Wenting**, **Margot**, **Lonne**, **Wouter**, **Stijn**, **Gowoon**, **David**, **Tom**, **Marie-Thérèse**, **Ankie**, **Damaris**, **Mario**, **Harry**, **Pilar**, **Ali**, **Govert** and everybody else that I might have forgotten.

I also want to thank my colleagues from BIOMED, who made my research stay in Hasselt extremely enjoyable and a great success.

Assia, it was a heroic battle, but we managed to tame the beast and the first articles are either in submission or close to being published. I really think we can pat each other on the back: “Go go Pyro Rangers!” Yes, with the beast, I mean the Pyrosequencer Q48 Autoprep, the monster that appeared at the end of our quest from BIOMED to Hasselt University, and the only thing that stood in-between us and our data. I would like to thank you for fighting this battle with me, and more importantly, for winning it together! Your enthusiasm, positive- and hard-working attitude definitely play an important part in this victory. Like Tim said: We are true “Pyro-neers”!

Melissa and **Ben**, I have never felt so welcomed and directly integrated into a new laboratory as compared to when I arrived for the first time to BIOMED. Mainly thanks to you! Melissa, everything you touch in the laboratory seems to turn into gold. Know that the next time when I am in desperate need for earthshattering data, you will be one of the first people I will contact to collaborate with. Ben, if you are ever in doubt whether becoming a scientist is your true ambition; perhaps you might consider finding a way to make money out of “memes”. For whoever reads this: All credits and copyrights of the “Pyro Rangers” go to Ben.

Of course, I also want to thank all my friends and colleagues from AG Haaf at the Institute of Human Genetics in Wuerzburg.

Diese **Felix** und Diese **Alex**! I think we can conclude that a whole lot of Greek food, half-liter beers and Ouzo are the perfect combination to achieve the highest level of productivity. I would like to thank you both for these memorable moments in-between our hard working days, for teaching me the ins and outs of limiting dilution bisulfite pyrosequencing (Felix) and for the help with the ICT-related issues I was facing in Wuerzburg (Alex) – AZERTY keyboards and the German version of Windows are simply not my favorite. Felix, if you might ever wonder what to buy me as a gift for obtaining my PhD degree: I am still in desperate need for a CD of “Die Flippers”. We should really start planning our trip to Amsterdam, invite Alex and “not limit our dilutions” at another Greek or Italian restaurant.

Nady, thank you for helping me out in the laboratory, for being approachable for any science-related question and for being my first go-to lunch mate during my time in Wuerzburg. The way you are developing your career is truly inspiring! No doubt, you will succeed! I like to believe that you are currently driving your Ferrari, Maserati, Rolls Royce or any other supercar to the laboratory in Qatar. Please do not forget that I am still waiting for the invitation to come over and drive these supercars ones together. Perhaps we can directly plan another lunch at a fancy restaurant and talk some more science.

Larissa, you were literally just one door away in case I was in need for advice, might it be anything work-related or even life in general. You are really one of the most open and approachable people that I met during my time in Wuerzburg with whom I could share anything, even the struggles of learning a new language like Spanish. Thank you for always being there whenever necessary!

Tabea, I do not know whether I should start with thanking you or with providing you with my apologies for involving you in all these complicated Germany-Netherlands-Belgium constructions and for making you ship materials from Wuerzburg to all over Europe – let us hope the concerned companies will never read this. I guess I owe you both! It cannot be emphasized enough that you really did an amazing job!

Ruth, a big thank you to you for always being available, for saving my day so many times and for making my life in Wuerzburg much easier! Your kindness, generosity, and willingness to help others stands as a shining example for every secretary. From helping me to find a place to stay in Wuerzburg, to making sure

I also could keep my place, to assisting me with filling-out legal documents, to organizing grant money and bills for my projects, to rescuing my e-mail account and so on; Not a single job was too big, for which I want to express my sincerest gratitude!

Indrajit, Caroline, **Michaela**, **Ramya**, **Anna, Julia, Julia, Julia, Paulina, Barbara** and **Laura**, I really enjoyed spending time and working together with you. Thank you for being such amazing colleagues, for introducing me to the German Biergarten, for the Weinprobe (which really could not be considered as a “tasting”) and for showing me the Weinfeste in Wuerzburg. It was a whole lot of fun!

A special mention also goes to **Daniel, Reinhard, Eva, Marcus, Tobias, Tamara, Julianne** and everybody else from the Institute of Human Genetics that have worked with me. Thank you for your assistance and invaluable advice during these years!

