ADHD Genetics in Mouse and Man

ADHS Genetik bei Maus und Mensch

Genética del TDAH en ratón y hombre

DISSERTATION

to obtain the degree of Ph.D. (Doctor of Philosophy) at Julius-Maximilians-Universität Würzburg on the authority of the

President Magnificus Prof. Dr. Alfred Forchel in accordance with the decision of

the Common Graduation Commission of the Graduate School of Life Sciences, Julius-Maximilians-Universität Würzburg, Germany

and

to obtain the degree of Doctor by the Universitat de Barcelona in the doctoral programme on the authority of the

Rector Magnificus, Dr. M. Joan Elias García in accordance with the decision of the Doctoral Commission of the Faculty of Biology, Universitat de Barcelona, Spain

by

Niall Patrick Mortimer
from Dublin, Ireland

This thesis was written under a jointly supervised double degree program of the Julius-Maximilians-Universität Würzburg, Würzburg, Germany and the Universitat de Barcelona, Barcelona, Spain

Submitted on: …………………………………………………………..……..

Office stamp

Members of the Thesis Committee

Chairperson: Prof. Dr. Caroline Kisker (Julius-Maximilians-Universität Würzburg)

Primary Supervisor: Prof. Dr. Klaus-Peter Lesch (Julius-Maximilians-Universität Würzburg

Supervisor (Second): PD Dr. Robert Blum (Julius-Maximilians-Universität Würzburg)

Supervisor (Third): Dr. Marta Ribases (Supervisor, Universitat de Barcelona)

Date of Public Defence: ……………………………………….…………

Date of Receipt of Certificates: …………………………………………….

Affidavit

I hereby confirm that my thesis entitled ADHD Genetics in Mouse and Man 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.

Boston, Massachusetts, USA, 07/07/20
Place, Date   Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation ADHS Genetik bei Maus und Mensch eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Boston, Massachusetts, USA, 07/07/20
Ort, Datum   Unterschrift

z

Acknowledgments

I would like to extend my sincerest thanks to Prof. Klaus-Peter Lesch and Dr Marta Ribasés for their joint mentorship and supervision of my research project. I would also like to thank the members of both their lab groups, in particular Dr Olga Rivero, who provided invaluable support throughout this project.

Thank you to my family, without the lifetime of encouragement from my grandparents, brother and in particular my parents Patrick and Mary, I would never have been able to even contemplate getting my doctorate. Finally I would like to thank my wife Roisin Mortimer who has made this last five years of study the happiest period of my life and whose support is never-ending.

Table of Contents

Summary

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder with an estimated heritability of around 70%. In order to fully understand ADHD biology it is necessary to incorporate multiple different types of research. In this thesis, both human and animal model research is described as both lines of research are required to elucidate the aetiology of ADHD and development new treatments. The role of a single gene, *Adhesion G protein-coupled receptor L3* (*ADGRL3*) was investigated using a knockout mouse model. *ADGRL3* has putative roles in neuronal migration and synapse function. Various polymorphisms in *ADGRL3* have been linked with an increased risk of attention deficit/hyperactivity disorder (ADHD) in human studies. *Adgrl3*-deficient mice were examined across multiple behavioural domains related to ADHD: locomotive activity, visuospatial and recognition memory, gait impulsivity, aggression, sociability and anxiety-like behaviour. The transcriptomic alterations caused by *Adgrl3*-depletion were analysed by RNA-sequencing of three ADHD-relevant brain regions: prefrontal cortex (PFC), hippocampus and striatum. Increased locomotive activity in *Adgrl3*$^{-/-}$ mice was observed across all tests with the specific gait analysis revealing subtle gait abnormalities. Spatial memory and learning domains were also impaired in these mice. Increased levels of impulsivity and sociability accompanying decreased aggression were also detected. None of these alterations were observed in *Adgrl3*$^{+/-}$ mice. The numbers of genes found to exhibit differential expression was relatively small in all brain regions sequenced. The absence of large scale gene expression dysregulation indicates a specific pathway of action, rather than a broad neurobiological perturbation. The PFC had the greatest number of differentially expressed genes and gene-set analysis of differential expression in this brain region detected a number of ADHD-relevant pathways including dopaminergic synapses as well as cocaine and amphetamine addiction. The most dysregulated gene in the PFC was *Slc6a3* which codes for the dopamine transporter, a molecule vital to current pharmacological treatment of ADHD. The behavioural and transcriptomic results described in this thesis further validate *Adgrl3* constitutive knockout mice as an experimental model of ADHD and provide neuroanatomical targets for future studies involving *ADGRL3* modified animal models.

The study of ADHD risk genes such as *ADGRL3* requires the gene to be first identified using human studies. These studies may be genome based such as genome wide association studies (GWAS) or transcriptome based using microarray or RNA sequencing technology. To

explore ADHD biology in humans the research described in this thesis includes both GWAS and trancriptomic data. A two-step transcriptome profiling was performed in peripheral blood mononuclear cells (PBMCs) of 143 ADHD subjects and 169 healthy controls. We combined GWAS and expression data in an expression-based Polygenic Risk Score (PRS) analysis in a total sample of 879 ADHD cases and 1919 controls from three different datasets. Through this exploratory study we found eight differentially expressed genes in ADHD and no support for the genetic background of the disorder playing a role in the aberrant expression levels identified. These results highlight promising candidate genes and gene pathways for ADHD and support the use of peripheral tissues to assess gene expression signatures for ADHD.

This thesis illustrates how both human and animal model research is required to increase our understanding of ADHD. The animal models provide biological insight into the targets identified in human studies and may themselves provide further relevant gene targets. Only by combining research from disparate sources can we develop the thorough understanding on ADHD biology required for treatment development, which is the ultimate goal of translational science research.

Zusammenfassung

Die Aufmerksamkeitsdefizit- / Hyperaktivitätsstörung (ADHS) ist eine neurologische Entwicklungsstörung mit einer geschätzten Erblichkeit von etwa 70%. Um die ADHS-Biologie vollständig verstehen zu können, müssen verschiedene Forschungsansätze verfolgt werden. In dieser Dissertation werden sowohl Forschungsansätze am Menschen als auch im Tiermodell beschrieben, da beide Forschungsansätze erforderlich sind, um die Ätiologie von ADHS aufzuklären und neue Therapien zu entwickeln. Die Rolle eines einzelnen Gens, des Adhesion G-Protein-gekoppelten Rezeptors L3 (ADGRL3), wurde unter Verwendung eines Knockout-Mausmodells untersucht. ADGRL3 spielt eine mutmaßliche Rolle bei der neuronalen Migration und der Synapsenfunktion. Verschiedene Polymorphismen in ADGRL3 wurden in Studien an Menschen mit einem erhöhten Risiko für Aufmerksamkeitsdefizit- / Hyperaktivitätsstörung (ADHS) in Verbindung gebracht. Adgrl3-defiziente Mäuse wurden in mehreren Verhaltensbereichen im Zusammenhang mit ADHS untersucht: Bewegungsaktivität, visuelles und Erkennungsgedächtnis, Gangimpulsivität, Aggression, Umgänglichkeit und angstartiges Verhalten. Die durch Adgrl3-Depletion verursachten transkriptomischen Veränderungen wurden durch RNA-Sequenzierung von drei ADHS-relevanten Hirnregionen analysiert: präfrontaler Cortex (PFC), Hippocampus und Striatum. Bei allen Tests wurde eine erhöhte Aktivität der Lokomotive bei Adgrl3 - / - Mäusen beobachtet, wobei die spezifische Ganganalyse subtile Gangstörungen aufdeckte. Das räumliche Gedächtnis und die Lerndomänen waren bei diesen Mäusen ebenfalls beeinträchtigt. Es wurde auch ein erhöhtes Maß an Impulsivität und Umgänglichkeit festgestellt, begleitet von verminderter Aggression. Keine dieser Veränderungen wurde bei Adgrl3 +/- Mäusen beobachtet. Die Anzahl der Gene, bei denen eine unterschiedliche Expression festgestellt wurde, war in allen sequenzierten Hirnregionen relativ gering. Das Fehlen einer Dysregulation der Genexpression in großem Maßstab weist eher auf einen spezifischen Wirkmechanismus als auf eine breite neurobiologische Störung hin. Die PFC hatte die größte Anzahl differentiell exprimierter Gene, und eine Gen-Set-Analyse der differentiellen Expression in dieser Hirnregion ergab eine Reihe von ADHS-relevanten Signalwegen, einschließlich dopaminerger Synapsen sowie Kokain- und Amphetaminsucht. Das am stärksten dysregulierte Gen in der PFC war *Slc6a3*, das für den Dopamintransporter kodiert. Dieses Gen ist bei der derzeitigen pharmakologischen Behandlung von ADHS von entscheidender Bedeutung. Die in dieser Arbeit beschriebenen Verhaltens- und Transkriptomergebnisse bestätigen die konstitutiven Adgrl3-Knockout-Mäuse als

experimentelles Modell für ADHS und liefern neuroanatomische Zielstrukturen für zukünftige Studien mit ADGRL3-modifizierten Tiermodellen.

Die Untersuchung von ADHS-Risikogenen wie ADGRL3 erfordert zunächst, dass das Gen in Studien im Menschen identifiziert wird. Diese Studien können genombasiert sein, z.B. wie genomweite Assoziationsstudie (GWAS), oder transkriptombasiert unter Verwendung von Microarray- oder RNA-Sequenzierungstechnologie. Um die ADHS-Biologie beim Menschen zu erforschen, umfassen die in dieser Arbeit beschriebenen Forschungsansätze sowohl GWAS- als auch trankriptomische Daten. Ein zweistufiges Transkriptom-Profiling wurde in mononukleären Zellen des peripheren Blutes (PBMCs) von 143 ADHS-Patienten und 169 gesunden Kontrollpersonen durchgeführt. Wir kombinierten GWAS- und Expressionsdaten in einer Expressions-basierten PRS-Analyse (Polygenic Risk Score) in einer Gesamtstichprobe von 879 ADHS-Fällen und 1919 Kontrollen aus drei verschiedenen Datensätzen. Durch diese Untersuchungen fanden wir acht differentiell exprimierte Gene bei ADHS und keinen Hinweis darauf, dass der genetische Hintergrund der Störung eine Rolle bei den identifizierten aberranten Expressionsniveaus spielt. Diese Ergebnisse weisen auf vielversprechende Kandidatengene und Genwege für ADHS hin und unterstützen die Verwendung peripherer Gewebe zur Beurteilung der Genexpressionssignaturen für ADHS.

Diese Arbeit zeigt, dass sowohl Forschungsansätze am Menschen als auch Tiermodelle erforderlich sind, um unser Verständnis von ADHS zu verbessern. Die Tiermodelle bieten biologische Einblicke in die in Studien an Menschen identifizierten Ziele und können selbst weitere relevante Genziele liefern. Nur durch die Kombination von Forschungsansätzen aus unterschiedlichen Quellen können wir ein tiefes Verständnis der ADHS-Biologie entwickeln, das für die Entwicklung von Behandlungsstrategien erforderlich ist. Dies ist das ultimative Ziel der translationalen wissenschaftlichen Forschung.

Resumen

El trastorno por déficit de atención con hiperactividad (TDAH) es un trastorno del desarrollo neural con una heredabilidad estimada de alrededor de un 70%. Para poder comprender plenamente la biología del TDAH, es necesario incorporar diversos tipos de investigación. En esta tesis, se describe la investigación en modelos tanto humanos como animales, ya que se requieren ambas líneas de investigación para aclarar la etiología del TDAH y poder desarrollar nuevos tratamientos. El papel de un solo gen, el receptor L3 acoplado a la proteína de adhesión G (ADGRL3) se ha investigado utilizando un modelo de ratón knock-out. El ADGRL3 tiene efectos putativos en la migración neuronal y en la función de la sinapsis. Varios polimorfismos en ADGRL3 se han relacionado con un mayor riesgo de trastorno por déficit de atención/ hiperactividad (TDAH) en estudios en humanos. Adicionalmente se han examinado ratones deficientes en ADGRL3 en varios ámbitos conductuales relacionados con el TDAH tales como la actividad locomotriz, la memoria visoespacial y de reconocimiento, la impulsividad de la marcha, la agresividad, la sociabilidad y los comportamientos similares a la ansiedad. Las modificaciones trabscriptómicas causadas por el agotamiento de ADGRL3 se han analizado por secuenciación del ARN de tres regiones del cerebro relevantes al TDAH: la corteza prefrontal (CPF), el hipocampo, y el estriado. Se ha observado una mayor actividad locomotriz en ratones ADGRL3 -/- en todas las pruebas con el análisis específico de la marcha que revela anomalías sutiles de la marcha. La memoria espacial y los dominios de aprendizaje también se han visto afectados en estos mismos ratones. También se detectaron niveles aumentados de impulsividad y sociabilidad que acompañan a la disminución de la agresividad. Ninguno de estos cambios se han observado en ratones ADGRL3 +/-. El número de genes encontrados que exhibieron una expresión diferencial ha sido relativamente bajo en todas las regiones del cerebro secuenciadas. La ausencia de desregulación de expresión génica a gran escala indica una vía de acción específica, en vez de una perturbación neurobiológica amplia. La corteza prefrontal tenía el mayor número de genes expresados diferencialmente y el análisis de conjuntos de genes de expresión diferencial en esta región del cerebro ha mostrado una serie de vías relevantes para el TDAH, incluyendo las sinapsis dopaminérgicas así como la adicción a la cocaína y a las anfetaminas. El gen más desregulado en la corteza prefrontal fue el Slc6a3, que codifica para el transportador de dopamina, una molécula esencial para el tratamiento farmacológico actual del TDAH. Los resultados conductuales y transcriptómicos descritos en esta tesis dan aún más validez a los ratones knock-out constitutivos de Adgrl3 como modelo experimental de TDAH y ofrecen objetivos neuroanatómicos para estudios futuros con modelos animales modificados con ADGRL3.

El estudio de genes de riesgo de TDAH como el ADGRL3 requiere que el gen se identifique primero mediante estudios en humanos. Estos estudios pueden basarse en el genoma, como GWAS (estudio extenso de asociación en todo el genoma) o en transcriptoma, usando microarrays o tecnología de secuenciación de ARN. Para explorar la biología del TDAH en humanos, la investigación descrita en esta tesis incluye datos GWAS y trancriptómicos. Se ha realizado un perfil de transcriptoma de dos fases en células mononucleares de sangre periférica (CMSP) de 143 sujetos con TDAH y 169 controles sanos. Hemos combinado GWAS y datos de expresión en un análisis de puntuación de riesgo poligénico con sede en expression genica en una muestra total de 879 casos de TDAH y 1919 controles de tres conjuntos de datos distintos. A través de este estudio exploratorio, hemos encontrado ocho genes expresados diferencialmente en el TDAH y además que no existe indicio de que el fondo genético del trastorno tiene un papel en los niveles de expresión aberrantes identificados. Estos resultados subrayan genes candidatos prometedores y vías genéticas para el TDAH y además apoyan el uso de tejidos periféricos para evaluar las firmas de expresión génica para el TDAH

Esta tesis muestra cómo se requiere la investigación en modelos humanos y animales para aumentar nuestra comprensión del TDAH. Los modelos animales proporcionan información biológica sobre los objetivos identificados en estudios en humanos y pueden proporcionar objetivos genéticos relevantes adicionales. Solo mediante la combinación de las investigaciones de fuentes dispares podemos desarrollar la comprensión exhaustiva de la biología del TDAH necesaria para el desarrollo del tratamiento, lo que es el objetivo principal de la investigación científica traslacional.

1. Introduction

1.1 Neurodevelopmental Disorders

1.1.1. Strategies to study the aetiology of neurodevelopmental disorders

Neurodevelopmental disorders are a group of diseases which can impact all aspects of mental health including cognitive processing, emotionality and sociability. These disorders and other mental illnesses are a leading cause of disability throughout the world and place a significant strain on society from both a personal and economic perspective (Lamsal & Zwicker 2016). Genomic and transcriptomic studies are a powerful research method for understanding the aetiology of neurodevelopmental disorders. These studies use biologic samples such as blood or saliva from individuals diagnosed with a particular disorder and control individuals without that condition, to identify differences in the genome or transcriptome which may impact an individual's risk of developing the disorder. Once these differences are identified, the genes and even the specific variants implicated can be modelled in animal and cellular-based models. The aim of this animal and cellular model-based research is to increase understanding of the neurobiology of neurodevelopmental disorders and develop new treatments to improve the quality of life of affected individuals. In this thesis we will describe novel genomic, transcriptomic and animal based research which has increased the understanding of neurodevelopmental disorders, in particular: attention-deficit/hyperactivity disorder (ADHD). The findings of this research may also be relevant to other neurodevelopmental disorders including autism spectrum disorder (ASD), which shares a significant genetic overlap with ADHD.

1.1.2. Attention-deficit/hyperactivity disorder (ADHD)

According to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM5) attention-deficit/hyperactivity disorder (ADHD) is a clinically heterogeneous neurodevelopmental disorder characterized by inattention, hyperactivity and increased impulsivity/emotionality (American Psychiatric Association 2013). Approximately 2-10% of school age children are affected by ADHD, making it the most common childhood psychiatric disorder worldwide (Hawi et al. 2015). There are currently no objective laboratory based tests which can diagnose ADHD. Individuals are primarily diagnosed based on age-inappropriate and disruptive levels of hyperactivity, impulsivity and inattention. While not part of the diagnostic criteria, individuals with ADHD often display mood instability and frustration intolerance (Haavik et al. 2010). ADHD diagnosis and treatment is often complicated by the high prevalence of comorbid conditions, in particular substance use disorder (SUD) and other psychiatric disorders such as depression and anxiety disorder (Reimherr et al. 2017; Skoglund

et al. 2017). Despite traditionally being regarded as primarily a childhood disorder, a number of longitudinal follow-up studies of ADHD children coupled with population-based epidemiological studies have estimated the prevalence of adult ADHD (aADHD) at 2.5-4.9% (Franke et al. 2012). The severity and presentation of ADHD may change during an individual's life with adults generally showing less hyperactivity and possibly reduced impulsivity (Franke et al. 2018). While often less severe than the childhood form, aADHD remains a significant personal and societal issue with the total number of adults with ADHD possibly outweighing the number of children. In 2012, the total economic cost of ADHD in the United States alone was estimated at between $143 and $266 billion per year (Doshi et al. 2012). A recent systematic review reported a consistent increase in suicidality of individuals with ADHD based on 26 studies conducted over four continents; 16% of adults with ADHD were reported to have previously attempted suicide (Balazs & Kereszteny 2017). Given the massive cost and personal hardship caused by ADHD there is a large gap in treatment capability. The most commonly used treatment for ADHD is methylphenidate (MPH). While effective in controlling some patients' symptoms MPH shows a large degree of variability across patients. More research is required to decipher the pharmacogenomics of MPH with a recent review by Soleimani and associates (2017) finding contradictory evidence on the role of *SLC6A3* polymorphisms in the MPH response of ADHD patients.

1.1.3. Genetics and environmental factors of ADHD

While environmental factors are known to influence the development and manifestation of ADHD, genetic variants are recognized as critical aetiological components. A meta-analysis of multiple large scale twin studies estimated the heritability of childhood and adolescent ADHD at between 0.7 and 0.8 (Nikolas & Burt 2010). Heritability in aADHD is estimated to be lower, at between 0.3 and 0.4, although some experts believe this to be an underestimation due to the self-rating system commonly used in these studies. The true heritability is likely to be as high as 0.7-0.8 (Brikell et al. 2015). Attempts to explain the genetic variance leading to the development of ADHD can be divided into two overarching hypotheses; the common disease common variant (CDCV) and common disease rare variant (CDRV) hypotheses (Hawi et al. 2015). The CDCV hypothesis focuses on multiple common polymorphisms with a frequency greater than 5%. Each polymorphism has a low level of penetrance but combined with many others across an individual's genome can give rise to a high genetic risk of developing ADHD. This hypothesis has been the basis of a large amount of molecular genetic research over the past two decades (Li et al. 2014). With the advent of high-throughput genetic screening

technology, researchers moved to hypothesis-free methods investigating the whole genome in large samples of patients and controls. These Genome-Wide Association Studies (GWAS) have successfully identified a number of polymorphisms which confer a slightly increased risk of developing a psychiatric disorders but fail to explain the vast majority of genetic heritability (Maher 2008). Recently the CDRV hypothesis, which claims common diseases such as ADHD are caused by a large variety of genetic variants that are rare in the overall population but have high penetrance when present in an individual has gained traction. One particular type of rare variant which has become a subject of intense research under the CDRV hypothesis is copy number variations (CNVs). CNVs are chromosomal deletions or duplications that span more than 1 kb and that can alter normal gene expression (Redon et al. 2006). There is increasing evidence that CNVs play a role in the pathogenesis of several neurodevelopmental disorders and, in the case of ASD, it has been demonstrated that patients carry an overall higher CNV number compared to healthy controls (Shishido et al. 2014). The relevance of CNVs to a particular phenotype could be linked to their position, length and heterozygosity. Rare and de novo CNVs have been implicated in increasing the risk of ADHD (Lionel et al. 2011). Some CNVs implicated in ADHD are also found to increase the risk of ASD and schizophrenia (Gudmundsson et al. 2019). The growing consensus among researchers is that the development of ADHD and other neurodevelopmental disorders is the result of a number of synergistic factors including CNVs and other rare variants in combination with common polymorphisms.

Epigenetic modification due to environmental stimuli may act in combination with an individual's genome to confer an increased risk of neurodevelopmental disorders. Certain developmental periods such as the pre/perinatal stages are suspected to be particularly susceptible to environmental stimuli. Premature birth and maternal smoking during pregnancy have been linked with increased risk of developing ADHD (Halmøy et al. 2012; Zhu et al. 2014). However when genetic and familial confounders are included in study models, the strength of these associations decrease, calling into question any causal relationship (Sciberras et al. 2017). A cohort study of over 1.5 million Swedish children failed to find an association between maternal antidepressant use during the first trimester and increased risk of ADHD (Sujan et al. 2017). Risk factors may also be specific for particular neurodevelopmental disorders with a recent study failing to find any shared pre/perinatal risk factors between ADHD and ASD (Oerlemans et al. 2016). After the antenatal period, a number of early life factors have also been linked to these neurodevelopmental disorders. These include an association between increased maternal anxiety and depression and ADHD among three year

olds (Vizzini et al. 2019). The direction of association between environmental factors and neurodevelopmental disorders is hard to determine and further mechanistic based research including the use of animal and cellular models is required to determine if there is a causal relationship behind any of these associations.

1.2. Human Biology Study Techniques
1.2.1. Linkage Analysis
Family based genetic studies commonly use a technique called linkage analysis to map genetic variants within a group of related individuals. Linkage analysis relates the sharing of particular genetic markers amongst family members with a phenotypic similarity amongst these individuals. Linkage analysis is a particularly powerful tool for the detection of highly penetrant genetic variants. Due to natural selection, genetic disorders which are caused by a single or small number of variants are typically rare disorders and the associated genetic loci are uncommon in the general population. However, these highly penetrant variants may be prevalent within a particular family with a strong history of the disorder and can therefore be detected using linkage analysis.

Traditional linkage analysis known as parametric linkage analysis assumes a specific method of inheritance and explicitly defines the link between a particular genetic variant and the phenotype being investigated. The strength of the evidence for linkage (cosegregation) between the genetic marker and the phenotype is represented as the logarithm of the odds (LOD) score (Morton et al. 1955). Non-parametric linkage analysis does not make a specific assumption about the model of inheritance. This method examines chromosomal regions to determine if the segregation of genetic markers and the phenotype occurs at different frequencies than would be expected by random identity by descent (IBD) sharing probabilities (Bailey-Wilson 2018). These non-parametric techniques are typically better at investigating more genetically complex disorders where the pattern of inheritance is unknown. Conventional parametric analysis is a more statistically powerful technique when the pattern of inheritance is known (Teare & Barrett 2005). Both linkage analysis techniques work best when used to investigate large families which contain phenotype and genotype information on all members. While linkage studies are typically associated with the investigation of single gene Mendelian disorders, the combination of parametric and particularly non-parametric linkage analysis techniques has provided valuable insights into the aetiology of genetically complex neurodevelopmental disorders including ADHD. The initial research which linked *ADGRL3* to

ADHD used a genetic linkage approach on large multigenerational families in a Columbian population isolate (Arcos-Burgos et al. 2010)

1.2.2. Genome-Wide Association Studies

While linkage studies are very effective at identifying high penetrance variants, these techniques are poor at identifying low penetrance variants which have a small impact on individual risk. While these low penetrance variants may not explain a large degree of the relative risk for one individual, they may have an important effect on the population-wide prevalence of the disorder due to high variant frequencies. In order to identify these common variants, larger samples of unrelated individuals must be collected and genotyped. Over the past two decades, GWAS of rapidly increasing sample sizes have been used to investigate complex disorders including neurodevelopment disorders.

Most GWAS focus on Single Nucleotide Polymorphisms (SNPs), which are the most common type of variation across populations. Rather than examining over 10 million common SNPs in non-African populations, GWAS typically uses the principle of linkage disequilibrium (LD) to reduce this number to a more manageable level of approximately 500,000 SNPs with roughly twice as many SNPs required for equally powered studies in African populations (Altshuler et al. 2008). LD is the non-random association between alleles at different loci. Alleles are said to be in LD if they are more commonly found together than would be expected if they were independently and randomly associated. Typically, alleles which are located more closely together on a chromosome are in stronger LD as they are less likely to be separated during chromosomal recombination. The distance required for LD to break down determines how many alleles must be assessed to detect a genetic haplotype. The HapMap project's identification of tagging SNPs which captured the majority of common genetic variation in most human populations coupled with the commercial development of dense genotype arrays that allow these SNPs to be genotyped in a single array led to the creation of biobanks and large case-control GWAS (Visscher et al. 2012).

GWAS is a hypothesis-free research method which does not require the input of prior biological information regarding the disorder being investigated. However, GWAS arrays are biased by their design to detect causal variants that are common in the population. The alleles on the array are chosen to have a population frequency of over 5%. Rare variants which may be causal and are in the same recombination interval as the common allele could be missed because they were not found to be in LD due to their low population frequency. The ability of GWAS to identify the variants underlying complex disorders is therefore dependent on these

disorders being the result of high-frequency low-penetrance alleles, as proposed by the CDCV hypothesis previously described.

One way to move past this assumption is to gather information on the non-common variants which are missed by traditional GWAS arrays. Newer and increasingly common forms of GWAS include whole exome and whole genome sequencing, which cover all an individuals' coding regions or the whole genome respectively, rather than depending on proxy tagging SNPs. More in-depth sequencing may allow burden tests which take account of multiple causative alleles in the same gene to increase the power to detect rare variants. The reducing cost of genome sequencing which started with the advent of next-generation sequencing technologies and continues with novel technologies such as the Oxford Nanopore sequencer has made the use of whole-exome (WES) and whole-genome sequencing (WGS) increasingly economically viable both for research and clinical care (Visscher et al. 2017). WGS is already providing valuable insights into the impact of rare variants of genetically complex human traits such as height. While a GWAS study encompassing over 250,000 individuals captured the majority of anticipated heritability in height across the population, a small amount of anticipated genetic heritability is still missing (Wood et al. 2014). WGS of a much smaller sample of 21,620 individuals was able to detect this missing heritability and by combining common and rare variants arrive at a heritability estimate of 0.79, a level which matches heritability estimates from twin studies (Wainschtein et al. 2019, Preprint). The combination of WES and linkage analysis has already been used to identify ADHD associated risk variants such as the rare variant (rs151326868) located within AAED1 (Corominas et al. 2018).

Large-scale GWAS have provided increasingly in-depth insights into the aetiology of complex disorders. Unsurprisingly, GWAS has confirmed the highly polygenic nature of complex disorders with over eleven thousand SNPs identified by GWAS reaching the genome-wide statistically significant p value threshold of 5×10^{-8} (Welter et al. 2014). Due to the large number of genes and genetic variants involved in complex disorders, the proportion of variance explained by a single variant is very small. The small effect size of variants means an extremely large sample size is required. As sample sizes have increased, the number of associated variants for each disorder has increased accordingly, for example a sample size of 3,322 cases and 3,587 controls was sufficient to identify one genomic variant associated with schizophrenia while a sample size of 36,989 cases and 113,075 controls identified 108 loci that meet genome-wide significance (Purcell et al. 2009; Ripke et al. 2014). The first genome-wide statistically significant variants associated with ADHD were recently found in a study of 20,183 diagnosed

ADHD cases and 35,191 controls (Demontis et al. 2019). Despite the success in identifying variants, the interpretation of GWAS results may be difficult. The association of a genomic variant with a disorder does not necessarily mean the gene in which the variant is located is directly involved in the mechanism which brings about the phenotype.

1.2.3. Polygenic Risk Score Analysis

Regardless of the many unknown causal pathways between variants and a phenotype, genetic variants can still provide predictive value. While each individual variant associated with a phenotype by GWAS typically contributes a very small relative risk, in composite these risk variants may have a significant predictive value for the probability of an individual developing the particular phenotype. Genome-wide polygenic risk scores (PRS) take into account the composite risk of variants across the genome to predict the probability of an individual developing a phenotype, such as a particular disorder. Different PRS take account of a widely divergent number of variants from dozens to several thousand genomic variants depending on which combination gives the highest predictive value (Sugrue & Desikan 2019).

A model used to develop PRS must take into account the number of variants included, the statistical model used to combine the variants and the generalisability of the score to the overall population. These models must also take account of the individual's age and underlying biology of the phenotype being investigated, for example an individual's PRS for childhood ADHD may be similar but not identical to the individual's PRS for aADHD (Rovira et al. 2019, bioRxiv, https://doi.org/10.1101/589614). Ideally PRS should also take account of the continuum of a disorder from mild to severe rather than a simple dichotomous classification of disorder case and control (Torkamani et al. 2018). As attempts to use PRS in the clinic become increasingly common, the inability to generalise PRS across different ethnicities is becoming an issue. The GWAS performed so far are highly biased towards individuals of European descent and since PRS are developed using training data from available GWAS samples, these scores have poorer predictive value for individuals who are not of European ancestry. This reduction in predictive value may be the result of different LD patterns or the presence of different causal variants in different ethnic populations (Márquez-Luna et al. 2017). Despite this issue, PRS for complex disorders including ADHD are becoming increasingly powerful tools for determining individuals' genetic susceptibility (Demontis et al. 2019). Regardless of the growing potential of PRS for predicting individuals' risk of developing a complex disorder, they must be considered in conjunction with other risks such as lifestyle and other environmental factors, as complex disorders are not completely caused by genetic variants.

1.2.4. Multi-omic Approaches

Genomics is the most widely studied of the "omic" disciplines. While identifying variation in genetic sequences across the population and linking these variations to increased risk of complex disorders has provided predictive estimates in the form of PRS and valuable insights into disease aetiology, genomic information alone does not provide a complete picture of the biology of these disorders. Genomic information provides a static picture of the individual genome which fails to take account of variation between cell types and the dynamic response of cells to internal and external stimulus.

Genetic sequences are transcribed into transcripts which collectively form the transcriptome. The transcriptome is discussed in greater detail in the following section but it is important to note that only a small subset of transcripts are translated into proteins. The difference between cells and their function is primarily manifested by the differences in their protein components, as the majority of cellular functions are performed by proteins. The collective study of all proteins in a biological structure at a particular time point is known as proteomics. Proteomics is one of the newest branches of -omic studies but advances in mass spectrometry technology are making increasingly detailed and extensive proteomic studies possible (Iwamoto & Shimada 2018). Coupling of proteomic and genomic studies will provide valuable insights into the impact of genetic variation on final protein form and function. The vast majority of current pharmacological interventions are targeted at proteins, these studies are therefore promising avenues for the identification of novel drug targets for pharmacological treatment (Frantzi et al. 2019).

In order to move from a genomic code to a functioning proteome, the expression of genes must be tightly controlled. Epigenomics is the study of non-sequence alternations which alter the expression of a genotype in a phenotype (Dupont et al. 2009). Epigenetics is the method through which cells with the same genetic code can differentiate into different cellular types. An abnormality in the epigenetics of a cell may lead to incorrect differentiation and disruption of cellular function. At a network level these disruptions may play a role in the aetiology of complex disorders including neurodevelopmental disorders (Meng et al. 2019). Epigenetic modifications include DNA and chromatin modifications via various mechanisms such as methylation, acetylation, phosphorylation, ubiquitylation, and sumolyation, all of which have a corresponding –omic field of study aimed at identifying these modifications across the entirety of a cell's epigenome (Weinhold 2006). Another branch of epigenetic research concerns RNA epigenetic modifications, in particular the role of non-coding RNAs. Two

classes of non-coding RNAs are particularly important for the modulation of genotype to phenotype, these non-coding RNAs are: microRNAs (miRNAs), a subset of RNA molecules less than 28 nucleotides in length and circular RNA (circRNA) named for their circular closed loop structure formed due the absence of 5′ caps or 3′ poly-A tails (Catalanotto et al. 2016, Barrett & Salzman 2016). Both these RNA classes act in conjunction with DNA and chromatin modifications to control the transcription of coding sequences of DNA into mRNA and the translation of mRNA into final protein products.

1.2.5. Transcriptomics

The study of all classes of RNA transcripts, both coding and non-coding, is known as transcriptomics. Transcriptomics provides another valuable layer of information which takes account of the unique characteristics of cells and how their RNA levels change over time and in response to internal and external stimuli.

Transcriptomic studies started in the 1990's with the creation of DNA microarrays. Microarrays are composed of thousands of nucleic acids which are bound to the array surface. These nucleic acid sequences are known as probes. These probes hybridise to complementary DNA sequences known as targets. When microarrays are used to investigate RNA levels these DNA sequences are created by the reverse transcription of RNA into complementary DNA (cDNA) which includes a fluorescently labelled probe or for Affymetrix microarrays such as the microarrays used in this thesis project, a biotin-labelled nucleotide. Once the probes and targets are hybridised, the non-complementary unbound targets are washed away. The level of fluorescence at each probe site is measured using a confocal microscope either immediately in the case of fluorescent-tagged targets, or for biotin-labelled samples following post-hybridisation with fluorescently labelled streptavidin. The strength of the fluorescence detected is proportional to the amount of hybridised DNA which in turn is proportional to the amount of original RNA in the sample (Epstein et al. 2000).

Microarrays have a number of limitations. The range in which fluorescence is linearly proportional to DNA fragment concentration is restricted by saturation at high concentrations and the kinetics of hybridisation favours no binding at low concentrations. The specificity of probes is also limited due to the complexity of human genomes and transcriptomes. This complexity leads to difficulty designing probes which do not bind homologs of the target gene, resulting in off-target hybridisation. The same issue makes it difficult to design probes which are capable of detecting exon-specific mRNA transcript levels. Finally, microarray probes can only target and detect known sequences, meaning unannotated genes and non-coding RNA's

which are frequently not included on the arrays are completely missed by this technology (Bumgarner et al. 2013).

In order to overcome the limitations associated with microarrays, RNA-sequencing has become an increasingly popular technology. With the creation and accessibility of next generation high-throughput sequencing methods, deep resolution RNA-sequencing has become an economically viable option for mapping and quantifying transcriptomes. Similarly to microarrays, the investigation of RNA levels by RNA-sequencing typically involves the conversion of RNA into cDNA. The conversion to cDNA can be performed on all RNA molecules including non-coding RNA or targeted to 3' polyadenylated (polyA+) tails in order to specifically select mRNA. The goal of the RNA-sequencing study must be considered as particular methods have certain advantages, for example ribosomal RNA depletion allows more detailed investigation of non-coding RNA molecules while PolyA+ selection gives superior exonic coverage which provides a more accurate quantification of gene expression (Zhao et al. 2018).

The most important consideration during the sequencing step is the read depth. This refers to the average number of times a particular nucleotide in the sample will be sequenced. The level of read depth will determine the statistical power for a given sample size. Low read depth will result in high technical variation while excessive reads will not significantly improve statistical power and result in unnecessary costs (Sims et al. 2014). Once the reads are produced the assembly of the transcriptome is a computational process typically involving the alignment of reads to a reference genome for the target organism being studied. (Schbath et al. 2012)

Conventional RNA-sequencing using short read lengths and single end sequencing is sufficient to investigate differential gene expression by measuring mRNA levels at a specific time point. Novel RNA-sequencing technologies including longer read lengths of approximately 100bp instead of 50bp and paired end sequencing from both the 3′ and 5′ end allow more in-depth analysis including exon specific quantification and identification of previously unknown transcripts (Alamancos et al. 2015). The design of RNA-sequencing studies to allow investigation beyond simple differential gene expression has led to numerous insights into the complexity of RNA splicing and the role of non-coding RNA molecules in the regulation of gene expression (Stark et al. 2019). The removal of the reverse transcription step by new direct RNA sequencing technology provides the potential for new transcriptome-wide insights into

RNA base pair modification and quantification of poly(A) tail lengths without the biases introduced by RNA conversion to cDNA (Workman et al. 2019; Depledge et al. 2019).

With regard to neurodevelopmental disorders, one limitation which modern transcriptomic techniques are unable to overcome is the inaccessibility of the primary tissue of interest, namely brain tissue. While access to post-mortem neural tissue samples is highly informative, the temporal dynamic nature of the transcriptome limits its utility in comparison to *in vivo* samples. Peripheral tissue, most commonly blood samples, are therefore used as a proxy for brain tissue in transcriptomics studies of living patients. Tissue-specificity means any differential expression patterns observed in human blood samples cannot be assumed to be replicated in the brain (Tylee et al. 2013). However, studies have shown significant correlation between gene expression across blood and neural tissue. Analysis of microarray results from whole blood and 16 different brain regions found a median non-parametric correlation of approximately 0.5 between whole blood and CNS transcripts. Genes which had higher expression levels in each tissue were found to have an even higher degree of correlation (Sullivan et al. 2006). Further research is required to identify correlations between the CNS transcriptome and peripheral tissues including the identification of the best proxy tissues. One study investigating specific blood cell types found a slightly higher level of correlation between the transcriptome of leukocytes and the CNS in comparison to whole blood samples (Cai et al. 2010). Even if differential expression patterns in blood samples from cases and controls are not always replicated in the brain, thus limiting the mechanistic insights which can be drawn, blood transcriptome signatures could still provide clinical utility as biomarkers for neurodevelopment disorders. For example, blood gene expression signatures assessed by microarray have been found to be differentiate between siblings with and without ASD (Kong et al. 2012).

The field of transcriptomics alongside other –omics disciplines, in particular genomics, is set to continue growing. This will provide novel biomarkers, mechanistic insights and further gene targets for investigation in animal and cellular models.

1.3. Adhesion G protein-coupled receptor L3 (ADGRL3)

Once a gene has been extensively linked to a neurodevelopmental disorder by human genetic or other –omic studies, the next step in understanding gene function is often the creation of an animal model. This thesis includes an extensive study of a mouse model of ADHD. This model involved the genetic modification of mice to be *Adhesion G protein-coupled receptor L3* (*Adgrl3*) deficient. A number of genetic studies in human populations have linked

polymorphisms in *ADGRL3,* also known as *Latrophilin 3 (LPHN3),* with an increased risk of ADHD. These human genetic studies, alongside previously published literature on *ADGRL3* animal and cellular models are described below.

1.3.1. *ADGRL3:* **Human Genetic Studies**

The association between *ADGRL3* and ADHD is one of the most robustly replicated genetics findings in ADHD genetic research. A large number of human genetic studies have implicated multiple polymorphisms in the *ADGRL3* gene with an increased risk of ADHD, SUD and modulation of stimulant response (Uhl et al. 2008a-b; Arcos-Burgos et al. 2010; Ribasés et al. 2011; Jain et al. 2012; Bruxel et al. 2015; Martinez et al. 2016; Gomez-Sanchez et al. 2017; Arcos-Burgos M et al. 2019; Kappel et al. 2019). The most recently published of these studies also found an association between a polymorphism in *ADGRL3* and an increased risk of developing ASD as well as ADHD (Kappel et al. 2019).

The first of these studies to link variants in *ADGRL3* with ADHD was a linkage study performed on 433 individuals from 18 large multigenerational families in the Paisa population of Antioquia, Colombia. This population forms a genetic isolate with a high prevalence of ADHD making it a perfect population for genetic linkage studies aimed at identifying novel genes which contribute to the aetiology of ADHD. The genetic association was originally narrowed to the 4q13.2 chromosomal region before the identification of a shared ADHD susceptibility haplotype in *ADGRL3* gene. Following meta-analysis using 5 international samples, three markers were identified which passed a test for heterogeneity and survived significance threshold correction for multiple testing. These markers were also found to modulate methylphenidate response (Arcos-Burgos et al. 2010).

Following the initial study reporting an association between ADHD and *ADGRL3* variants, Ribasés et al. (2011) performed an independent case control study aimed at replicating this association in aADHD. While the original study primarily involved children, this replication study involved aADHD patients. 43 SNPs covering the *ADGRL3* gene were investigated in 334 adults with ADHD and 334 adult control subjects. This study found further evidence of the link between *ADGRL3* and ADHD by identifying five SNPs (rs1868790, rs2122643, rs6858066, rs4860106, rs13115125) linked with aADHD (Ribasés et al. 2011).

Further human genetic studies investigated the interaction between *ADGRL3* variants and variants at other genetic loci. Multiple two-locus interactions between five SNPs in *ADGRL3* and four SNPs in the 11q chromosome region, which increased the risk and severity of ADHD

have been reported. These interacting SNPs were located inside or near the 11q chromosomal region containing the *Dopamine Receptor D2 (DRD2)* and *neural cell adhesion molecule 1 (NCAM1)* genes (Acosta et al. 2011). Another study using the original Paisa population found a significant association between rs6551665 and rs877137 located in the 11q chromosomal region. The 11q haplotype was a modulator of the *ADGRL3* variant as it was not found to increase ADHD risk in the absence of the *ADGRL3* susceptibility variant. The presence of risk variants at both sites resulted in a large 2.46-fold risk increase in developing ADHD (Jain et al. 2012).

ADHD is frequently found to be co-morbid with other disorders, such as SUD. The lifetime risk of developing SUD for an individual with ADHD is approximately twice as high as a member of the general population and four times as high for an individual with ADHD and co-morbid conduct disorder (Wilens et al. 2011; Zulauf et al. 2014). The frequent co-morbidity of ADHD and SUD, coupled with findings from twin studies and human genetic studies, strongly suggests an overlap in the genetic variants underlying both these disorders (Palacio et al. 2004; Chang et al. 2012). In order to investigate if the *ADGRL3* variants previously associated with ADHD were predictive for the presence of SUD with or without a diagnosis of ADHD, a recursive-partitioning framework (classification tree analysis) was assembled using genotype data from family-based, case-control, and longitudinal samples from four independent cohorts around the world ($n = 2698$). A fixed-effects meta-analysis using results from the Advanced Recursive Partitioning Analysis (ARPA) of the four cohorts found an overall predictive accuracy of 0.727 (95% CI = 0.710–0.744). rs4860437 was the SNP with the highest predictive value in the ARPA model for each of the three studies for which information was available. Interestingly, in Caucasian populations, rs4860437 is in complete LD with rs6551665 and rs1947274, two of the three SNPs identified in the original study linking *ADGRL3* to ADHD. The inclusion of genetic information alongside baseline data increased the accuracy of predicting SUD development at a later time point in a 10-year longitudinal study of children with ADHD. The ability to identify children, particularly those with a prior diagnosis of ADHD, who have a high risk of developing SUD due to genetic liability would provide great clinical utility through the targeted management of these patients (Arcos-Burgos M et al. 2019).

None of the studies described above found an association between a coding variant in *ADGRL3* and ADHD or SUD. The known associations therefore come from a number of non-coding variants. The non-coding variants may impact the expression levels of *ADGRL3* or the levels of specific *ADGRL3* isoforms. An *in silico* cross-species comparison of evolutionarily

conserved regions (ECR) combined with chromatin annotation markers identified genomic regions which have a high probability of being enhancer regions. 834 individuals from the Colombian multigenerational family in which *ADGRL3* was originally linked to ADHD, were tested to identify common variants in ECRs which were predicted to impact gene regulation. Following correction for multiple testing, 6 of the chosen ECRs contained SNPs which were significantly linked to ADHD. It is possible that non-coding variants may not affect the genes that they are located in, but rather have effects on genes elsewhere in the genome. However, as described in the animal and cellular models section, there is functional evidence from the Martinez et al. (2016) study that the noncoding variant rs2271338 affects *ADGRL3* expression, further linking this gene to ADHD.

The absence of currently identified naturally occurring coding variants in humans greatly limits our ability to investigate *ADGRL3* in the human population. The clinical importance and evolutionarily conserved nature of *ADGRL3* makes it an attractive target for modelling in genetically modified organisms.

1.3.2. ADGRL3: Biological and Molecular Description
Latrophilins (LPHNs) are a subfamily of G-coupled protein receptors (GCPRs). They play an integral role in the receipt, transduction and response of cells to external stimuli (Südhof 2011). Research into GCPRs has greatly increased our understanding of cellular response mechanisms, which in turn has led to the development of multiple new pharmaceutical treatments (Hauser et al. 2017).