Then I would like to thank everybody from the Laboratory of Translational Neuroscience in Wuerzburg, and specifically:

Johanna, our paths in Barcelona just missed each other and when you came back to Wuerzburg, I was just about to leave and move to Maastricht. In spite of that, it always felt like we have known each other for years. Thank you for being such an extremely joyful and kind person, for introducing me to Kiliani (after my time in Wuerzburg – I clearly did something wrong there), for helping me out in the laboratory and for providing me a place to sleep whenever necessary. Know that you are always welcome and that you always will have a place to stay here with us!

Charline, Maria, and **Julia,** thank you for showing me the ins and outs of working with stem cells during my time in Wuerzburg. I have learned a lot from you and I am very grateful for the time and effort that you have invested in teaching me.

I also would like to thank and to express my appreciation to everybody from PEBC1 and PEBC2 at IDIBELL during my time in Barcelona.

Edilene, my favorite Brazilian! A special mention and thank you also goes to you. The energy and enthusiasm that radiates from you is truly admirable and unique! You are such an amazing and kind person. Thank you for being so patient with me and for always being there for me during my time in Barcelona. And, of course, for taking care of my ReN(zo)cells VM, whenever I decided to travel back to the

Netherlands – I bet people are still trying to figure out whether we came up with this name ourselves. It really was a blast getting to know you and I truly hope our paths will cross once again soon!

Davide, who would ever expect that it is possible to obtain a PhD degree by just putting water into water the entire day? If you are still in doubt, please find proof of that here. Thank you for being a great friend and for being the most stereotypical Italian person in existence – Let's face it: During our chats, your hands and fingers were always doing the talking. Our time together in Barcelona was really too short! We should meet up once again soon to catch-up on life while enjoying a game of "Exploding Kittens" and your mother's home-made Limoncello.

Sònia, **David**, **Catia**, **Cristina**, **Humberto**, **Raquel**, **Fer**, **Pere**, **Miguel**, **Aida**, **Sebas**, **Carles**, **Maxime**, **Holger**, **Manu**, **Laia**, **Carmen**, **Anna**, **Marta**, **Rute** and **Olga**. Thanks for making me feel home regardless of my very poor Spanish or Catalan back then, for guiding me, for sharing all your expertise and for the amazing gatherings outside of the laboratory. My time in Barcelona has really been one of the highlight during my time as a PhD- and as a research master student. Some of the Spanish habits that I acquired during my time with you, such as the time and size of meals for example, will be forever engrained in my daily routines. Thank you for the amazing memories! I look back with a lot of joy and a smile on my face!

I am also grateful to my wonderful Bachelor and Master students: **Dasha**, **Natalie, Julianne**, **Manas**, **Sonia** and **Ziva**. Thank you not only for your contribution to my projects, but also for everything that I have learned from you. I hope that you all enjoyed your internships and that you learned as much from me as I learned from you! Please keep me up-to-date about all your future endeavors; I am sure you will all achieve great success!

Of course, I also would like to thank everybody outside of academia and my work, and specifically my closest friends and family.

Jonas, **Koen**, **Tijs**, **Fieneke**, **Niels**, **Wouter**, **Chaira, Tom**, **Marlon**, **Joeri**, **Sam**, **Brian** and **Thomas**, you really helped me to find the right balance between work and life, and were always there so I could take my mind off work and clear my head. I want to thank you for being part of my life, for all the joyful moments during parties, gatherings, festivals or visits, and for the post-cards or messages that I received over the years while being in the distance – This really kept me going.

This also applies to **Michiel**. We started our studies at the same time in Maastricht and soon we both will be doctors – Although you are the only one that can really treat patients, but that is just a minor detail. I think we can be proud! Who would have ever expected that?

Then fom the bottom of my heart, I am extremely grateful to my parents, **Willem** and **Ingeborg,** and to my brother, **Silvano**. I am so blessed to have you in my life. I want to thank you deeply for all what you have done for me, for your endless support and encouragement, for all the sacrifices that you made, for helping me move around all the time, for all the guidance that you gave me and for molding me into the person that I have become today. Without you I would not have made it this far! Really!

Anaís, you have been my solid rock, my other better half and the person that has always been there for me during these years. You are the one who knows me best, both during good and during bad times, the one that cheers me up and the one that picks me up to put me back on my feet whenever necessary. Thanks for helping me to put things in perspective, for believing in me, for reminding me of my own capabilities and for simply being the person who you are. I am really looking forward to the time and adventures that lay ahead of us!

Finally, I would like take a moment to acknowledge my grand- and godparents, **Jan** and **Maria**, who have always been a true inspiration and a strong drive behind my success. Your memory will be forever with me.