The Latrophilin family of proteins was initially discovered due its ability to bind α-Latrotoxin, an extremely potent neurotoxin found in black widow spider venom (Davletov et al. 1996). From this family of proteins, *ADGRL3* is of particular interest in relation to neurodevelopmental disorders. The production of ADGRL3 by neuronal cells, its localisation at synapses and its putative roles in neuronal migration and synapse development provide potential mechanisms through which *ADGRL3* genetic polymorphisms may increase the risk of developing ADHD (O'Sullivan et al. 2014). ADGRL3 forms trans-synaptic connections with a number of protein partners including fibronectin leucine-rich-repeat transmembrane proteins (FLRTs) and teneurins (Fig. 1) (Burbach & Meijer, 2019). ADGRL3 has also been shown to form a trimeric complex with FLRT3 and UNC5. This complex supports transcellular adhesion and glutamatergic synapse development, both vital processes for brain development and function (Jackson et al. 2015). The disruption of ADGRL3's ability to bind its transmembrane interaction partners has been shown to reduce excitatory synapse development

while not impacting inhibitory synapse development (Sando et al. 2019). Excitatory-inhibitory balance has been suggested as a central mechanism across neuropsychiatric disorders (Sohal & Rubenstein, 2019).
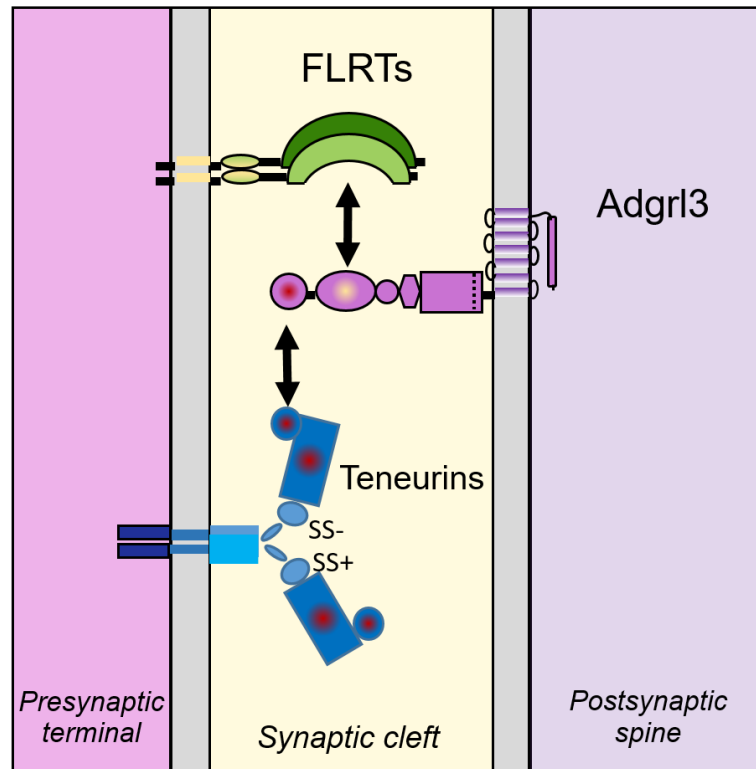
Figure 1. **Illustration of postsynaptic Adgrl3 protein interacting with FLRTs and Teneurins in the synaptic cleft to promote synapse development.** Image adapted from Sando et al. 2019

1.3.3. ADGRL3: Model Organisms

Animal and cellular models have been used extensively in biomedical research. In fact, they have been proven to be a fundamental tool to both study the pathogenesis of an illness and test the effects of possible treatments. The models described below allow the possibility to run large scale control-based trials, avoiding the ethical concern of testing new compounds on human subjects and decreasing the financial investment needed. Each of the models discussed below has greatly contributed to the understanding of *ADGRL3* and its role in neurodevelopmental disorders particularly ADHD.

While multiple human studies have shown associations between polymorphisms in *ADGRL3* and risk of developing ADHD, no functional variants have been identified in *ADGRL3* coding regions so far (Orsini et al. 2016). The lack of currently characterised, naturally occurring coding variants in the human population greatly limits our ability to directly investigate ADGRL3 in humans. As such, the creation of *ADGRL3* genetically modified animal models is essential to decipher the role and function of ADGRL3.

1.3.4. ADGRL3: Rodents

Rodent models, in particular the mouse, are the most commonly used model organism for studies of human biology and disease. Their short generation time, small size and social nature allows a large number of animals to be housed in a small facility compared to larger mammals. The first findings from an *Adgrl3* knockout mouse were reported by Wallis et al. (2012). When tested at 4, 8 and 12 weeks old, *Adgrl3* null mice displayed hyperactivity in the open field test. In the same study, *Adgrl3* mice were administered cocaine to test if they had a differential locomotive response to psychostimulants. A differential response by *Adgrl3* null mice to cocaine would be of particular interest due to both the high co-morbidity of ADHD and SUD and the link between *ADGRL3* polymorphisms and the increased risk of developing SUD. *Adgrl3* null mice displayed an increased response to the higher (20 mg/kg) dosage of cocaine in comparison to the wild-type controls.

Gene expression in *Adgrl3* null and wild-type P0 pups was investigated using qPCR. Genes involved in neuronal differentiation and survival were investigated alongside genes known to play a role in the dopaminergic and serotonergic neurotransmitter systems. Other ADHD candidate genes and loci which were known to interact with *Adgrl3* were also studied. *5-Htt, 5-Ht2a, Dat1, Drd4, Ncam, Nurr1,* and *TH* were found to be differentially expressed in *Adgrl3* null mice compared to controls. Neurochemical assays revealed an increased level of dopamine and serotonin in *Adgrl3* null mice. Disruption of neurotransmission is a common mechanism

in the pathology of neurodevelopmental disorders, and the discovery of abnormal gene expression and neurotransmitters levels in *Adgrl3* null mice is highly supportive of its potential as an ADHD candidate gene (Li et al. 2006; Faraone & Larsson al. 2019).

Adgrl3$^{-/-}$ mice performed more lever presses in a paradigm designed to investigate motivation to work for food, while the rotarod test failed to find any differential motor coordination. *Adgrl3*$^{-/-}$ mice had a greater swim time and had increased latency to immobility in the Porsolt swim test, possibly due to reduced depressive behaviour or as a result of hyperactivity (Orsini et al. 2016). Further investigation with a depression model such as the sucrose consumption test, which is less susceptible to confounding by hyperactivity, could help determine if *Adgrl3* null mice have genuinely different depressive behaviour (Perona et al. 2008).

RNA-sequencing was performed on the ADHD-relevant brain regions; prefrontal cortex, striatum, and hippocampus of *Adgrl3*$^{-/-}$ and *Adgrl3*$^{+/+}$ mice collected at 4 days, 28 days and 6 months. Across region and time point analysis 11 differentially expressed genes were found. Interestingly, 2 of these 11 genes (*Pcdhgb8* upregulated and *Pcdhb9* downregulated) are members of the protocadherin family of calcium dependent cell-cell adhesion molecules (Orsini et al. 2016). The cadherin superfamily including *Cdh13* has been linked by multiple studies to numerous neurodevelopmental disorders including ADHD and ASD (Rivero et al. 2015). Independent analysis of different brain regions and time points revealed a large number of differentially expressed genes. Among them were the ADHD candidate genes *Htr2c* (5-hydroxytryptamine (serotonin) receptor 2c) and *Dat1*, which were overexpressed in 6 month old *Adgrl3* null mice. While *Htr2c* was overexpressed in the cortex and hippocampus, *Dat1* was overexpressed in the striatum. *Dat1* is the only candidate gene from the original *Adgrl3* mutant mouse study which was also found to be significantly dysregulated in the whole transcriptome study.

A recently published study reported behavioural and molecular results from a novel *Adgrl3* knockout rat model. Measurements of home-cage activity found *Adgrl3* null rats to be significantly hyperactive in comparison to wild type rats. *Adgrl3* null rats showed an increased acoustic startle response (ASR) but no significant difference in tactile startle responses (TSR). The increased ASR may be indicative of a hyper-reactivity in *Adgrl3* null rats associated with an inability to appropriately filter stimuli (Koch & Schnitzler 1997). The ability to filter sensory stimuli is critical for attention and is frequently impaired in ADHD patients (Holstein et al. 2013). In the open field test, *Adgrl3* null rats showed reduced activity in response to

amphetamine relative to WT rats. No significant differences were found in monoamine, major metabolites or *adgrl1*, *adgrl2* or *flrt3* levels in any of the brain regions tested. Differential protein levels were found for five dopaminergic markers (Upregulated: striatal tyrosine hydroxylase (Th), aromatic L-amino acid decarboxylase (Aadc) dopamine transporter (Dat); Downregulated: dopamine D1 receptor (Drd1) and dopamine- and cAMP-regulated neuronal phosphoprotein (Darpp-32)). Together the results of this paper show a strong similarity with the findings from the mouse model and provide further evidence for the association between *ADGRL3* and ADHD, with a potential causative link through the dopaminergic pathway (Regan et al. 2019)

1.3.5. ADGRL3: Zebrafish

The zebrafish is an increasingly common model organism for biomedical research. Zebrafish are particularly useful for developmental studies as they develop externally and are transparent at larval stages. Morpholino (MO) injection has been used to model *Adgrl3* disruption in zebrafish. *Adgrl3.1* MO larvae displayed a hyperactive phenotype under both light and dark conditions. The increase in total distance travelled was the result of an increase in average speed throughout the trial rather than a reduction in resting period (Lange et al. 2012). As seen in the majority of ADHD patients, methylphenidate (MPH) administration resulted in a reduction in hyperactivity in *dgrl3.1* MO larvae. Another less commonly used ADHD drug, the selective norepinephrine reuptake inhibitor atomoxetine (ATO), was also found to reduce the total distance swam by *Adgrl3.1* MO zebrafish.

In *Adgrl3.1* MO zebrafish, the number of dopaminergic neurons in the posterior tuberculum (PT) were found to be reduced at 3 and 6 days post fertilisation (dpf). The PT is a region of the zebrafish brain which contains neurons that are functionally homologous to those found in regions of the mammalian midbrain such as the ventral tegmental area (VTA) and substantia nigra (SN). At 6 dpf the organisation of the *Adgrl3.1* MO zebrafish PT was severely disrupted across all seven sublayers.

Prenatal exposure to acetaminophen, a common pain-relief medication, is a potential risk factor which has been linked by epidemiological research to an increased risk of developing ADHD. Exposure of *Adgrl3.1* MO larvae to acetaminophen failed to show any effect on levels of hyperactivity in wild-type zebrafish. While this study confirmed the hyperactive phenotype of *Adgrl3.1* MO zebrafish it did not detect any additive or synergistic effects between developmental acetaminophen exposure and reduced *Adgrl3.1* expression. Although the results of this study are not supportive of acetaminophen as an ADHD risk factor, this zebrafish model

remains a powerful potential tool for examining ADHD gene by environment risk factors and future drug treatments (Reuter et al. 2016).

1.3.6. ADGRL3: *Drosophila*

Drosophila melanogaster has greatly contributed to neuroscience research and is a widely used model system (Bier 2005). Fruit flies are small and easy to maintain with a very short generation time which allows large-scale low-cost experiments. In *Drosophila*, the UAS-GAL4 system can be used in conjunction with RNA-mediated interference (RNAi) constructs to drastically reduce expression levels of target genes in specific cell types. The human genome contains three genes each coding for a different member of the latrophilin subfamily of GCPRs, while in Drosophila there is only a single latrophilin gene and protein. This UAS-GAL4/RNAi method was used to create *Latrophilin* knockdown flies which were found to have increased levels of locomotion and reduced time spent sleeping in both the 12h:12h light-dark cycle and 24h dark cycle periods compared to wild types as measured in the *Drosophila* Activity Monitor (DAM) system. Both phenotypes were much more prominent in dark. The presence of a light sensitive hyperactivity phenotype is indicative of a possible disruption of the dopaminergic system. Interestingly, gene knockdown of *dopamine transporter* (*DAT*), resulted in the same phenotypes seen in *latrophilin* knockdown *Drosophila*. The administration of MPH resulted in a significant reduction of hyperactivity and sleepless phenotype in *latrophilin* knockdown *Drosophila* (van der Voet et al. 2015).

1.3.7. ADGRL3: Cross-Model Analysis of Evolutionary Conserved Regions in *ADGRL3*

The effect of protective vs risk allele/haplotypes on ADGRL3 expression was assessed using luciferase assays in four different cell lines. A haplotype in ECR 47 was found to reduce luciferase activity by ≈40% in the B35 neuroblastoma and U87 astrocytoma cell lines. A transgenic zebrafish line was created in which ECR47 was used to drive green fluorescent protein (GFP) expression. ECR47-driven GFP expression was found to share specific patterns with endogenous *Adgrl3.1* expression in the fore, mid and hindbrain while differing in the telencephalon and retina. Limitations of the transgenic tools used to create the zebrafish line prevented a comparison of ECR47 protective and risk haplotypes in zebrafish. Electromobility Shift Assays showed that one of the SNPs (rs2271338) in the ECR47 risk haplotype prevented binding of the Yan Yang 1 (YY1) transcription factor which may impact *ADGRL3* expression levels in a tissue and time specific manner. Expression quantitative trait loci analysis of post-mortem human brain tissue from neuropathologically confirmed controls has linked the

rs2271338 AA risk genotype allele to decreased thalamic expression of *ADGRL3* (Martinez et al. 2016).

1.3.8. ADGRL3: Animal and Cellular Models Summary

Since the initial study linking a common variant of *ADGRL3* to increased ADHD susceptibility in 2010, several animal and cellular models have been created to further our understanding of this gene and its role in the development of ADHD. Many phenotypes reported in these *Adgrl3* models are common across species. All animal models display increased hyperactivity, which is reduced following MPH administration in zebrafish and Drosophila. Neurobiological and molecular findings in ADGRL3 depleted mice and zebrafish indicate an impairment of the dopaminergic system. Abnormalities in the dopaminergic system are commonly associated with ADHD and provide a potential mechanism through which *ADGRL3* variants may impact neurodevelopment and increase the risk of developing ADHD. The evidence provided by the various models, coupled with the numerous genetic studies linking *ADGRL3* variants to ADHD, make it one of the most strongly supported ADHD-candidate genes. The in-depth analysis of the *Adgrl3* constitutive knockout mouse described in this thesis provides further insight into the behavioural and neurobiological aspects of this model, therefore increasing our understanding of not only *ADGRL3*, but also ADHD in general.

2. Methods
All methods were performed with the guidance and support of lab members at the University of Würzburg and Vall d'Hebron Research Institute. Work which was specifically performed by/in collaboration with another researcher is noted below.

The experiments described in methods section 2.1 have previously been reported on in Mortimer et al. 2019.

2.1 *Adgrl3* Knockout Mouse Studies
2.1.1. Establishment of *Adgrl3*-deficient mouse line and confirmation of protein deletion

The creation of the knockout mouse line (B6; 129S4-Lphn3Gt (FHCRC-GT-S17-5H1)Sor) investigated in this thesis has previously been described in detail in a paper by Wallis et al., 2012. In order to establish the mouse line at University of Würzburg's Centre for Experimental Molecular Medicine (ZEMM), frozen sperm from *Adgrl3*$^{-/-}$ male mice was shipped from the Mutant Mouse Regional Resource Centres at The Institute for Genomic Medicine, Texas A&M to Würzburg. The frozen sperm was then implanted in C57BL/6J dams by *in vitro* fertilization.

The group at Texas A&M confirmed disruption of the *Adgrl3* mRNA transcript using RT-PCR probes. In order to confirm that this genetic disruption resulted in an abolition of Adgrl3 protein we used Western blots to directly measure protein levels in the mouse brain. Western blots were performed by Olga Rivero at the University of Würzburg. Samples were collected from the cortex and hippocampus of all three *Adgrl3* mice genotypes (*Adgrl3*$^{+/+}$, *Adgrl3*$^{+/-}$ and *Adgrl3*$^{-/-}$). The tissue samples were homogenized by sonication. Protease and phosphatase inhibitor tablets (Roche Diagnostics Deutschland GmbH, Mannheim, Germany) were dissolved in RIPA (Radio-Immunoprecipitation Assay) buffer (Sigma-Aldrich Chemie Gmbh, Munich, Germany) which along with sonication promoted cell lysis whilst avoiding protein degradation. Total protein levels were quantified and each Western blot sample was composed of 30µg of total protein. Samples were loaded on 4-12% Bis-Tris polyacrylamide gels (Life Technologies, Darmstadt, Germany) and electrophoreses was performed under reducing and denaturing conditions. Following electrophoreses, samples were transferred to a 0.45 µm pore nitrocellulose membrane by electroblotting (Life Technologies). Membranes were probed using a sheep anti-human ADGRL3 (labelled LPHN3) antibody (1:400, R&D systems, Wiesbaden-Nordenstadt, Germany) and a loading control of mouse anti-α-tubulin (1:5000, Sigma-Aldrich) antibody. Following primary antibody incubation, membranes were incubated with the following secondary antibodies: fluorescent donkey anti-sheep Alexa Fluor 790

(1:15000, Jackson Immunoresearch, Cambridgeshire, UK) and donkey anti-mouse 690 (1:10000, Li-Cor Biosciences GmbH, Bad Homburg, Germany). Fluorescence detection was performed on the Odyssey infrared imaging system using Image Studio Lite 3.1 software (Li-Cor Biosciences). Arbitrary fluorescent units were normalized against the internal loading control of tubulin. Adgrl3 relative protein levels were expressed as a percentage of average control values.

2.1.2. Animal housing conditions

Animals were housed under controlled conditions in the Animal Core Facility of the ZEMM Mice were accommodated at $21\pm1°C$ and $50\pm5\%$ humidity under a 12 h light and 12 h dark cycle. During the continuous performance test (CPT), food was restricted as described in the CPT testing protocol. During all other times, food and water was provided *ab libitum*. All mice were stored in Makrolon type III cages under group-housing conditions except for male mice prior to testing in the resident intruder paradigm; these mice were single housed for a minimum of three weeks prior to testing. All experiments and housing conditions were designed to inflict the minimum amount of stress and suffering possible on the mice. Housing and testing was carried out in accordance with the European Community guidelines for animal care and use. All protocols involving animals were approved by both the University of Würzburg board and the Government of Lower Franconia (license 55.2-2531.01-30/14). The behavioural testing began when mice were 6 weeks old and finalised on week 13, except for the CPT which was performed on 6 to 9 month old mice, which had previously been behaviourally tested. Tests were performed from the least strenuous to most strenuous paradigms to reduce animal suffering and stress. There was a minimum of a week between the start of behavioural tests. In order to ensure uniformity of results, all tests were performed between 10:00-18:00 during the mouse light phase.

2.1.3. Behavioural paradigms

Adgrl3 deficient mice are primarily a model for ADHD, therefore all behavioural tests were carried out in order to assess behavioural features associated with ADHD or other relevant comorbidities. The experimenter handled all animals daily from birth to ensure habituation. An initial battery of behavioural tests was performed on *Adgrl3*[+/+] (n=28, 12 ♂ and 16 ♀), *Adgrl3*[+/-] (n=20, 10 ♂ and 10 ♀) and *Adgrl3*[-/-] (n=19, 11 ♂ and 8 ♀) mice. Emotional dysregulation including anxiety-like behaviour was assessed using the standard light-dark box (Crawley and Goodwin, 1980; Onaivi and Martin, 1989). Anxiety like behaviour was also assessed in the open-field test, though the primary role of this testing paradigm was to examine hyperactivity

(Post et al., 2011; Seibenhener and Wooten, 2015). Memory and learning was assessed in the novel object recognition task and Barnes Maze, with the Barnes Maze particularly focused on spatial memory (Leger et al., 2013; Lueptow, 2017; Pitts, 2018).). Following an absence of clear phenotype in *Adgrl3$^{+/-}$* mice, further tests were performed on *Adgrl3$^{+/+}$* and *Adgrl3$^{-/-}$* mice only. While the open field test assessed gross level abnormalities in movement, Gaitlab analysis allowed detection of fine level locomotor deficits (*Adgrl3$^{+/+}$* n=13, 7 ♂ and 6 ♀ and *Adgrl3$^{-/-}$* n=12, 8 ♂ and 4 ♀) (Hetze et al., 2012). Sociality including social memory and aggression were investigated using the social interaction test (*Adgrl3$^{+/+}$* n=12, 8 ♂ and 4 ♀ and *Adgrl3$^{-/-}$* n=12, 9 ♂ and 3 ♀) and the resident-intruder paradigm respectively (All male: *Adgrl3$^{+/+}$* n=10 and *Adgrl3$^{-/-}$* n=9) (Koolhaas et al., 2013; Moy et al., 2004). Impulsivity and attention were investigated using the continuous performance test (CPT) which strongly resembles the testing paradigm utilised in human subjects (All male: *Adgrl3$^{+/+}$* n=7 and *Adgrl3$^{-/-}$* n=9) (Caballero-Puntiverio et al., 2019).

2.1.4. Light-dark box

This test utilises both light (2/3 of total space) and dark (1/3 of total space) areas separated by an insert. It aims to examine anxiety like behaviour through the opposing motivations to explore novel spaces versus avoid brightly lit spaces. The LDB test is carried out in an opaque white box (50×50×40 cm) which is semi-permeable to infrared light (TSE Systems, Bad Homburg, Germany). Illuminance in the light portion of the box is approximately 100 lx while the dark portion has an illuminance of under 5 lx. Testing started with the mouse in the dark section and continued for 10 mins whilst being recorded by an infrared-sensitive CCD camera controlled by VideoMot2 software (TSE Systems) and located over the centre of the box. Parameters measured included the time spent in each compartment, latency to enter the lit compartment and the total distance travelled.

2.1.5. Open-field

In addition to the LDB, the open field (OF) test was also used to assess anxiety like behaviour, although its primary function was to measure locomotion/hyperactivity levels. The OF test is carried out in an opaque quadratic box ($50 \times 50 \times 40$ cm) with an illuminated floor; illuminance went from 50 lx at the walls to 100 lx at the centre. The test began with a mouse placed in one corner of the box close to the walls and continued for 30 min. Automatic recording of mouse movements was performed by an infrared-sensitive CCD camera, positioned above the centre of the box and controlled by VideoMot2 software (TSE Systems). Parameters analysed included the portion of time spent in the centre area in comparison to the edges as a measure

of anxiety-like behaviour and both the total distance travelled and the duration of vertical rears as measures of locomotor activity.

2.1.6. Gaitlab

The Gaitlab system developed by Viewpoint Behavioural Technology (Lyon, France) provides a fine-level quantitative gait analysis for rodents (Chedly et al., 2017). This technology is based on the CatWalk Method (Caballero-Garrido et al., 2017). Mice spent four days training to run smoothly over a narrow glass corridor (7×90 cm). Four trials followed the training sessions, during which the gait and ambulatory movements of each mouse were recorded using a high-speed camera (150 frames/s). Trials in which the position of mouse paws or movement speed were unclear were discarded from the Gait analysis. Stride length was defined as the vertical distance moved by a paw. The time during which a paw was in contact with the ground was defined as stance time while conversely the swing time was defined as the time during which a paw is in the air. Hind/rear limb gaps refers to the vertical distance between a left-right pair of limbs, while the horizontal distance between the limb pairs was defined as the respective limb pairs base of support (BOS). The data from the high-speed camera was assessed using the specialised Gaitlab analysis software.

2.1.7. Novel object recognition task

Mouse memory and ability to discriminate between objects was assessed in the novel object recognition (NOR) task. During the first NOR trial (T1), two identical aluminium counters were placed in opposite corners of the OF box. 24 h later the second NOR trial (T2) was performed. During T2, a transparent glass bottle with a blue lid replaced one of the counters, the counter being replaced alternated between mice to prevent bias in results as a result of side preference. Olfactory cues were removed by cleaning both the objects and the OF box with 70% ethanol after every trial. Automatic recording of mouse movements was performed by an infrared-sensitive CCD camera, positioned above the centre of the box and controlled by VideoMot2 software (TSE Systems). The amount of time mice spent in the immediate area around each object was recorded. In the second trial, the difference between the time spent interacting with the novel object and the time spent interacting with the familiar object was defined as the preference level for the novel object. The following data were assessed; objection interaction in each trial, preference for the novel object in T2 and total distance travelled in each trial.

2.1.8. Barnes maze

Mice's visuospatial learning and memory were assessed in the Barnes maze (BM). The BM is composed of a 120 cm diameter circular grey platform containing 40 holes, each 5 cm in diameter located close to the platforms edge. The escape box is attached to one of the holes. On the walls surrounding the platform there were visual cues to allow the mice to spatially orientate. Prior to the start of each trial, a mouse was placed in the middle of the platform inside a tall cylinder which obscured the mouse's view of the platform and walls, therefore preventing orientation prior to trial commencement. Once the mouse was released from the cylinder, the trial began and lasted until the mouse entered the escape box or 3 min elapsed. The first phase of testing known as the acquisition phase involved 10 trials over 4 days during which the escape was in the same position each time. $Adgrl3^{+/-}$ mice did not proceed to the reverse learning phase which followed the acquisition phase. During the reversal phase, the position of the escape box was changed and $Adgrl3^{+/+}$ (n=19) and $Adgrl3^{-/-}$ (n=13) mice were trained over 6 trials in 2 days to identify the new escape hole. Automatic recording of mouse movements was performed by an infrared-sensitive CCD camera, positioned above the centre of the box and controlled by VideoMot2 software (TSE Systems). The distance travelled was automatically recorded as the latency to escape which signified the end of the trial. The number of primary errors, defined as the number of times a mouse poked its head into an incorrect hole, was manually scored.

2.1.9. Social interaction

Sociability and desire for social novelty were assessed in the social interaction (SI) test. Similarly to the NOR test, the SI test was also performed in the OF box. A novel cohort of mice were used for the test. Immediately prior to testing, mice were allowed to habituate to the OF for 10 min. Each trial was also 10 min long. During the first trial known as the sociability test (T1) mice were exposed to two identical small wire cages in opposite corners of the OF, one of which was empty while the other contained an unfamiliar wildtype mouse. The mouse-containing and empty cage position were alternated between trials to prevent possible effects of side preference. The second trial known as the social novelty trial (T2) was performed 24 h later. During this trial mice were again exposed to the mouse from the first trial (familiar mouse) and now also to a second novel, unfamiliar mouse. The same small wire cages were used in T1 and T2. Automatic recording of mouse movements was performed by an infrared-sensitive CCD camera, positioned above the centre of the box and controlled by VideoMot2 software (TSE Systems). Sociability was defined during T1 as the difference between time spent interacting with the wild-type mouse and the time spent interacting with the empty cage. During T2, social novelty was defined as the difference between the time spent interacting with

the novel mouse and the time spent interacting with the familiar mouse. Alongside interaction times, the total distance travelled by mice was also recorded in both trials.

2.1.10. Resident-intruder

The resident-intruder paradigm is designed to investigate aggression as a result of territorial intrusion. As male mice are more aggressive than female mice, the resident-intruder paradigm is performed exclusively on male mice (Koolhaas et al., 2013). In order to induce strong territorial behaviour, resident mice were housed by themselves for a minimum of three weeks prior to testing. While single housing is stressful for mice, the isolation is reduced by the mice's ability to hear, smell and partially view other mice housed in the same room. All novel intruder mice were wildtype and group-housed under standard conditions. Resident mice were paired with a smaller intruder mouse in order to increase the probability of eliciting aggressive behaviour towards the intruder. Trials were 10 min long from the introduction of the intruder mouse and were recorded on a Logitech webcam and Windows computer system. A physical struggle initiated by the resident towards the intruder mouse was defined as an attack. Each trial was scored for the presence or absence of an attack, the time until the first attack and the total number of attacks.

2.1.11. Continuous performance test (CPT)
Animals

During touchscreen training, mouse body weight was restricted to 85-90% of free-feeding body weight by food restriction. In order to ensure no excessive weight loss occurred, mice were weighed daily. Water was available *ad libitum* at all times.

Apparatus

Touchscreen operant chambers (Campden Instruments Ltd, Salford, UK) were used to conduct the CPT. All chambers composed of a single touchscreen on one wall opposite a liquid delivery magazine on the opposing wall, a house light and tone generator with a metal grid floor. Sound-attenuated ventilated boxes were used to prevent noise from surroundings interfering with the test. Whisker operating system (Cardinal and Aitken, 2010) and ABET II Touch software (Lafayette Instrument, Lafayette, USA) controlled the touchscreen chambers as well as recording mouse behaviour.

Continuous performance test (CPT)

The CPT was performed by Aet O'Leary at Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital of Frankfurt. The CPT involves a two session habituation stage followed by three training stages and finally four probe sessions. Each mouse was trained in the same touchscreen chamber throughout the CPT. No more than one session was performed per day.

Habituation stage: The first stage of CPT involved habitation to the touchscreen chamber over two 20-min sessions. The reward delivery tray was filled with 600 µl of strawberry milkshake to also allow mice to habituate to the reward during these sessions, (Müllermilch Erdbeer, Molkerei Alois Müller GmbH & Co. KG, Fischach-Aretsried, Germany). During the entire session the house light remained off. In order to advance to the training stage mice must consume all of the reward within the 20-min session.

Training Stage 1 – Touch: During the first training stage, mice were exposed to the response area of a white square (3.5 x 3.5 cm) on the touchscreen inside a white frame located in the horizontal centre, 2 cm from the grid floor. At the beginning of the trial, 20 µl of the reward were added to the illuminated reward tray. Once the mouse entered the reward tray, the light was switched off. The mouse was given 2 s to consume the reward before the 3 s inter-trial interval (ITI) began. The white square was displayed in the response area for 10 s following the ITI. During stage 1, a hit was recorded if the mouse touched the response area during the limited hold (LH) period of 10 s whilst the stimulus was displayed or 0.5 s after the stimulus. Across all CPT stages the LH was defined as the stimulus duration plus 0.5 s. If the hit occurred while the stimulus was still present, it was removed and 20 µl of the reward were delivered to the illuminated tray along with a 1s tone at 1000 Hz. Once the mouse's head entered the tray the light was turned off, there was a 2 s pause for consumption and another 3 s ITI began. Failure to touch the response during the LH was defined and recorded as a miss and no reward was administered. The ITI was restarted if the response area was touched prior to the end of the ITI. The total session lasted until either 45 min elapsed or 100 rewards were administered. Mice advanced to the next stage if they successfully earned 60 or more rewards during the 45-min session.

Training Stage 2 – S+ introduction: During training stage 2, mice were exposed to the S+ target stimulus of a black and white image of horizontal stripes (3.5 x 3.5 cm) which replaced the white square. S+ stimulus exposer lasted only 5 s. The reward was administered or not

administered in response to a hit or miss respectively, in the same manner as during training stage 1. The criterion for advancement remained the same.

Training Stage 3 – Two-Stimulus Discrimination: Training stage 3 involved the introduction of a novel non-target stimulus, a black and white image of a snowflake (S-), which was presented in the response area 50% of the time. Alongside the introduction of a non-target stimulus the duration of stimulus exposure was shortened further to 3s. Hits were recorded and rewarded only if the mouse touched the S+ within the LH. A correct rejection (CR) was recorded if the mouse did not touch the screen during the LH of an S- exposure. CR was followed by an ITI without reward administration. A touch response to S- exposure constituted a false alarm. In response to a false alarm, the S- was removed and another ITI began, correction trials involving the S- repeated until the mouse performed a correct rejection. Data from correction trials were excluded from analysis. Training stage 3 sessions ended when either 100 rewards were administered or 45 min elapsed. All mice completed a minimum of seven training stage 3 sessions.

Performance sensitivity (d') was defined as the ability to discriminate between S+ and S- as defined in the CPT data analysis section. A performance sensitivity (d') over 0.6 in two consecutive session was required for inclusion in baseline recording. Two *Adgrl3*$^{-/-}$ mice failed to reach a d' of 0.6 and were therefore excluded.

Baseline responding was defined as the mean of the last two sessions from training stage 3. Once d' criterion was reached, animals proceeded to probe sessions with greater task difficulty. Each probe session was followed by two consecutive sessions under training stage 3 conditions to ensure mice reverted to baseline responding levels.

Probe 1 – Stimulus duration: The first probe was carried out over two sessions during which the duration of stimulus exposure was shortened by varying degrees to increase attentional load. Stimulus durations during the first session varied in a pseudorandom pattern from 1 s to 2 s in 0.5 s intervals (1, 1.5, and 2s). Stimulus durations were shortened to 1, 0.75, 0.5, and 0.25s during the second session. The level of performance at 1 s was taken as the mean for this stimulus interval over both sessions.

Probe 2 – Stimulus Contrast: During probe 2, the ability of mice to perform in challenging perceptual conditions was assessed by varying the contrast of the stimuli displayed. While total luminance was kept constant, the black areas' luminance was reduced while white area luminance was increased. Stimuli contrast varied across 12.5, 25, 50, and 100% contrast.

<u>Probe 3 – Long ITI</u>: Probe 3 investigated the ability of mice to control impulsive responding by increasing the ITI interval from 3 s to 7 s.

<u>Probe 4 – S+ probability</u>: Probe 4 assessed performance differences when the probability of S+ presentation was reduced from 50% to 30%.

<u>CPT data analysis</u>. During training stages 1-3, the number of sessions required to reach advancement criteria was recorded. For each of the sessions during training stage 3 and the probe trials, the following parameters were determined: Hit rate (HR=Hits/(Hits+Misses)), False alarm rate (FAR=False alarms/(False alarms+Correct rejection)), sensitivity d' (d'=z(HR)-z(FAR)) and response bias c, defined as the tendency to favour one way of responding to stimuli over another; (c=0.5[z(HR)+z(FAR)]). During training stage 3, the number of ITI touches of the response area were also recorded, as was the latency to reward retrieval.

2.1.12. Behavioural data analysis

Most behavioural data was initially analysed for genotype effects by two-sided univariate analysis of variance (ANOVA). Exceptions were the analysis of genotype differences in baseline level reward retrieval latency during the CPT which was investigated by t-test and the difference in presence/absence of attack rates across genotypes during the resident-intruder trial which was analysed using Fisher's exact test. The threshold for statistical significance was defined as $p < 0.05$. Following identification of statistically significant differences by ANOVA, *post-hoc* comparison between groups were performed using adjusted p-value thresholds for multiple correction. The relevant *post-hoc* test utilised is reported for each analysis. Behavioural tests which involved multiple trials were analysed using a repeated measures ANOVA, with genotype as the between-subject factor while time or testing phase was the repeated measure. Parametric data assumptions were met for all behavioural data and both statistical analysis and graphing was performed using GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA).

2.1.13. Prefrontal cortex, hippocampus and striatum dissection and RNA-sequencing

The prefrontal cortex, hippocampus and striatum were selected for RNA-sequencing on the basis of their relevance to ADHD. Behaviourally naïve mice between P65 and P76 (mean=P67.15, SD=3.46) were chosen for sequencing. In total, 24 mice evenly distributed between *Adgrl3*$^{+/+}$ and *Adgrl3*$^{-/-}$ genotypes were sequenced. Mice were euthanized with isoflurane followed by cervical dislocation. Immediately following animal sacrificing, brains

were flash frozen in isopentane which was cooled with dry ice. Tissue was then stored at −80 °C until dissection on a pre-cooled plate. Brains were dissected by Tatjana Ganster, a collaborator at the University of Würzburg. The chosen brain regions were segregated and RNA was isolated using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Hilden, Germany). The commonly used Illumina protocol (Illumina GmbH, Munich, Germany) was followed for cDNA library preparation. Libraries were sequenced by the Core Unit Systems Medicine at the University of Würzburg using the Nextseq sequencing platform with High Output 2×75 cycle kits that generated 75bp long single end reads. Following exclusion of the original cDNA templates approximately 400 million raw reads remained for further processing and analysis.

2.1.14. RNA-Seq data analysis pathway

Data analysis was performed in collaboration with a bioinformatics researcher from Core Unit Systems Medicine at the University of Würzburg. FASTQC 0.11.4 was used to assess raw read quality, duplicate numbers and the presence of adapter sequences. Once detected, cutadapt 1.16 was used to remove Illumina TruSeq adaptor sequences and refined reads were further refined whilst maintaining a quality drop value below a mean of Q20 (Martin, 2011). The STAR 2.5.2b aligner specifically designed for short reads was used to align these processed sequences with the murine genome (mm16, GRCm38) using genome and annotation files from GENCODE (Dobin et al., 2013). The proportion of reads which mapped to the reference genome was $90\pm1\%$ for all samples. BEDtools 2.25.0 subcommand intersect was used to align sequences to specific genes and quantify these alignments (Quinlan and Hall, 2010). Read counts were globally normalised followed by analysis of each brain region independently using DESeq2 1.16.1 to detect genotype dependent differential expression of genes (Love et al., 2014). The threshold for differential gene expression was set using Benjamini-Hochberg correction for multiple testing to $p<0.05$. Gross level data was graphed post-normalisation using a MA plot created by DESeq2 function plotMA. The NCBI Gene Expression Omnibus contains all RNA-Seq data produced in this thesis and is available for access through the GEO series accession number GSE117357 (https://www.ncbi. nlm.nih.gov/geo/query/acc.cgi?acc=GSE117357) (Edgar et al., 2002).

2.1.15. Gene set enrichment analysis (GSEA) and pathway analysis

Following detection of differentially expressed genes, gene set enrichment analysis (GSEA) and pathway analysis was performed to aid interpretation of the biological relevance of genes. Enricher was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) based

GSEA (Chen et al., 2013; Kuleshov et al., 2016). For this analysis, genes with a differential gene expression of less than 50% magnitude were excluded. Gene sets were analysed separately based on brain region and the direction of differential expression. Analysis was restricted to PFC downregulated genes are this was the only gene set with a size greater than the pre-determined minimum requirement of 10 genes. Statistically significant over-representation was assessed using the Fisher's Exact Test with imputation-based correction for multiple testing.

Annotated terms associated with biological functions and disease pathways which contained an over-representation of differentially expressed genes were identified using the Ingenuity Pathway Analysis software (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/ products/ingenuitypathway-analysis). All significantly differentially expressed genes in a brain region were included in a single analysis as IPA software is capable of taking account of differences in the magnitude and direction of differential gene expression as represented by log fold changes (logFC). The threshold for statistically significant enrichment was set at 0.05 following Benjamini-Hochberg correction of the Fisher's Exact Test. Identified pathways required at least 2 differentially expressed genes to be considered in order to avoid spurious results.

In addition, IPA was used to construct the top 25 gene networks using both the direction and magnitude of gene's logFC. All networks with a score (-log10(p-value)) greater than 8 were considered statistically relevant. IPA also investigated the representation of differentially expressed genes in a list of previously associated ADHD genes. This previously published list contains 436 ADHD associated genes identified in the ADHD gene database (http://adhd.psych. ac.cn/index.do) and from a comprehensive search of reviews on ADHD genetic and pharmacogenetics (Pagerols et al., 2018).

2.2 Human ADHD Transcriptomic and Genomic Study Methods

2.2.1. Clinical Sample Collection

All patients included in the ADHD sample were required to fully satisfy the DSM-IV criteria for ADHD diagnosis after evaluation using the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I & II) and Conners´ Adult ADHD Diagnostic Interview for DSM-IV (CAADID-I & II). Further description of the tests used for clinical assessment is available in Ribasés et al (2009). The control sample was composed of individuals who neither currently nor in the past had a diagnosis of ADHD nor showed any signs of ADHD symptomatology such as trouble paying attention or persistent fidgeting. All individuals were at least 18 years old, born in Spain and of Caucasian ethnicity. Exclusion criteria included an IQ below 70 or a diagnosis of a pervasive developmental disorder, schizophrenia or other psychotic disorder. Any significant neurological or systemic disease which may explain ADHD symptoms was also grounds for exclusion. The presence of comorbid oppositional defiant disorder, conduct disorder, depression and anxiety disorders were not permitted nor was any current or past history of neurologic, metabolic, cardiac, liver, kidney, or respiratory disease. The Ethics Committee of the Hospital Universitari Vall d'Hebron approved the study and all tests were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained in written form from all individuals included in the study in accordance with the Helsinki Declaration. Admission of study participants and collection of blood samples was performed by clinical staff at University Hospital Vall d'Hebron

2.2.2. RNA isolation

The initial hypothesis-free microarray analysis included blood samples of 94 medication-naïve ADHD individuals (60.6% male with a mean age of 34.8 years, sd=11.29) and 124 control individuals (55.6% male with a mean age of 36.7 years, sd=10.0). The real-time quantitative Polymerase Chain Reaction (qPCR) replication sample included 30 medication-naïve ADHD individuals (60% male with mean age of 34.0 years, sd=11.6) and 29 control samples (56.6% male with a mean age of 43.9 years, sd=16.7). The second independent microarray analysis was carried out on 49 medication-naïve ADHD individuals (57.1% male with a mean age of 30.3 years, sd=11.2) and 45 control individuals (55.6% male with a mean age of 51.7 years, sd=21.4). Total RNA was extracted from peripheral blood mononuclear cells (PBMCs) using the Ficoll density gradient method and the RNeasy Midi kit (Qiagen, Hilden, Germany). RNA quality was assayed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and reported as a RNA Integrity Number (RIN). RNA samples were isolated in collaboration with other researchers at Vall d'Hebron Research Institute

2.2.3. Hypothesis free gene expression analysis of peripheral blood mononuclear cells (PBMCs) using microarrays

Following RNA extraction and isolation, samples were reverse transcribed and amplified using the Ambion WT Expression Kit (Life Technologies, Carlsbad, CA, USA). The cDNA produced was subsequently fragmented and labelled using the GeneChip WT Terminal Labelling and Hybridization Kit (Life Technologies, Carlsbad, CA, USA). Samples were hybridized to the Genechip Human Gene 1.1 ST 96-Array plate for microarray analysis (Life Technologies, Carlsbad, CA, USA). The experiment included samples from three independent batches containing 8-9 amplification plates each. The Gene Titan Affymetrix microarray platform was used for microarray data generation and processing. The Robust Multichip Average (RMA) function in the oligo R package was used to background correct, normalize and summarize probe values for each batch independently. Probes which did not match to genes were removed and the normalised gene expression matrix was extracted. The gene expression data of the three batches was then combined while using the Combat R-package to take account of the batch effect. Probes matching to genes on the X or Y chromosome were removed and the difference in gene expression between ADHD cases and controls was determined using the linear model features of the limma R-package including gender and RIN as covariates. Benjamini-Hochberg correction was applied for multiple correction analysis on the 19184 probes used in the final model.

2.2.4. Gene-Set Enrichment and Pathway Analysis

The results of the microarray analysis were analysed for biological relevance using the pathway analysis software, IPA [Ingenuity Pathway Analysis software (IPA) (Ingenuity Systems, Redwood City, CA, USA]. All genes with a nominally significant p-value in the microarray analysis were considered differentially expressed in the pathway analysis. IPA was used to test for an over-representation of differentially expressed genes in annotated terms representing various biological and disease pathways. The Fisher's Exact Test followed by Benjamini-Hochberg correction to a threshold of 0.05 was used to determine statistically significant pathway enrichment. The top 25 gene networks were also constructed by IPA taking into account the direction and degree of genes´ log fold changes. Gene networks with a network score (-log10(p-value)) over 8 were considered statistically significant. IPA was also used to test for an over-representation of differentially expressed genes which were previously implicated in ADHD. The 436 genes included in the list of ADHD associated genes was compiled from the combination of the ADHDgene database (http://adhd.psych.ac.cn/index.do)

and a comprehensive search of published reviews of ADHD genetic and pharmacogenetic studies. The full list of genes is available in Pagerols et al. (2018) supplemental information 1.

2.2.5. Replication of a subset of differentially expressed genes using reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Following the microarray analysis the four genes with the most statistically significant differences in expression between ADHD cases and controls were chosen for further study. These genes were *Lysine Methyltransferase 5A (KMT5A), Nuclear RNA Export Factor 1 (NXF1), Kruppel Like Factor 4 (KLF4)* and *LRR Binding FLII Interacting Protein 1 (LRRFIP1)*. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was used to conduct the gene expression replication study using a novel clinical sample of 30 ADHD cases and 29 controls. Up to 2 µg of total RNA were reverse transcribed using the High Capacity cDNA Reverse Transcription Kit following the manufacturer's guidelines [Applied Biosystems; Foster City, California, USA]. qPCR reactions were run in triplicate using 4 µL of cDNA with a concentration of 9.75 ng/µL, 0.5 µL of RNAse free water, 5 µL of TaqMan® Gene Expression Master Mix and 1 µL of the relevant TaqMan Gene Expression Assay, in a final volume of 10 µL per reaction [Applied Biosystems; Foster City, California, USA]. The amplification efficiency of each gene assay was calculated using the dilution series standard curve method. cDNA was amplified over 40 cycles and the fluorescence data recorded using an Applied Biosystems 7900HT Fast Real-Time PCR system set at the default thermal cycling conditions specified by the manufacturer [Applied Biosystems; Foster City, California, USA]. In total six qPCR plates were included in the experiment. The threshold cycle (CT) was defined automatically for each plate using the Applied Biosystems RQ manager 1.2.1 software. Triplicate technical replicates which showed a standard deviation in CT value greater than 0.3 were flagged and the outlying replicate was removed from the data. If the remaining two replicates retained a standard deviation greater than 0.3, the sample was removed from the analysis. The comparison of gene expression between ADHD cases and controls was performed using the generalized linear model function of the MCMC.qpcr R package. *PES1* was chosen as a reference gene after checking its stability and linearity in the initial microarray analysis and prior qPCR experiments. *PES1* was included as a Bayesian prior in the linear model alongside assay amplification efficiencies while qPCR plate, gender and RIN values were included as covariates. The statistical association test was two-sided and Bonferroni correction was applied to control for multiple-testing.

2.2.6. Replication of significantly differentially expressed genes with second independent microarray samples

Due to restrictions on available sample quantities, it was not possible to perform qPCR experiments for all 21 genes found to be significantly differentially expressed in the first microarray. In order to investigate all 21 genes we used data from a second microarray composed of independent ADHD and control samples. The same microarray data processing pathway was followed for the second replication microarray however analysis was restricted to the 21 genes found to be significantly differentially expressed in the first microarray. The threshold for statistical significance was set using Benjamini-Hochberg correction, taking into account these 21 genes. The correlation between microarrays for all probes' logFC for case versus control differential expression was assessed using a Pearson correlation test.

2.2.7. Combination of expression Quantitative Trait Loci (eQTL) in blood from a publicly available database and log fold change data from our microarray analysis to form a Polygenic Risk Score

We performed a polygenic risk score (PRS) analysis on all genes for which microarray information was available and a subset analysis for the 8 genes which were replicated in the secondary microarray replication study. Using the publicly available Blood eQTL database we identified eQTLs in 6746 of the 15795 genes for which gene expression was available from the microarray analysis (Westera et al. (2013). The blood eQTL database contained 432854 eQTLs for these genes. For the 8 replicated genes, we identified eQTLs for 4 (*KMT5A*, *IL7R*, *SLA* and *EGR2*). In total, these genes had 716 eQTLs linked to them.

Independent SNPs were obtained by clumping using PLINK 2.0. SNPs in linkage disequilibrium ($r^2 > 0.2$) with a more statistically significant SNP within 250kb were removed (Chang et al. 2015). The largest available genome wide association study of 356 ADHD cases and 1414 controls was used as the reference genome for clumping, prior to analysis all genotyped markers underwent strict quality control according to a standardized pipeline.

For each of these independent SNPs, a combined z-score was determined by adding a z score for eQTL (SNP effect on gene expression) and a z-score for gene expression (Gene expression difference between ADHD cases and controls). The sign of both the eQTL effect and the gene expression difference are taken into account to identify the risk allele which increases the potential risk of ADHD. The z-combined value provided the weighting applied to each SNP for the polygenic risk score analysis (PRS).
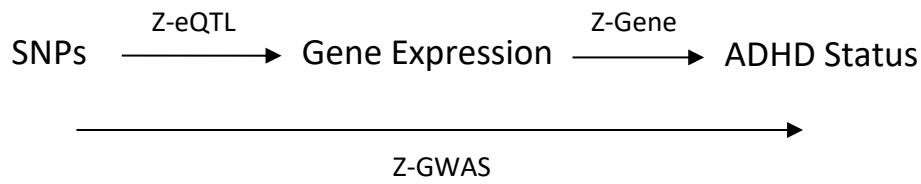
Figure 2. Graphical description of the model used to weight SNPs for polygenic risk score analysis.

PRS analysis was performed using PRSice-2 (https://choishingwan.github.io/PRSice/) on adult samples from three independent GWAS studies. The first dataset (GWAS1) contained 417 ADHD Cases (68.82% male, with a mean age of 32.9 years, sd=10.5) and 428 controls (76.4% male, with a mean age of 43.9, sd=14.2). GWAS2 was composed of 106 ADHD cases (66.98% male, with a mean age of 32.1 years, sd=10.6) and 77 controls (51.95% male, with a mean age of 46.3 years, sd=11.7) and GWAS3 included 335 ADHD cases (68.96% male, with a mean age of 32.2 years, sd=10.8) and 1356 controls (54.72% male, with a mean age of 54.7 years, sd=16.2).

Each SNP had its corresponding eQTL p-value assigned to it and default p-value thresholds were used to construct PRSs that were tested for association with ADHD status. A meta-analysis of the PRS results from the three GWAS studies was performed at each p-value threshold to increase the statistical power of the analysis.

3.   Results

## 3.1 *Adgrl3* Knockout Mouse Model Results
The results described below have previously been reported in Mortimer et al. 2019

3.1.1. *Adgrl3*$^{-/-}$ mice show a complete absence of ADGRL3 protein while ADGRL3 levels are partially reduced in *Adgrl3*$^{+/-}$ mice

Western Blot analysis of prefrontal cortex lysates from *Adgrl3*$^{+/+}$ mice clearly show a dense band corresponding to the p120 subunit of *ADGRL3*. *Adgrl3*$^{+/-}$ mice lysates showed an intermediate density band while no band was detected in the lysates from *Adgrl3*$^{-/-}$ mice (Fig. 3). These results clearly show *ADGRL3* protein levels are controlled in a *Adgrl3* gene dosage dependent manner.

3.1.2. Anxiety-like behaviour as investigated in the light-dark box is unaffected by *Adgrl3* inactivation

Analysis of *Adgrl3*$^{+/+}$, *Adgrl3*$^{+/-}$ and *Adgrl3*$^{-/-}$ mice failed to show any significant genotype effect in two common measures of anxiety-like behaviour; latency to enter the light region (Fig. 4a) and total time spent in the light region (Fig. 4b), (Latency $F_{(2,65)}=1.958$; $p=0.1493$, Time $F_{(2,65)}=0.1617$; $p=0.8510$). In contrast, there was a highly significant genotype effect on the global distance travelled (Fig. 4c) during the LDB ($F_{(2,65)}=9.033$; $p=0.0003$). *Post-hoc* analysis showed *Adgrl3*$^{-/-}$ mice had increased locomotor activity in comparison to both *Adgrl3*$^{+/+}$ and *Adgrl3*$^{+/-}$ mice which were not significantly different to one another (*Adgrl3*$^{-/-}$ vs *Adgrl3*$^{+/+}$ ($p=0.0004$) and *Adgrl3*$^{-/-}$ vs *Adgrl3*$^{+/-}$ ($p=0.0043$), Tukey's multiple comparison test).
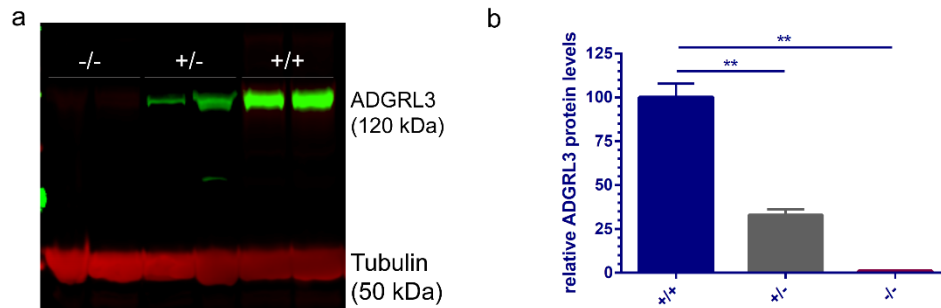
Figure 3. Western Blot Analysis: ADGRL3 levels from mouse brain prefrontal cortex. (a) Lysates shows a prominent band of approximately 120 kDa in *Adgrl3⁺/⁺* animals, which corresponds to the p120 subunit of ADGRL3 protein. **(b)** Tubulin was used as the internal loading control to normalise arbitrary fluorescent units. Relative protein levels were expressed as percentage of average *Adgrl3⁺/⁺* expression levels. The 120 kDa protein band was undetectable in *Adgrl3⁻/⁻* mice, while an intermediate strength band was visible in lysates from *Adgrl3⁺/⁻* mice. Error Bars represent Standard Error of the Mean (SEM). Data was analysed by one-way ANOVA followed by Tukey's multiple comparison test; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.
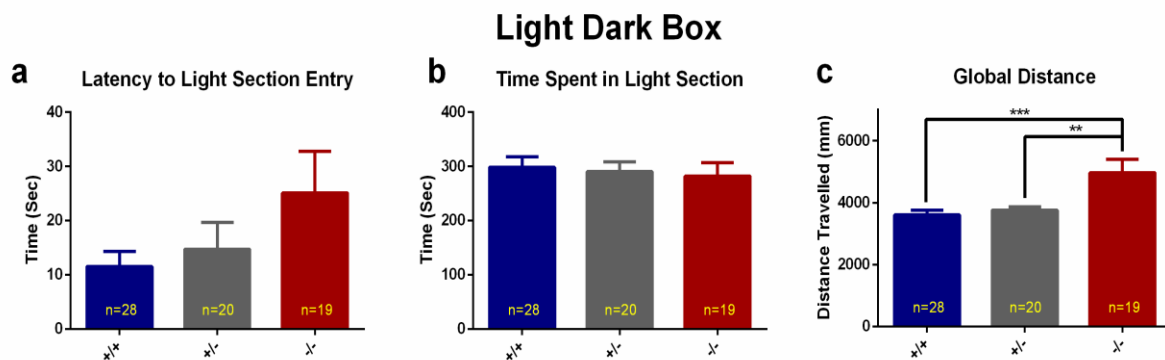
Figure 4. Light Dark Box: No significant effect of genotype on mouse anxiety levels (a) The latency to enter the light region was not significantly different between mouse genotypes **(b)** nor was a significant genotype effect found on the total amount of time mice spent in the light section. **(c)** A significant overall genotype effect was found on the total distance travelled by mice with *Adgrl3⁻/⁻* mice travelling a significantly greater distance than heterozygous or control mice. Error Bars represent Standard Error of the Mean. Data was analysed by one-way ANOVA followed by Tukey's multiple comparison test; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.
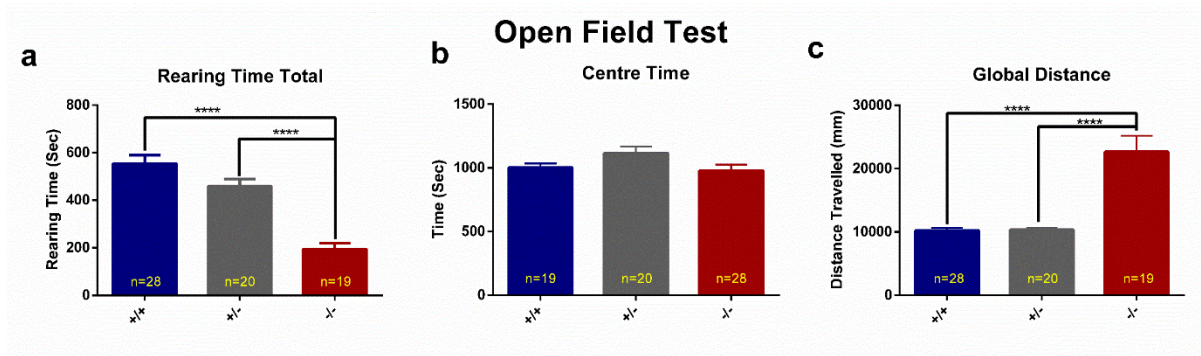
3.1.3 *Adgrl3$^{-/-}$* mice display hyperactivity through increased locomotion in the open field

Mouse genotype had a significant effect on the amount of time mice spent rearing (Fig. 5a, F(2,64)=31.16; p < 0.0001). Both *Adgrl3$^{+/-}$* and *Adgrl3$^{+/+}$* mice on average spent over two fold the amount of time rearing of *Adgrl3$^{-/-}$* (*Adgrl3$^{-/-}$* vs *Adgrl3$^{+/-}$* (p < 0.0001), *Adgrl3$^{-/-}$* vs *Adgrl3$^{+/+}$* (p < 0.0001), Tukey's multiple comparison test). In accordance with the results of the LDB, no significant genotype effect was found in the percentage time mice spent in the central region of the OF (Fig. 5b) another common measure of anxiety (F(2,64)=2.746; p=0.0717). Genotype was found to have a large effect on the distance travelled (Fig. 5c) (F(2,64)=28.06; p < 0.0001). On average, the total distance travelled by *Adgrl3$^{+/-}$* and *Adgrl3$^{+/+}$* mice was less than half the total distance travelled by *Adgrl3$^{-/-}$* mice (*Adgrl3$^{-/-}$* vs *Adgrl3$^{+/-}$* (p < 0.0001), *Adgrl3$^{-/-}$* vs *Adgrl3$^{+/+}$* (p < 0.0001), Tukey's multiple comparison test). Hyperactivity is the likely explanation for both increased locomotion and reduced rearing time of *Adgrl3$^{-/-}$* mice. This hyperactivity was maintained throughout the OF test as showed by 5 min interval analysis of the distance travelled (Fig. 5d). While the distance travelled by *Adgrl3$^{+/+}$* and *Adgrl3$^{+/-}$* mice decreased over the 30 min trial, *Adgrl3$^{-/-}$* locomotion remained constant throughout the OF. The drop in distance travelled by *Adgrl3$^{+/+}$* and *Adgrl3$^{+/-}$* mice is the result of normal habituation which failed to occur in *Adgrl3$^{-/-}$* mice.

3.1.4. Fine-level gait analysis finds reduction in stance time and a wider base of support in *Adgrl3$^{-/-}$* mice

Gait disturbances in *Adgrl3$^{-/-}$* mice was suggested on the basis of qualitative visual observations of home cage activity and movement during the OF test. In order to further investigate these observations in an unbiased quantitative manner we performed a gait analysis of *Adgrl3$^{+/+}$* and *Adgrl3$^{-/-}$* mice using the Gaitlab system. All gait attributes measured are described in the methods section. Stride frequency (Fig. 6a) was not significantly different between genotypes (F (1,96)=2.882; Genotype: p=0.0928) nor was stride length (Fig. 6b) (F (1,96)=0.0268; p=0.8704). *Adgrl3$^{-/-}$* mice did show a significantly reduced stance time in comparison to *Adgrl3$^{+/+}$* mice (Fig. 6c) (F(1,96)=9.949; p=0.0021). When stance time was analysed at the individual paw level, the left hind paw's stance time was found to be significantly shorter for *Adgrl3$^{-/-}$* mice (*Adgrl3$^{+/+}$* vs *Adgrl3$^{-/-}$* LF p=0.4549, RF p=0.2349, LH p=0.0421, RH p=0.9944, Sidak's multiple comparisons test). Swing time was not significantly different between genotypes (Fig. 6d) (F(1,96)=2.857; p=0.0942). Limb gaps (Fig. 6e) were not significantly different between genotypes (F (1,45)=0.2331; Genotype p=0.6316). *Adgrl3$^{-/-}$* mice's base of support (BOS) was significantly wider than *Adgrl3$^{-/-}$* mice's BOS (Fig. 6f, F(1,45)=19.89;

Genotype $p < 0.0001$). *Post-hoc* analysis of fore and hind limb's BOS with appropriate multiple comparison correction showed significant genotype differences for both limb pairs' BOS (*Adgrl3*$^{+/+}$ vs *Adgrl3*$^{-/-}$ Fore Limb p=0.0048, Hind Limb p=0.0069, Sidak's multiple comparison test). The reduced stance time of *Adgrl3*$^{-/-}$ mice is in alignment with the increased hyperactivity observed across trials while the drastically reduced rearing activity seen in the OF may be associated with the abnormal BOS.

Figure 5. Open Field Test: *Adgrl3⁻/⁻* **mice display hyperactivity alongside reduced rearing activity** (a) Time spent rearing was markedly reduced in *Adgrl3⁻/⁻*. (b) Genotype did not have a statistically significant effect on the amount of time mice spent in the central region of the box (c) A significant genotype effect was found on the total distance travelled with *Adgrl3⁻/⁻* mice travelling significantly further. Error Bars represent Standard Error of the Mean. Data were analysed by one-way ANOVA followed by Tukey's multiple comparison test; * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.
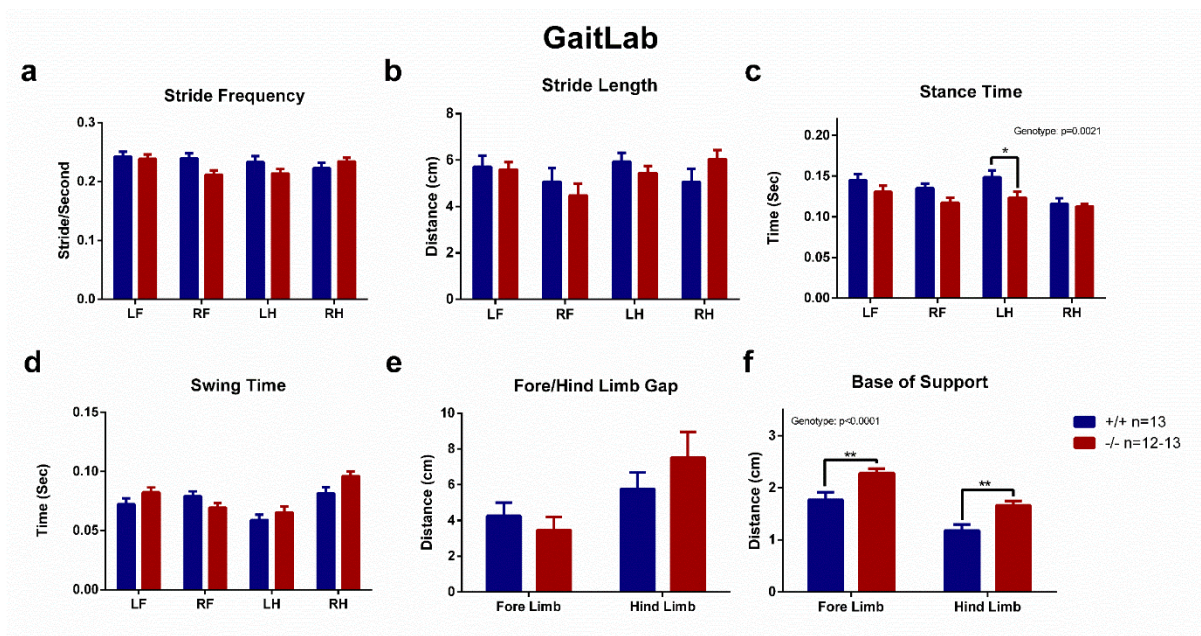
Figure 6. Gaitlab Analysis: *Adgrl3⁻/⁻* **mice showed reduced stance time (indicative of increased hyperactivity) and an increased base of support (suggestive of unstable footing).** Abbreviatons: LF=Left Fore Limb, RF= Right Fore Limb, LH= Left Hind Limb & RH= Right Hind Limb. (a) & (b) No significant genotype effect on stride frequency (a) or length (b). (c) *Adgrl3⁻/⁻* mice stance time was significantly shorter representing a reduced period of time with their paws on the ground. Individual paw-level analysis showed the LH paw stance time of *Adgrl3⁻/⁻* mice was significantly shorter than in *Adgrl3⁺/⁺* mice. (e) No significant genotype effect was found on the swing time (time paws were in the air) (e) The vertical distance (Gap) between fore and hind limb pairs was not significantly different between genotypes. (f) The horizontal cross body distance (Base of Support) was significantly greater in *Adgrl3⁻/⁻* mice than *Adgrl3⁺/⁺* mice. Error Bars represent Standard Error of the Mean. Data was analysed by two-way ANOVA followed by Sidak's multiple comparison test; * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.

3.1.5. The novel object recognition task found memory impairments in *Adgrl3*$^{-/-}$ mice associated with a reduction in capacity to discriminate between novel and familiar objects

Between the object exploration (T1) and objection recognition (T2) trials, no significant difference was found in the total amount of time mice spent interacting with the objects (Fig. 7a) (Trial $F_{(1,65)}=0.0175$; $p=0.8952$). Across both trials, mice did not show a significant genotype effect in the percentage of time spent exploring objects (Genotype $F_{(1,65)}=0.0253$; $p=0.7775$). The time spent interacting with objects varied significantly in a trial by genotype interaction (Trial-by-Genotype Interaction $F_{(2,65)}=5.054$). *Post-hoc* analysis showed *Adgrl3*$^{+/+}$ mice spent significantly more time interacting with the objects during the second trial than during the first trial, while *Adgrl3*$^{+/-}$ and *Adgrl3*$^{-/-}$ mice interaction remained constant across both trials (Exploration vs Recognition Trials: *Adgrl3*$^{+/+}$ $p=0.0162$, *Adgrl3*$^{+/-}$ $p=0.7751$, *Adgrl3*$^{-/-}$ $p=0.5005$, Sidak's multiple comparisons test). During T2, the novel object had a significantly increased amount of interaction in comparison to the familiar object (Fig. 7b) ($F_{(1,130)}=15.70$; Object $p=0.0001$). *Post-hoc* analysis showed this novel object preference was specific to *Adgrl3*$^{+/+}$ and *Adgrl3*$^{+/-}$ mice as *Adgrl3*$^{-/-}$ mice did not spend significantly different periods of time interacting with either the novel or familiar object (Novel vs Familiar Object: *Adgrl3*$^{+/+}$ $p=0.0008$, *Adgrl3*$^{+/-}$ $p=0.0269$, *Adgrl3*$^{-/-}$ $p=0.8175$, Sidak's multiple comparisons test). While *Adgrl3*$^{-/-}$ mice showed similar object exploration levels across trials they did not show the typical preference for the novel object in T2. Unaltered exploration indicates unimpaired motivational levels in *Adgrl3*$^{-/-}$ mice therefore suggesting the reduction in novel object preference is a consequence of memory impairment. During T2, the total distance travelled by mice was significantly less than during T1 (Fig. 7c) (Trial 1) (Trial $F_{(1,65)}=6.678$; $p=0.0120$). *Post-hoc* comparisons showed this reduction was specific to *Adgrl3*$^{+/-}$ mice (Trial *Adgrl3*$^{+/+}$ $p=0.9859$, *Adgrl3*$^{+/-}$ $p=0.0484$, *Adgrl3*$^{-/-}$ $p=0.3490$, Sidak's multiple comparisons test). The global distance travelled across the two trials was found to be significantly different across genotypes (Genotype $F_{(2,65)}=35.89$). In each trial, *Adgrl3*$^{+/+}$ and *Adgrl3*$^{+/-}$ travelled less than 50% of the distance travelled by *Adgrl3*$^{-/-}$ mice (*Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{+/-}$ $p=0.6653$, *Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{-/-}$ $p < 0.0001$, *Adgrl3*$^{+/-}$ vs. *Adgrl3*$^{-/-}$ $p < 0.0001$, Sidak's multiple comparisons test).
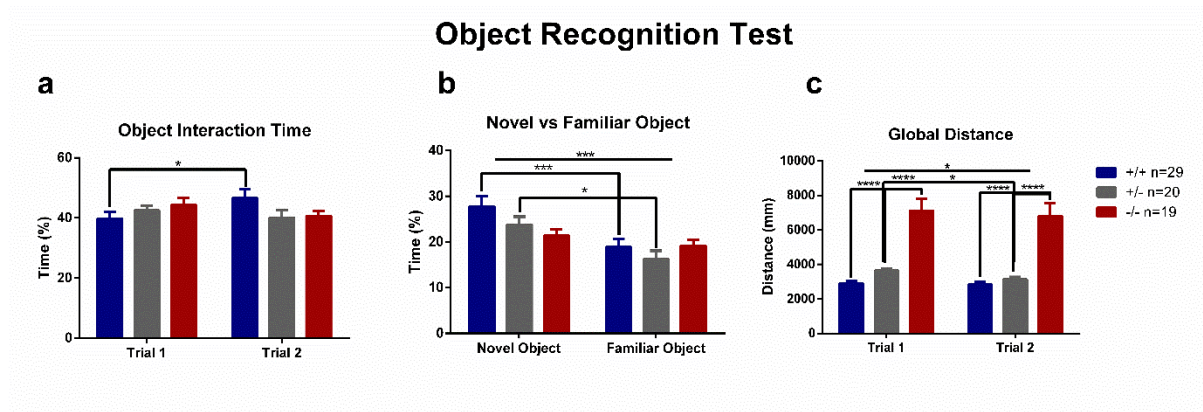
Figure 7. Object Recognition Test: Reduced object recognition suggestive of a memory impairment was observed in *Adgrl3⁻/⁻* mice. (a) There was no significant difference in total interaction time between trials (object recognition (Trial 2) vs object exploration (Trial 1)). A significant trial by genotype interaction was revealed with *Adgrl3⁺/⁺* showing a greater interaction in Trial 2 than Trial 1 which was not seen in *Adgrl3⁻/⁻* mice. (b) In T2, a significant overall preference for the novel object was detected. This significant preference was confined to *Adgrl3⁺/⁺* and *Adgrl3⁺/⁻* mice while *Adgrl3⁻/⁻* mice did not show any preference for the novel object. (c) Inter-trial comparisons showed a significantly reduced distance travelled during T2. Multiple comparison analysis revealed this reduction in distance was significant only for *Adgrl3⁺/⁻* mice. Across both trials, a significant genotype effect was detected with *Adgrl3⁻/⁻* mice travelling a greater distance than *Adgrl3⁺/⁻* or *Adgrl3⁻/⁻* mice. Error Bars represent Standard Error of the Mean. Data were analysed by two-way ANOVA followed by Sidak's multiple comparison test; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.1.6. Visuospatial memory as assessed in the Barnes Maze shows an impairment in *Adgrl3*$^{-/-}$ mice

All mice showed the capability of learning where the escape box was located both during the primary acquisition phase (Fig. 8a–c) and secondary reversal phase (Fig. 8d–f), as evidenced by the reduction in escape latency, primary error count and global distance travelled as the trials progressed (Latency, Primary Error and Global Distance: two-way repeated measures ANOVA: Session in Acquisition Phase $p < 0.0001$ and Session in Reversal Phase $p < 0.0001$). The latency to escape and primary error count did not show evidence of a significant trial by genotype interaction during either acquisition or reversal phases. Global distance travelled did show a significant trial by genotype interaction in the acquisition phase as a result of *Adgrl3*$^{-/-}$ mice travelling a greater distance during T1-T4 (Session by Genotype Interaction $F(18,585)=2.721$; $p=0.0002$). A significant genotype effect was found in the escape latency during the acquisition phase (Fig. 8a) with *post-hoc* analysis showing *Adgrl3*$^{-/-}$ mice had a significantly increased latency over heterozygous or wild-type mice ($F(2,65)=5.769$; Genotype $p=0.0049$, *Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{-/-}$ $p=0.0054$, *Adgrl3*$^{+/-}$ vs. *Adgrl3*$^{+/+}$ $p=0.0260$, Tukey's multiple comparisons test). A genotype effect was also found in the number of primary errors (Fig. 8b) with *post-hoc* analysis showing *Adgrl3*$^{-/-}$ mice also had a significantly increased number of primary errors in comparison to the other two genotype, while *Adgrl3*$^{+/-}$ mice made significantly more errors than *Adgrl3*$^{+/+}$ mice ($F(2,65)=5.501$; Genotype $p < 0.0001$, *Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{-/-}$ $p=0.0092$, *Adgrl3*$^{+/-}$ vs. *Adgrl3*$^{+/+}$ $p=0.0469$, Tukey's multiple comparisons test). Finally during the acquisition phase, a genotype effect was found in global distance travelled, with appropriate *post-hoc* analysis showing *Adgrl3*$^{-/-}$ mice travelled a greater distance (Fig. 8c) than *Adgrl3*$^{+/-}$ or *Adgrl3*$^{+/+}$ mice while *Adgrl3*$^{+/+}$ mice travelled a greater distance than *Adgrl3*$^{+/-}$ mice (Genotype $F(2,65)=16.46$; $p < 0.0001$, *Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{-/-}$ $p=0.0017$, *Adgrl3*$^{+/-}$ vs. *Adgrl3*$^{-/-}$ $p < 0.0001$, *Adgrl3*$^{+/+}$ vs. *Adgrl3*$^{+/-}$ $p=0.0028$, Tukey's multiple comparisons test). During the reversal phase a similar pattern of results was found between *Adgrl3*$^{+/+}$ and *Adgrl3*$^{-/-}$ mice, with significant genotype differences in escape latency and global distance (Fig. 8d and f) (Escape Latency: $F(1,30)=4.347$, Distance Travelled $F(1,30)=5.59$; $p=0.0247$). The number of primary errors was not significantly different between *Adgrl3*$^{+/+}$ and *Adgrl3*$^{-/-}$ mice (Fig. 8e) (Primary Errors $F(1,30)=0.8047$; $p=0.3768$). The reduced ability of *Adgrl3*$^{-/-}$ mice to learn the position of the escape box is indicative of an impairment in visuospatial memory processing and consolidation which may be the result of a hippocampal-based memory deficiency.
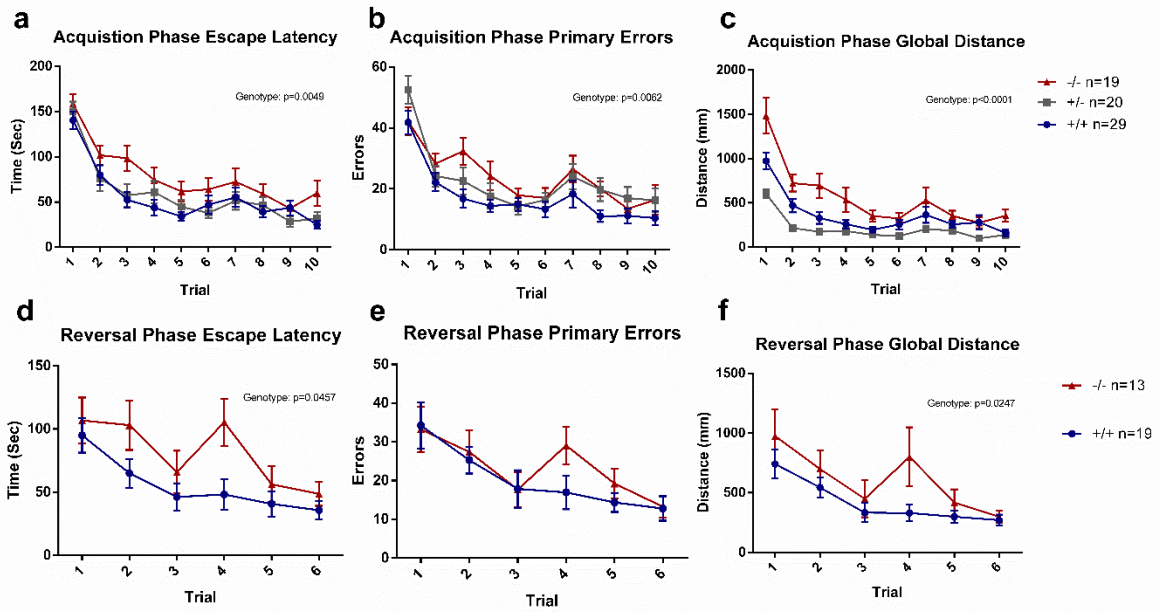
Figure 8. Barnes Maze: Increased escape latency and primary error count of *Adgrl3⁻/⁻* mice is indicative of a spatial memory deficit. Acquisition Phase: (a) Escape latency was significantly increased in *Adgrl3⁻/⁻* mice compared to heterozygous or wild-type mice. (b) *Adgrl3⁻/⁻* mice showed a significantly greater primary error count in comparison to wild-type but not heterozygous mice. *Adgrl3⁺/⁻* mice made significantly more errors than *Adgrl3⁺/⁺* mice. (c) A significant genotype effect was observed on distance travelled with *Adgrl3⁻/⁻* mice traveling further than *Adgrl3⁺/⁺* or *Adgrl3⁺/⁻* mice while *Adgrl3⁺/⁺* mice travelled a greater distance than *Adgrl3⁺/⁻* mice. **Reversal Phase:** (d) Escape latency was significantly greater in *Adgrl3⁻/⁻* mice compared to *Adgrl3⁺/⁺* mice (e) Primary error count did not show a significant genotype effect. (f) *Adgrl3⁻/⁻* travelled a significantly greater distance than *Adgrl3⁺/⁺* mice. Error Bars represent Standard Error of the Mean. Data was analysed by two-way ANOVA followed by Sidak's multiple comparison test; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.1.7. The social interaction test revealed *Adgrl3⁻ᐟ⁻* mice have higher levels of sociability but deficits in social memory

During the social novelty trial (T2), mice showed an overall increase in interaction time with the familiar or novel mouse in comparison to interaction time during the sociability trial (T1), this increase in interaction was present in both *Adgrl3⁺ᐟ⁺* and *Adgrl3⁻ᐟ⁻* mice (T1) (Fig. 9a) (Trial $F(1,22)=26.03$; $p < 0.0001$, *Adgrl3⁺ᐟ⁺* $p=0.0003$, *Adgrl3⁻ᐟ⁻* $p=0.0275$, Sidak's multiple comparisons test). No significant genotype effect was found in the comparison of total interaction time across T1 and T2 (Genotype $F(1,22)=0.0870$; $p=0.7709$). During T1, mice spent a greater amount of time interacting with the cage containing the wild-type mouse than time spent interacting with the empty cage, this increased interaction signifies a significant preference for sociability (Fig. 9b) (Cage $F(1,44)=4.366$; $p=0.0425$). A significant genotype by cage interaction effect was detected with *post-hoc* analysis showing the preference for sociability was only present in *Adgrl3⁻ᐟ⁻* mice, which spent significantly more time interacting with the mouse cage, while *Adgrl3⁺ᐟ⁺* mice did not show a significant preference for sociability (Genotype by Cage Interaction $(F(1,44)=6.972$; $p=0.0114$, *Adgrl3⁺ᐟ⁺* $p=0.9092$, *Adgrl3⁻ᐟ⁻* $p=0.0034$, Sidak's multiple comparisons test). In T2, mice spent a significantly greater amount of time interacting with the novel mouse than the familiar mouse, indicating a preference for social novelty (Fig. 9c) (Cage $F(1,44)=4.869$; $p=0.0326$). The preference for social novelty was specific to *Adgrl3⁺ᐟ⁺* mice with no significant increase in novel mouse interaction over familiar mouse interaction seen in *Adgrl3⁻ᐟ⁻* (*Adgrl3⁺ᐟ⁺* $p=0.0278$, *Adgrl3⁻ᐟ⁻* $p=0.8213$, Sidak's multiple comparisons test). The results from both trials show an increased sociability of *Adgrl3⁻ᐟ⁻* mice alongside a reduction in preference for social novelty. In combination, these results are highly indicative of an impairment in social memory consolidation (Kogan et al., 2000). Inter-trial analysis showed a reduction in the total distance travelled during T2 in comparison to T1 (Fig. 9d) (Trial $F(1,22)=8.489$; $p=0.0080$). *Post-hoc* analysis showed this decrease in locomotion was restricted to *Adgrl3⁻ᐟ⁻* mice with no significant difference in *Adgrl3⁺ᐟ⁺* mice (*Adgrl3⁺ᐟ⁺* $p=0.3424$, *Adgrl3⁻ᐟ⁻* $p=0.0224$, Sidak's multiple comparisons test).
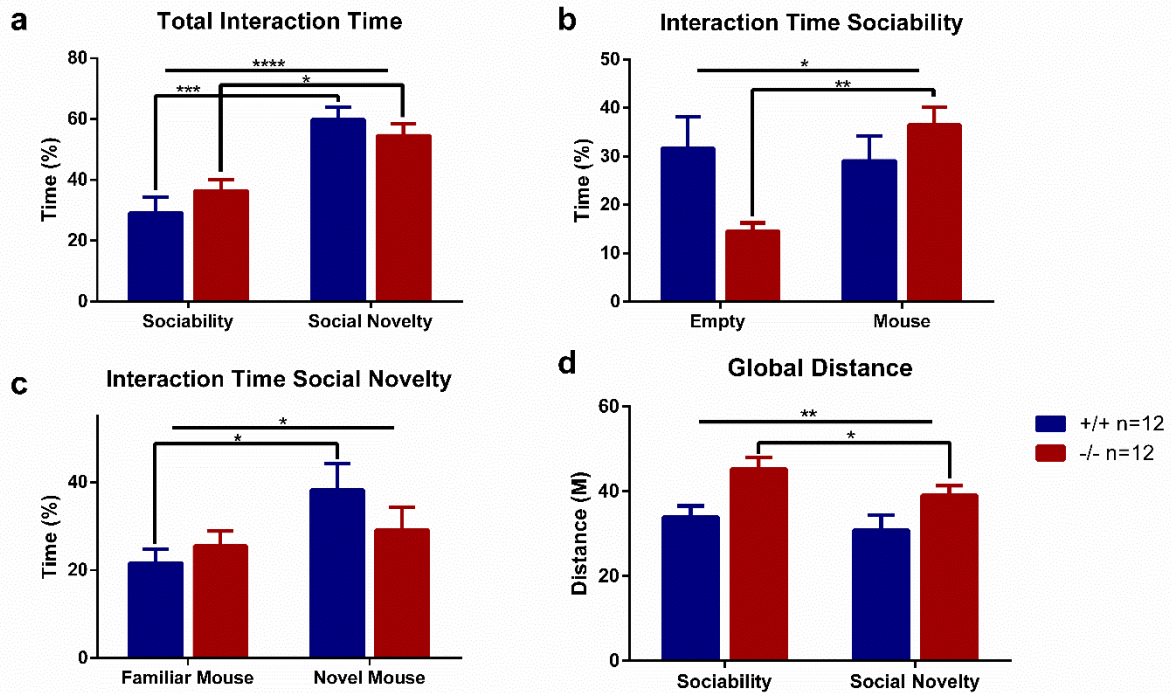
Figure 9. Social Interaction Test: Increased sociability and reduced preference for social novelty in *Adgrl3⁻/⁻* mice is indicative of a social memory impairment. (a) Total interaction time was greater in the social novelty trial (T2) than the sociability trial (T1) with both genotypes showing increased interaction. (b) During T1 a significant genotype by cage interaction effect was observed. *Adgrl3⁻/⁻* mice showed increased time spent interacting with the mouse cage relative to the empty cage. (c) During T2, *Adgrl3⁺/⁺* mice showed an overall preference for the novel mouse which was not seen in *Adgrl3⁻/⁻*. (d) Distance travelled significantly decreased in T2 from T1. This reduction was confined to *Adgrl3⁻/⁻* mice, which exhibited significantly increased locomotion overall. Error Bars represent Standard Error of the Mean. Data were analysed by two-way repeated measures ANOVA followed by Sidak's multiple comparison test; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.1.8. The resident-intruder paradigm shows a dramatic decrease in the aggression of *Adgrl3⁻/⁻* mice

Prior to performing the resident-intruder test, it was planned to record the total number of resident attacks, the latency to primary attack and the presence or absence of resident attack for each mouse. Due to the complete absence of aggression from $Adgrl3^{-/-}$ mice it was only possible to record and analyse the presence/absence of attack parameter. This parameter revealed a strongly significant genotype effect (Fig. 10) (Fisher's Exact Test p=0.0007) with 80% of $Adgrl3^{+/+}$ resident mice attacking the intruder mouse, in comparison to no attacks by $Adgrl3^{-/-}$ mice. In these mice, aggression is therefore strongly impacted by the inactivation of $Adgrl3$.
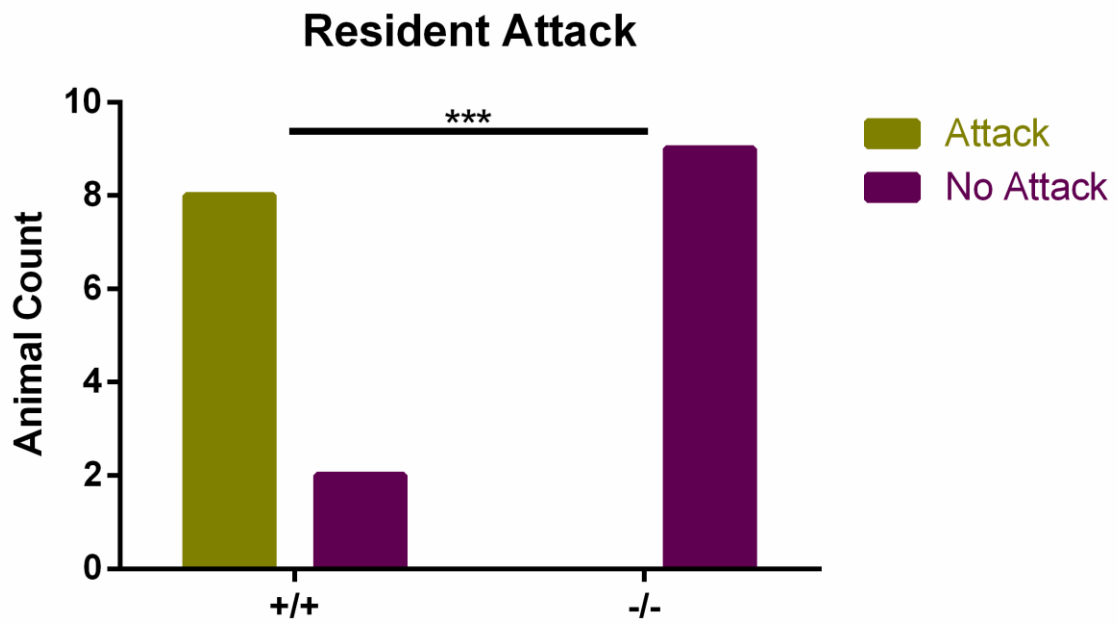
Figure 10. Resident Intruder Test: Reduced aggression of *Adgrl3*^{-/-} mice. There was a complete absence of aggression in *Adgrl3*^{-/-} mice. *Adgrl3*^{-/-} mice had a significantly lower rate of attack. Data was analysed by Fisher's Exact Test ; * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.

3.1.9. The continuous performance test (CPT) reveals increased levels of impulsivity and reduced motivation for food in *Adgrl3⁻/⁻* mice

Analysis of the training stages did not show any evidence of general cognitive deficits which may have been identified by a genotype effect on the number of sessions required to reach the advancement criterion (Fig. 11a). The overall performance of mice increased as Stage 3 training progressed (Fig. 11b), this improvement was captured by a significant increase in both HR and sensitivity (d') (Session: $F_{(6,84)}=11.92$; HR, $p < 0.0001$; d' $F_{(6,84)}=27.50$; $p < 0.0001$) coupled with a reduction in FAR (Session: $F_{(6,84)}=9.94$; $p < 0.0001$). Response bias (c) levels remained constant across stage 3 training sessions. A significant genotype effect was found in stage 3 with FAR being significantly higher for *Adgrl3⁻/⁻* mice than *Adgrl3⁺/⁺* mice (Genotype: $F_{(1,84)}=5.92$; $p=0.0289$). In contrast, no significant genotype effect was found on HR, sensitivity and response bias during stage 3.

Baseline performance was defined as the mean of the last two CPT training sessions (Fig. 11c). For baseline performance; HR, FAR, d' and c were not found to be significantly different between genotypes (Fig. 12). Meanwhile, on average *Adgrl3⁻/⁻* mice performed an over two fold greater number of screen touches during ITIs than *Adgrl3⁺/⁺* mice ($t_{(16)}=3.451$; $p=0.0033$), a finding indicative of higher impulsivity. While correct and incorrect response latencies were not significantly different between *Adgrl3⁺/⁺* and *Adgrl3⁻/⁻* mice, *Adgrl3⁻/⁻* mice showed significantly longer reward retrieval latency in comparison to *Adgrl3⁻/⁻* mice ($t_{(14)}=2.947$; $p=0.0106$). This reduction in reward retrieval latency may be the result of reduced motivation for food in *Adgrl3⁻/⁻* mice.

Mice which reached the advancement criterion during stage 3 training proceeded to probe sessions during which the impact of baseline parameter modifications was examined (Fig. 13). During the probe 1 (variable stimulus duration) sessions, HR and d' decreased as stimulus durations were reduced (Stimulus Duration: HR, $F_{(5,84)}=6.28$; $p < 0.0001$ and d', $F_{(5,84)}=16.88$; $p < 0.0001$). These decreases did not show a significant genotype effect. A significant genotype effect was found for FAR which increased in both genotypes as stimulus durations reduced (Stimulus Duration: $F_{(5,84)}=4.27$; $p=0.0017$) but a greater increase was observed in *Adgrl3⁻/⁻* mice (Genotype: $F_{(1,84)}=5.42$; $p=0.0223$). During probe 2 (variable stimulus contrast) as stimulus contrast decreased, HR remained constant, while d' and c decreased (d': $F_{(3,56)}=27.89$; $p < 0.0001$ and c:$F_{(3,56)}=5.79$; $p=0.0016$, respectively) and FAR increased (Stimulus Contrast: $F_{(3,56)}=24.62$; $p < 0.0001$). A significant genotype effect was found for HR, FAR and c; with *Adgrl3⁻/⁻* mice showing higher levels for HR and FAR

alongside a reduced c (HR: Genotype: $F_{(1,56)}=9.35$; p=0.0034, FAR: Genotype: $F_{(1,56)}=10.81$; p=0.0017, c, Genotype: $F_{(1,56)}=13.26$; p < 0.001). Probe 3 (long ITI), did not show significant effects of either genotype or parameter variation on task performance (Fig. 13). The FAR reduced during probe 4 (S+ probability) after the S+ probability was reduced to 30% (S+ probability: $F_{(1,30)}=4.40$; p=0.044). During probe 4, $Adgrl3^{-/-}$ mice had a higher overall FAR than $Adgrl3^{+/+}$ mice (Genotype: $F_{(1,30)}=4.71$; p=0.0380). Higher FAR of $Adgrl3^{-/-}$ mice FAR was observed across the majority of sessions and is highly indicative of increased impulsivity in $Adgrl3^{-/-}$ mice.

Figure 11. Continuous performance test (CPT) – Training Sessions: Increased false alarm rate in *Adgrl3-/-* mice during acquisition stage and higher inter-trial interval touches and reward retrieval latency at baseline performance (a) The required number of sessions to reach progress criterion was not significantly different between genotypes (b) Significant genotype effect during acquisition stage with *Adgrl3-/-* mice showing a higher false alarm rate than *Adgrl3+/+* mice. (c) Baseline performance parameters show significantly higher inter-trial interval touches (ITI) and reward retrieval latency in *Adgrl3-/-* mice. All mice were trained for 7 sessions. Error Bars represent Standard Error of the Mean. Data was analysed by unpaired t-test (b) or two-way ANOVA and t-test (a and c); * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.
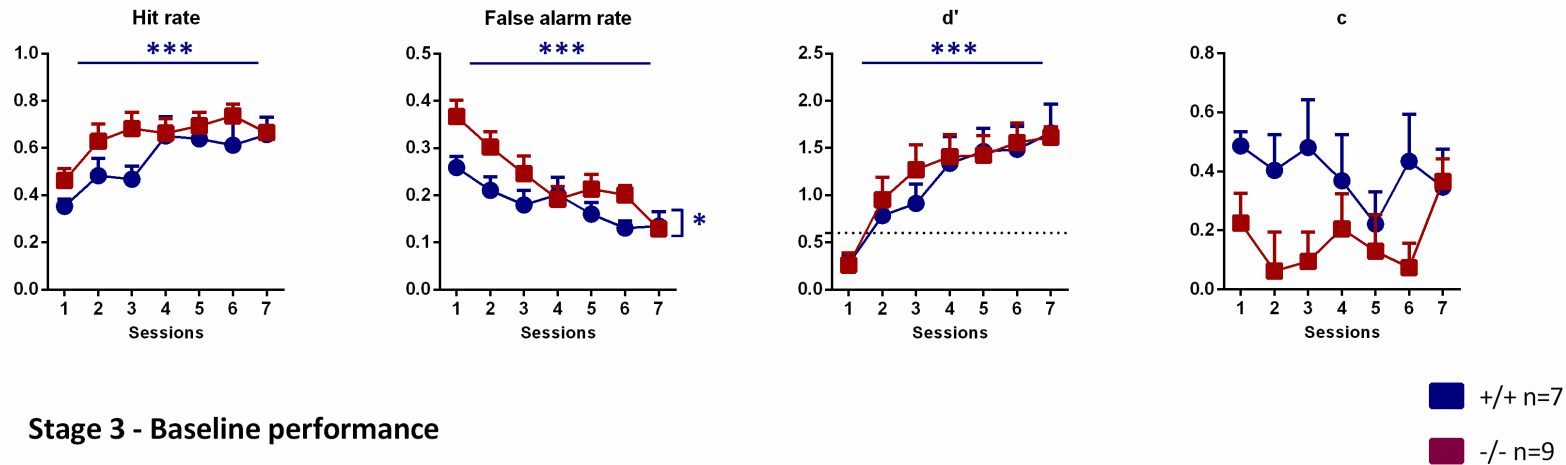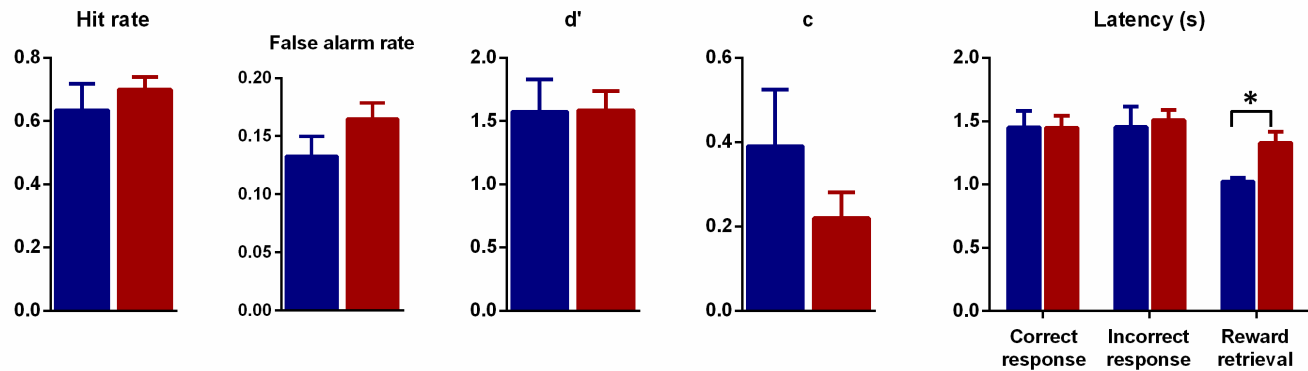
Figure 12. CPT probe trials 1 & 2: Higher false alarm rate in *Adgrl3*$^{-/-}$ mice during Probe 1 and Probe 2 sessions with increased hit rate during Probe 2 and lower c.
False alarm rate was higher for *Adgrl3*$^{-/-}$ mice in Probe 1 (variable stimulus duration) and Probe 2 (variable stimulus contrast) in comparison to *Adgrl3*$^{+/+}$ mice. Significant genotype effect during Probe 2 with *Adgrl3*$^{-/-}$ mice displaying an increased hit rate and a decreased response bias (c) in comparison to *Adgrl3* $^{+/+}$ mice. Error Bars represent Standard Error of the Mean. Data were analysed by two-way ANOVA; * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.
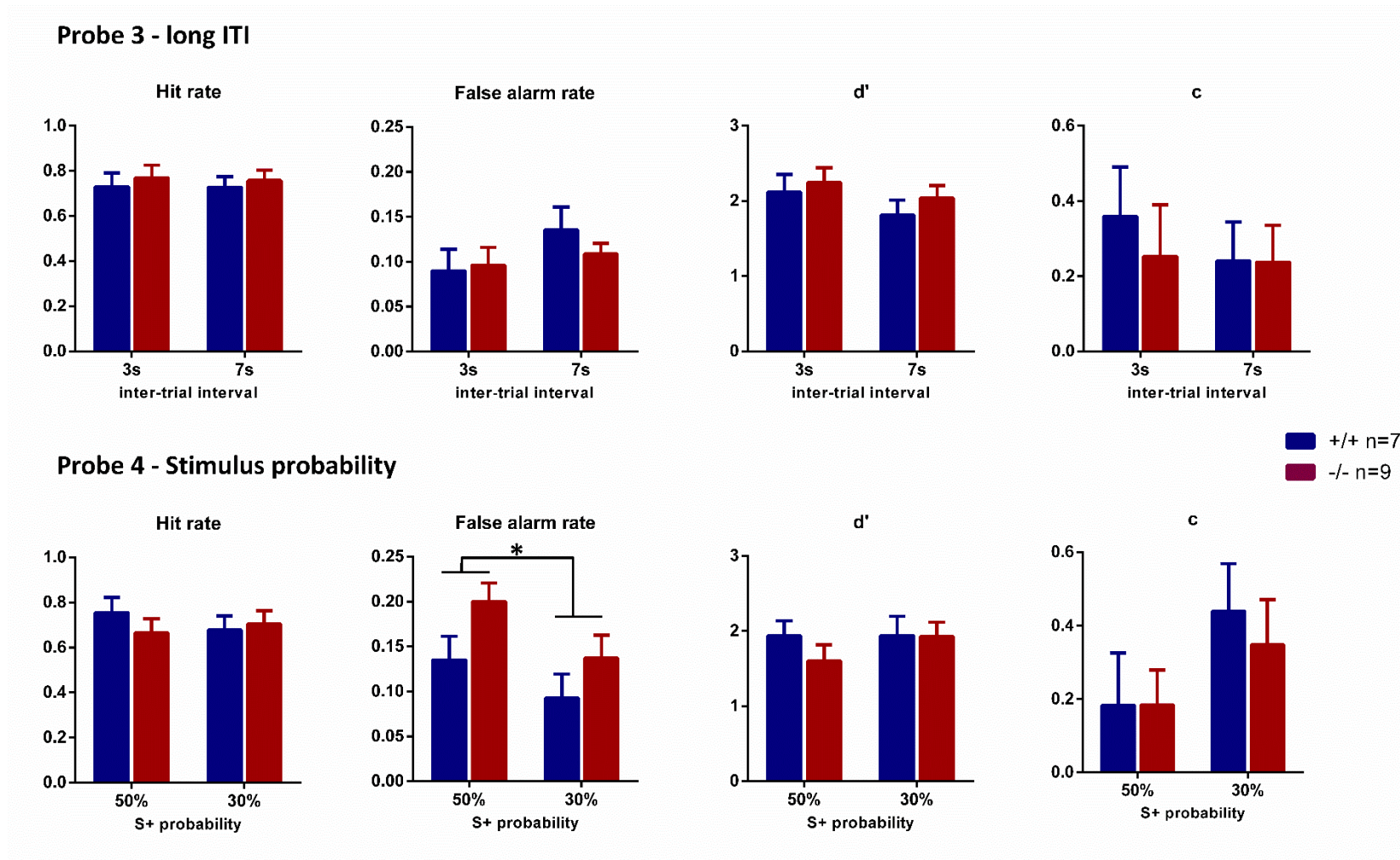
Figure 13. CPT probe trials 3 & 4. *Adgrl3^{-/-}* **mice showed an increased false alarm rate in Probe 4. (S+ probability)** No significant genotype effects observed in Probe 3 (variable ITI). Significantly higher false alarm rate in *Adgrl3^{-/-}* mice during Probe 4 (S+ probability) in comparison to *Adgrl3^{+/+}* mice. Error Bars represent Standard Error of the Mean. Data were analysed by two-way ANOVA; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

3.1.10. Transcriptomic analysis of brain regions reveals differential gene expression

Gross level differences in gene expression patterns were detected by principal component analysis of all sequenced samples which clearly separated into three distinct clusters which corresponded to the PFC, hippocampus and striatum (Figs. 14–15). Each brain region was then analysed independently to detect statistically significant differential gene expression between $Adgrl3^{+/+}$ and $Adgrl3^{-/-}$ mice. The region with the greatest number of statistically significant differentially expressed genes (DEG) was the PFC with 180 genes. 115 (63.9%) of these DEG were upregulated in the $Adgrl3^{-/-}$ PFC in comparison to $Adgrl3^{+/+}$ samples. 24 (13.33%) of these DEG had a greater than two-fold difference in gene expression levels (Table 1). While the majority of PFC DEG were upregulated only two of these genes had an average log fold changes (logFC) greater than 1. These genes were *interleukin 31* (*Il31*) and *starch binding domain 1* (*Stbd1*) (logFC *Il31*=1.86, logFC *Stbd1*=1.97). While fewer genes were significantly downregulated in the PFC, 22 of the 65 DEG had average expression level in $Adgrl3^{-/-}$ mice of less than 50% mean expression levels in $Adgrl3^{+/+}$ mice. A 50% reduction is represented by than a logFC < -1. The gene which codes for the dopamine transporter (DAT), *solute carrier family 6, member 3 (Slc6a3)*, had a logFC=−3.03 which made it the most downregulated gene in any of the RNA-sequenced brain regions.

Across $Adgrl3^{+/+}$ and $Adgrl3^{-/-}$ mice, 36 DEG were detected in the hippocampus. 23 (63.9%) of these genes were upregulated in the $Adgrl3^{-/-}$ hippocampus relative to $Adgrl3^{+/+}$ mice expression levels. Two of these genes had a mean logFC level of greater than 1 in magnitude. *Stbd1*, which was also significantly upregulated in the PFC, had a logFC=1.82 in the $Adgrl3^{-/-}$ hippocampus compared to $Adgrl3^{+/+}$ mice. *Perilipin 4* (*Plin4*) was the only other gene with a logFC=|1| with a logFC=−1.11 in the $Adgrl3^{-/-}$ hippocampus in comparison to $Adgrl3^{+/+}$ mice expression levels (Table 1). The smallest number of DEG (22) were found in the striatum. The number of DEG was divided equally between upregulated and downregulated genes with 11 each. Six of these genes had a logFC>|1| (Table 1). Five of these genes had increased expression levels in the $Adgrl3^{-/-}$ striatum compared to the $Adgrl3^{+/+}$ striatum. *Pro-melanin-concentrating hormone (Pmch)* was the most upregulated gene in the striatum analysis and also had the largest magnitude logFC detected in any of the sequenced brain regions (logFC=5.370832). *Pmch* codes for a complex gene product with is proteolytically cleaved into a minimum of three known protein products: Melanin-Concentrating Hormone (MCH), Neuropeptide-Glutamic

Acid-Isoleucine (NEI), and Neuropeptide-Glycine-Glutamic acid (NGE). *Shroom family member 3 (Shroom3)*, was the only gene with a logFC < 1 in the *Adgrl3⁻/⁻* striatum (logFC=-1.096714321). A subsample of three *Adgrl3⁻/⁻* mice (n=3) had much greater levels of expression neuropeptide and neuro-hormone coding genes than those seen in the other seven *Adgrl3⁻/⁻* mice. The positions of these three samples within the principal component analysis shows that these mice had general striatum gene expression profiles similar to other *Adgrl3⁺/⁺* and *Adgrl3⁻/⁻* mice (Figs. 14–15).

Taken in combination, the results from the three brain regions reveal 208 DEG across *Adgrl3⁺/⁺* and *Adgrl3⁻/⁻* mouse samples, with the majority of these genes upregulated (134 genes (64.4%)). Most DEG (188 genes (90.4%)) were distinct to one brain region with the majority of these region-specific genes being upregulated (66.0%). The remaining twenty genes were divided equally between genes differentially expressed in two brains regions and in all three brain regions (10 each). Upregulation and downregulation was evenly divided within both groups of genes (5 each). The direction of logFC was uniform across brain regions for all DEG implicated in multiple regions.

Differential Gene Expression Absolute log$_2$ Fold Change ≥ |1|			
Gene Symbol	**Gene Name**	**log$_2$FoldChange**	**P-value (adj)**
PFC			
Slc6a3	*solute carrier family 6 (neurotransmitter transporter, dopamine), member 3*	-3.030248464	0.000126768
Fermt1	*fermitin family member 1*	-2.85447429	0.027745238
Barhl2	*BarH-like 2 (Drosophila)*	-2.341735495	0.012046358
Scgn	*secretagogin, EF-hand calcium binding protein*	-2.228703229	0.000343187
Chrna10	*cholinergic receptor, nicotinic, alpha polypeptide 10*	-2.170113334	0.045224968
Chrnb4	*cholinergic receptor, nicotinic, beta polypeptide 4*	-1.975645914	0.028005592
Trh	*thyrotropin releasing hormone*	-1.895089286	0.000348037
Doc2g	*double C2, gamma*	-1.868514734	0.020227139
Tbx21	*T-box 21*	-1.736850578	0.026824467
Cdhr1	*cadherin-related family member 1*	-1.685605756	0.036880589
Frmd7	*FERM domain containing 7*	-1.655572263	0.008079282
Th	*tyrosine hydroxylase*	-1.578888447	0.002841042
Aqp1	*aquaporin 1*	-1.557704669	0.012046358
Fgf16	*fibroblast growth factor 16*	-1.428303169	0.028005592
Adamts19	*a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 19*	-1.303170033	0.030038299
Prss56	*protease, serine 56*	-1.302985856	0.026824467
Hspb8	*heat shock protein 8*	-1.188774118	0.002246439
Slc9a4	*solute carrier family 9 (sodium/hydrogen exchanger), member 4*	-1.125190849	0.00157127
Ppef2	*protein phosphatase, EF hand calcium-binding domain 2*	-1.124102836	0.0318905
Vipr2	*vasoactive intestinal peptide receptor 2*	-1.097318756	0.01722175
Nmb	*neuromedin B*	-1.066365486	0.01869699
Il31	*interleukin 31*	1.85630251	0.012046358
Stbd1	*starch binding domain 1*	1.969471871	1.41998E-05
Hippocampus			
Plin4	*perilipin 4*	-1.112291168	0.016872478
Stbd1	*starch binding domain 1*	1.818306565	5.34443E-07
Striatum			
Shroom3	*shroom family member 3*	-1.096714321	0.000224201
Stbd1	*starch binding domain 1*	1.800834865	8.41806E-14
Avp	*arginine vasopressin*	3.200658475	0.000248666
Oxt	*Oxytocin*	4.034462022	0.000248666
Hcrt	*Hypocretin*	5.117557719	0.000248666
Pmch	*pro-melanin-concentrating hormone*	5.370831844	9.32844E-06

Table 1. All Differential Expressed Genes with a log$_2$ Fold Change ≥ |1|. List of all genes in the pre-frontal cortex, hippocampus or striatum which passed the multiple comparison adjusted p-value threshold of 0.05 and had a greater than 50% increase or decrease in expression found in *Adgrl3$^{-/-}$* mice relative to *Adgrl3$^{+/+}$* mice for the respective brain region.
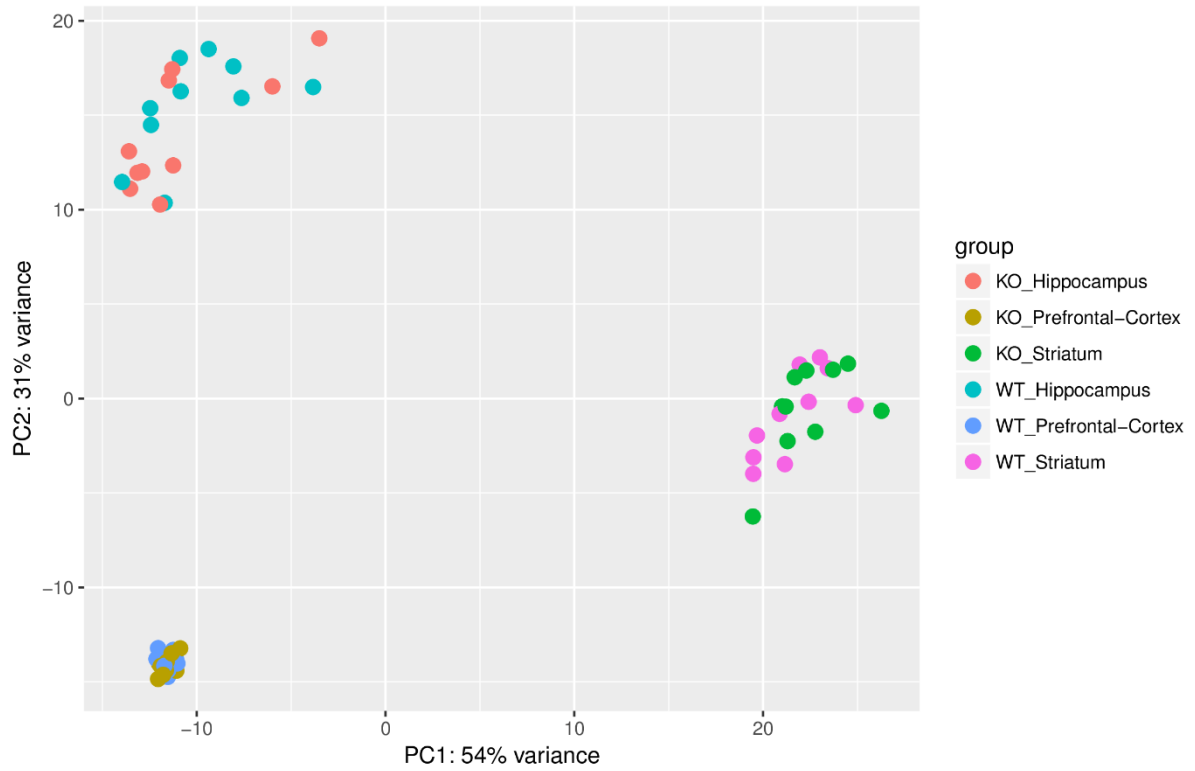
Figure 14. Principal Component Analysis of RNA-sequencing results reveals three distinct clusters. The samples corresponding to the hippocampus, PFC and striatum form three independent clusters composed of both *Adgrl3*$^{+/+}$ and *Adgrl3*$^{-/-}$ mouse samples.
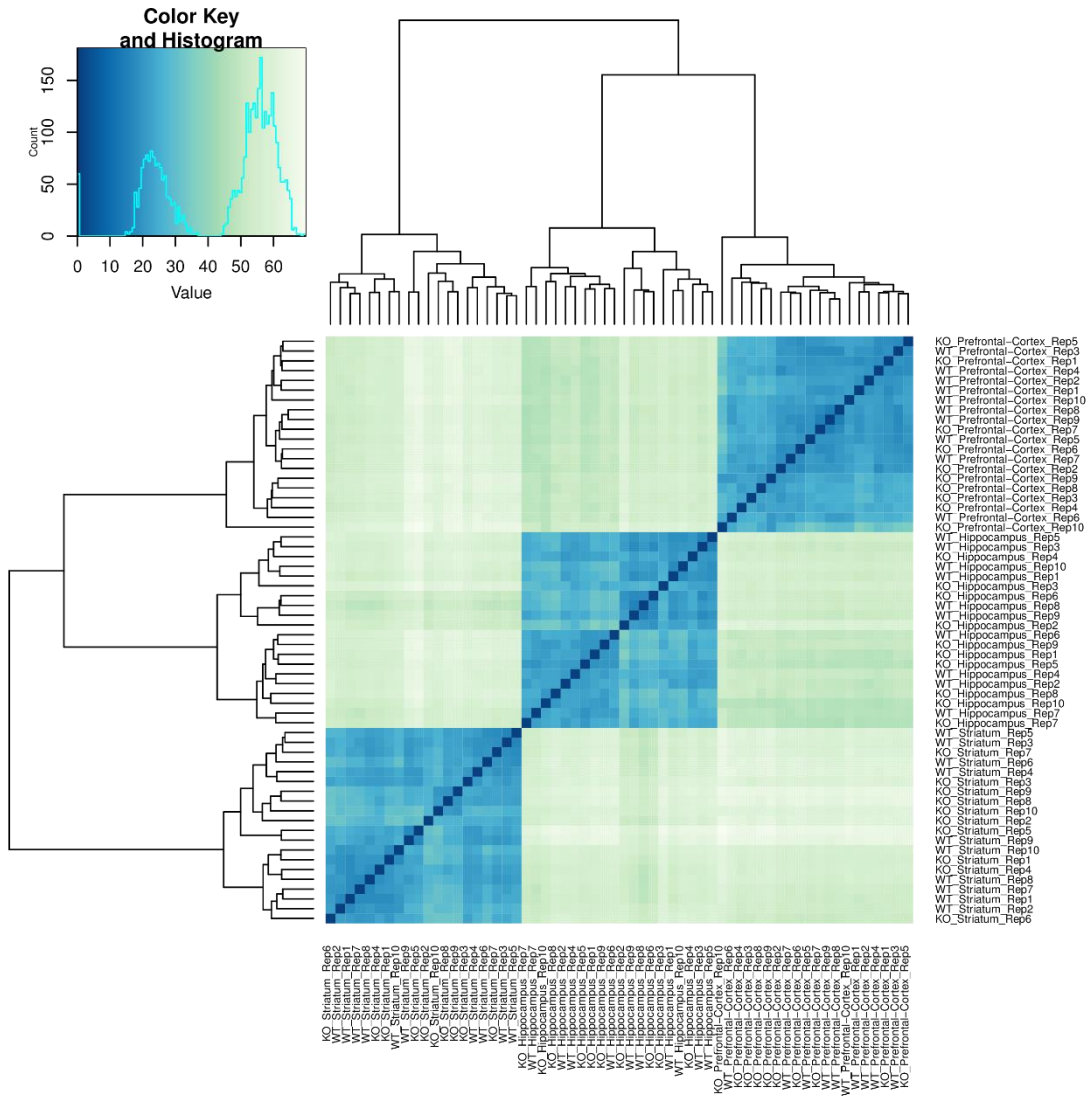
Figure 15. Heatmap of RNA-Sequencing Results reveals three distinct sample clusters. Heat-map clusters represented by blue squares correspond to the hippocampus, pre-frontal cortex and striatum.

3.1.11. Gene set enrichment and pathway analysis reveals statistically significant enrichment of multiple biologically relevant gene sets

Only one of the gene sets met the pre-determined requirements for GSEA using Enricher, this gene set was the PFC downregulated genes (Chen et al., 2013; Kuleshov et al., 2016). Following appropriate levels of multiple testing correction, 5 statistically significant KEGG pathways were identified. These pathways were; cocaine addiction, amphetamine addiction, neuroactive ligand receptor interaction, dopaminergic synapse and Parkinson's disease (Table 2).

All three brain regions were analysed using the IPA analysis software. The most significantly enriched biological and disease associated annotated terms were for the PFC, with behaviour (Fig. 16) and neurological diseases being the top two terms (Benjamini-Hochberg testing correction: Behaviour p=0.000135, Neurological Diseases p=0.000305).

PFC - Down Regulated - Gene Set Enrichment Analysis		
Term	**Adjusted P-value**	**Genes**
Cocaine addiction	0.022277215	*TH;SLC6A3*
Amphetamine addiction	0.022277215	*TH;SLC6A3*
Neuroactive ligand-receptor interaction	0.022277215	*CHRNB4;VIPR2;CHRNA10*
Dopaminergic synapse	0.044262629	*TH;SLC6A3*
Parkinson's disease	0.044262629	*TH;SLC6A3*

Table 2. Enriched gene sets in the PFC detected using Enricher. Genes with a logFC <-1 in the *Adgrl3* -/- mouse PFC were significantly enriched for the five KEGG 2016 Pathways listed. All five pathways are highly relevant to neurobiology and potentially linked to the neuropathology of ADHD.
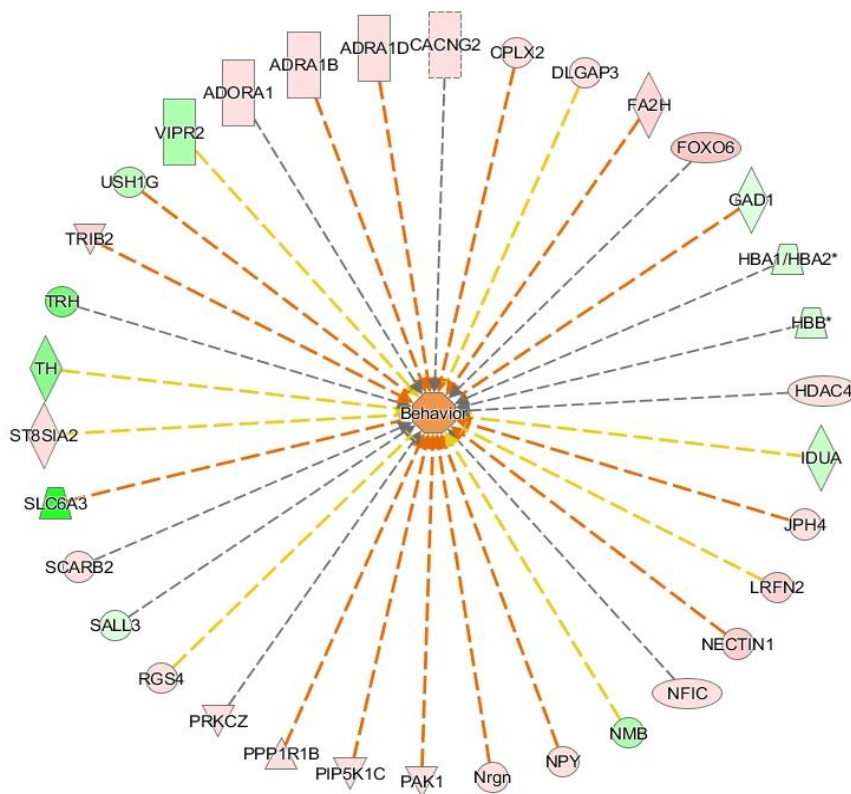
Figure 16. Behaviour was the top IPA annotated term for differential expressed genes in the prefrontal cortex. Analysis of differential expressed genes from the prefrontal cortex of *Adgrl3* -/- mice revealed behaviour as the top annotated term. Legend: Green genes are down regulated in *Adgrl3* -/- mice and red genes are upregulated. Colour Intensity=Magnitude of fold change

3.2 Human ADHD Transcriptomic and Genomic Results

3.2.1 Differential gene expression between ADHD patients and controls in PMBCs

Following standard quality control and removal of duplicate microarray probes matching to the same gene, expression data was available for 18227 genes. 1793 of these genes showed a nominal level of differential expression between the 94 medication-naive ADHD patients and 124 control PMBCs. Following adjustment of the p-value threshold for multiple comparisons using the Benjamini-Hochberg method, expression differences of 21 of these genes remained statistically significant (Table 1). The majority (71.4%) of these genes were found to have reduced expression levels in ADHD patients relative to controls.

ADHD vs Control Patient Microarray – Differentially Expressed Genes

Gene Symbol	Gene Name	logFC	P.Value	adj.P.Val
KMT5A	lysine methyltransferase 5A	-0.119591645	7.65E-08	0.00146811
NXF1	nuclear RNA export factor 1	-0.106359194	4.24E-07	0.00406769
KLF4	Kruppel like factor 4	-0.376415819	7.33E-07	0.00468753
LRRFIP1	LRR binding FLII interacting protein 1	-0.197091082	3.28E-06	0.01573658
PPP1R9B	protein phosphatase 1 regulatory subunit 9B	-0.112164431	5.03E-06	0.01610229
RAB11FIP1	RAB11 family interacting protein 1	-0.197793054	6.03E-06	0.01653422
ABCG1	ATP binding cassette subfamily G member 1	-0.186203941	1.25E-05	0.02680522
ETFDH	electron transfer flavoprotein dehydrogenase	0.132578513	1.63E-05	0.03124746
C1QA	complement C1q A chain	-0.176501627	2.24E-05	0.03900868
ZC3H3	zinc finger CCCH-type containing 3	-0.113782	2.95E-05	0.0471418
SLA	Src like adaptor	0.130275595	3.95E-05	0.04802145
TEPSIN	TEPSIN. adaptor related protein complex 4 accessory protein	-0.105315433	4.54E-05	0.04802145
TNFSF8	TNF superfamily member 8	0.142393478	4.62E-05	0.04802145
SMAP2	small ArfGAP2	0.110449573	4.72E-05	0.04802145
ZBTB7A	zinc finger and BTB domain containing 7A	-0.120483466	5.08E-05	0.04802145
TXNIP	thioredoxin interacting protein	0.08387852	5.13E-05	0.04802145
SNORA38	small nucleolar RNA. H/ACA box 38	-0.203435872	5.32E-05	0.04802145
EGR2	early growth response 2	-0.458507358	5.54E-05	0.04802145
IL7R	interleukin 7 receptor	0.132259767	5.71E-05	0.04802145
TAGLN	Transgelin	-0.273617959	5.98E-05	0.04802145
SIDT2	SID1 transmembrane family member 2	-0.21403533	6.01E-05	0.04802145

Table 3. ADHD vs Control Patient Primary Microarray – Differentially Expressed Genes. Microarray analysis of differential gene expression from 94 medication-naive ADHD patients and 124 control subjects' peripheral blood mononuclear cells (PMBCs) found 21 differentially expressed genes following multiple comparison correction using the Benjamini-Hochberg method

3.2.2 Nominally differentially expressed genes formed highly biologically relevant gene networks and were over represented in a collection of ADHD-linked genes

Network analysis of all nominally differentially expressed genes using IPA formed 25 gene networks, all of which had a network score (-log10(p-value)) greater than 8, which was set as the threshold for relevance (Table 4). The gene network with the most significant network score of 44 had the annotation: Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Developmental Disorder. This network contained 35 genes, all of which were found to be nominally differentially expressed in the microarray experiment (Fig. 17). Enrichment analysis revealed a significant over representation of genes previously linked to ADHD, 48 of the 1793 differentially expressed genes were contained in the ADHD database (Table 5, Fisher's Exact Test p-value= 0.00214). A total of 26 canonical pathways were also found to be enriched for nominally differentially expressed genes (Table 6). RAR activation contained 33 nominally differentially expressed genes and was found to be the most enriched canonical pathway (p-value=0.00240).

Enrichr analysis found three gene ontology biological processes to be significantly enriched for nominally differentially expressed genes following adjustment for multiple testing (Table 7). The three gene ontology processes involved gene expression regulation with the top process titled "Regulation of transcription, DNA-templated (GO:0006355)" (p-value=0.00001897). None of the KEGG 2019 Human pathways were found to be significantly enriched following correction.

Top 25 Genes Networks Formed by Nominally Differentially Expressed Genes

Gene Network Annotation - Top Diseases and Functions	Network Score (-log10(p-value)	Focus Genes (n)
Cell-To-Cell Signaling and Interaction, Nervous System Development and Function, Developmental Disorder	44	35
Cellular Function and Maintenance, Connective Tissue Disorders, Developmental Disorder	42	34
Cell Cycle, Gene Expression, Developmental Disorder	42	34
Cell Morphology, Cellular Function and Maintenance, Embryonic Development	42	34
Cellular Development, Cellular Movement, Hematological System Development and Function	42	34
RNA Post-Transcriptional Modification, Nervous System Development and Function, Cardiovascular Disease	39	33
Cell Cycle, DNA Replication, Recombination, and Repair, Amino Acid Metabolism	39	33
Cell Cycle, Cell Morphology, Cellular Assembly and Organization	39	33
RNA Post-Transcriptional Modification, Connective Tissue Disorders, Developmental Disorder	39	33
RNA Post-Transcriptional Modification, Amino Acid Metabolism, Small Molecule Biochemistry	39	33
RNA Damage and Repair, Cellular Compromise, Cellular Development	37	32
Digestive System Development and Function, Embryonic Development, Organismal Development	37	32
Cell Death and Survival, Digestive System Development and Function, Hematological System Development and Function	37	32
RNA Post-Transcriptional Modification, Amino Acid Metabolism, Cardiovascular Disease	35	31
Gene Expression, Developmental Disorder, Embryonic Development	35	31
Post-Translational Modification, Connective Tissue Disorders, Developmental Disorder	35	31
Cell Morphology, Hematological System Development and Function, Hematopoiesis	33	30
Developmental Disorder, Embryonic Development, Organismal Development	33	30
Cellular Compromise, Cellular Function and Maintenance, Cancer	31	29
Cancer, Hematological Disease, Immunological Disease	31	29
Developmental Disorder, Hereditary Disorder, Metabolic Disease	29	28
Cellular Function and Maintenance, Cellular Assembly and Organization, Tissue Development	29	28
Organismal Survival, Cancer, Organismal Injury and Abnormalities	29	28
Nucleic Acid Metabolism, Small Molecule Biochemistry, Hematological Disease	27	27
Developmental Disorder, Hereditary Disorder, Organismal Injury and Abnormalities	27	27

Table 4. Top 25 Gene Networks formed by nominally differentially expressed genes. Ingenuity pathway analysis of all nominally differentially expressed genes formed 25 gene networks all of which reached the threshold for statistical significance.

Figure 17. Cell-To-Cell Signalling and Interaction - Nervous System Development and Function. The most significant gene network created by IPA analysis of nominally differentially expressed genes had the annotation Cell-To-Cell Signalling and Interaction Nervous System Development and Function and a network score (-log10(p-value) of 44. Legend: Red=Up-regulation, Green=Suppression, Colour Intensity=Magnitude of fold change

Nominally Differentially Expressed Genes Previously Implicated in ADHD or MPH response	
Gene Symbol	**Gene Name**
ADGRL3	adhesion G protein-coupled receptor L3
AK8	adenylate kinase 8
ATXN2	ataxin 2
AUTS2	AUTS2, activator of transcription and developmental regulator
CCSER1	coiled-coil serine rich protein 1
CDH23	cadherin related 23
CLOCK	clock circadian regulator
CLYBL	citrate lyase beta like
CPLX4	complexin 4
DPH6	diphthamine biosynthesis 6
EGFR	epidermal growth factor receptor
ELOC	elongin C
EREG	Epiregulin
FADS1	fatty acid desaturase 1
FANCL	Fanconi anemia complementation group L
GNAT2	G protein subunit alpha transducin 2
GSK3B	glycogen synthase kinase 3 beta
HKDC1	hexokinase domain containing 1
HTR1F	5-hydroxytryptamine receptor 1F
ITGAE	integrin subunit alpha E
LARP7	La ribonucleoprotein domain family member 7
LINGO2	leucine rich repeat and Ig domain containing 2
LOXL2	lysyl oxidase like 2
METTL3	methyltransferase like 3
MYBPC1	myosin binding protein C, slow type
MYO5B	myosin VB
NCAN	Neurocan
NEUROD6	neuronal differentiation 6
NR4A2	nuclear receptor subfamily 4 group A member 2
NTF3	neurotrophin 3
PEX5L	peroxisomal biogenesis factor 5 like
PRKAG2	protein kinase AMP-activated non-catalytic subunit gamma 2
PRKD1	protein kinase D1
PTPRG	protein tyrosine phosphatase, receptor type G
RNF144B	ring finger protein 144B
SEM1	SEM1, 26S proteasome complex subunit
SH2B1	SH2B adaptor protein 1
SLC6A2	solute carrier family 6 member 2
SLC9A9	solute carrier family 9 member A9
TIAM2	T-cell lymphoma invasion and metastasis 2
TLE4	transducin like enhancer of split 4
TPH2	tryptophan hydroxylase 2
TRIM32	tripartite motif containing 32
TSHZ2	teashirt zinc finger homeobox 2
TTC12	tetratricopeptide repeat domain 12
VAMP2	vesicle associated membrane protein 2
ZNF75A	zinc finger protein 75a
ZNF805	zinc finger protein 805

Table 5. Nominally Differentially Expressed Genes Previously Implicated in ADHD or MPH response. 48 of the 1793 nominally differentially expressed genes were contained in an ADHD database totalling 436 genes, this represented a statistically significant over-representation of genes (Fisher's Exact Test p-value= 0.00214). Full list of ADHD genes available in Pagerols et al. 2018 (Supplemental Information 1).

Statistically Significant Enrichment of Canonical Pathways	
Ingenuity Canonical Pathways	**Adjusted p-value**
RAR Activation	0.002398833
B Cell Receptor Signaling	0.003981072
Phospholipase C Signaling	0.009549926
Protein Kinase A Signaling	0.009549926
Huntington's Disease Signaling	0.012022644
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	0.013489629
Estrogen Receptor Signaling	0.014454398
Melanocyte Development and Pigmentation Signaling	0.016595869
Insulin Receptor Signaling	0.016595869
PDGF Signaling	0.016595869
PTEN Signaling	0.023988329
Neuregulin Signaling	0.031622777
Integrin Signaling	0.033884416
PI3K/AKT Signaling	0.033884416
ERK/MAPK Signaling	0.033884416
ERK5 Signaling	0.033884416
IL-3 Signaling	0.034673685
Ephrin Receptor Signaling	0.03801894
D-myo-inositol (1.4.5)-trisphosphate Degradation	0.038904514
Glucocorticoid Receptor Signaling	0.041686938
14-3-3-mediated Signaling	0.042657952
FLT3 Signaling in Hematopoietic Progenitor Cells	0.043651583
1D-myo-inositol Hexakisphosphate Biosynthesis II (Mammalian)	0.043651583
P2Y Purigenic Receptor Signaling Pathway	0.046773514
IL-4 Signaling	0.046773514
ErbB Signaling	0.046773514

Table 6. Statistically Significant Enrichment of Canonical Pathways. A total of 26 canonical pathways were identified by IPA to have an overrepresentation of nominally differentially expressed genes.

Statistically Significant Enrichment of Gene Ontology Processes		
Name	**P-value**	**Adjusted p-value**
Regulation of transcription, DNA-templated (GO:0006355)	4.89E-09	0.00001897
Regulation of nucleic acid-templated transcription (GO:1903506)	4.46E-06	0.005762
Regulation of gene expression (GO:0010468)	8.71E-06	0.008441

Table 7. Statistically Significant Enrichment of Gene Ontology Processes. Three gene ontology processes were identified by Enrichr as having an over representation of nominally differentially expressed genes

3.2.3 Replication of *Kruppel-Like Factor 4* differential expression in qPCR of a novel group of samples

The four genes; *Lysine Methyltransferase 5A (KMT5A), Nuclear RNA Export Factor 1 (NXF1), Kruppel Like Factor 4 (KLF4)* and *LRR Binding FLII Interacting Protein 1 (LRRFIP1)* with the strongest evidence of differential expression in the microarray study were chosen for targeted expression analysis in a novel sample of 30 medication-naïve ADHD individuals and 29 control samples. The standard curve method confirmed all four qPCR assays displayed acceptable levels of amplification efficiency and linearity. Following quality control the final number of ADHD and control samples differed for each gene (Table 6). *KLF4* (padj.mcmc= 5.643227e-05) was the only gene which reached the Benjamini-Hochberg adjusted p-value threshold for differential expression in the replication sample. The average log fold change in expression of each gene in ADHD cases relative to controls is presented in Fig. 18.

qPCR Replication Study – Sample Size and Results							
Gene	ADHD (n)	Control (n)	Log$_2$ Fold Change	Lower 95% Confidence Interval	Upper 95% Confidence Interval	pval. Markov chain Monte Carlo (mcmc)	p-adjusted.mcmc
KLF4	30	29	-0.833387324	-1.1997777	-0.429047614	1.41E+09	5.64E+08
KMT5A	29	28	-0.026937044	-0.4060032	0.33278274	0.885542373	0.885542373
LRR	29	28	0.123793755	-0.2265267	0.504820261	0.49874233	0.664989774
NXF1	30	28	0.178856893	-0.1808184	0.532474391	0.330062079	0.660124159
PES1	30	29	0.211853786	-0.1918116	0.580712612	NA	NA

Table 8. qPCR Replication Study – Sample Size and Results. qPCR results show *KLF4* expression in ADHD PMBCs was significantly lower than expression levels in control samples.
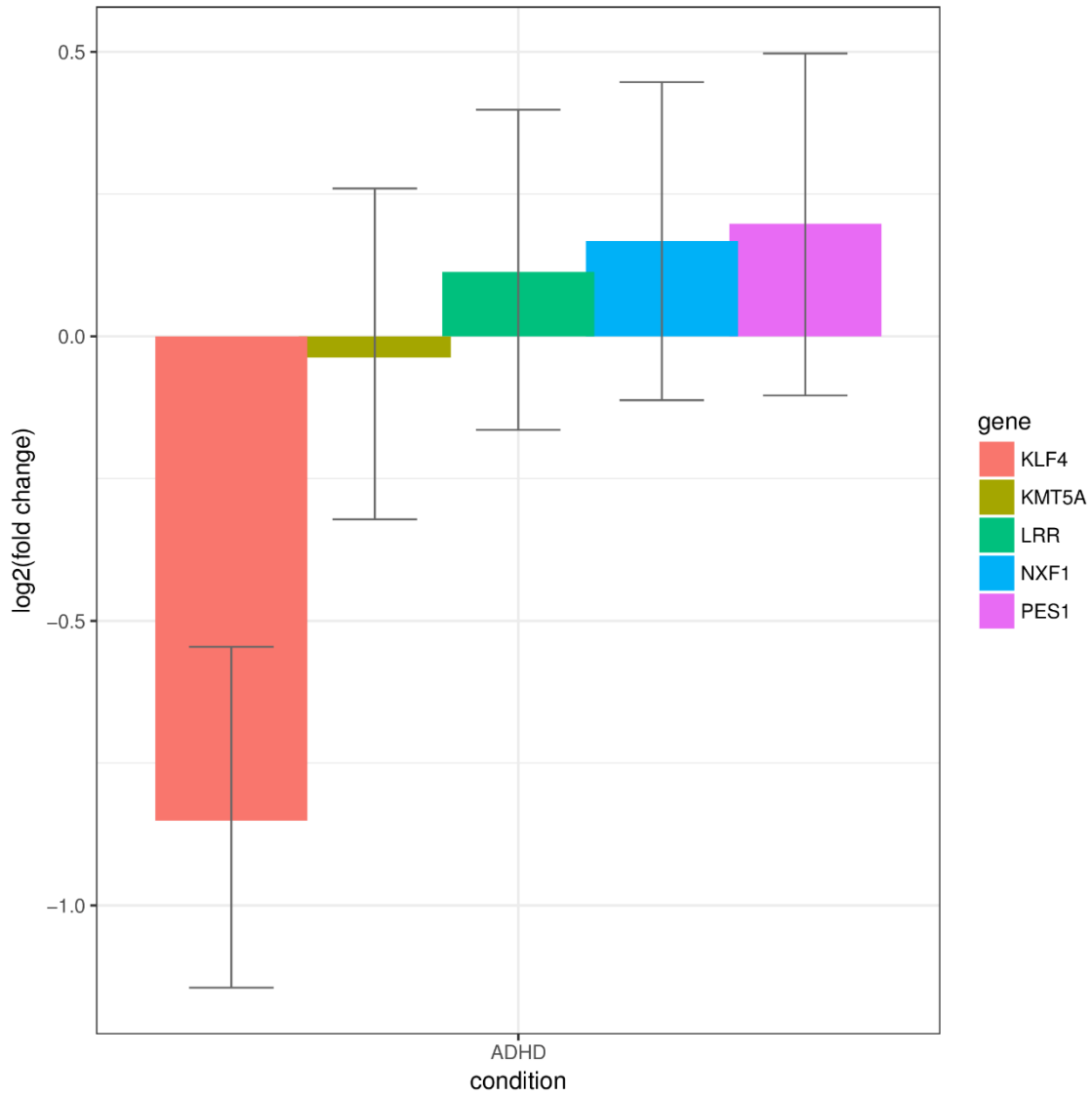
Figure 18. qPCR Replication Study shows *KLF4* expression is significantly lower in ADHD samples. qPCR results show *KLF4* expression in ADHD PMBCs was significantly lower than expression levels in control samples.

3.2.4 Correlation and replication of gene expression results in a second microarray

A comparison between the primary microarray analysis and a secondary analysis performed on a smaller sample set found a significant correlation between the log fold change (logFC) for probes' differential expression across studies (Fig. 19, Pearson correlation r=0.3152139, p<2.2e−16).

The logFC correlation between the two microarray studies was particularly high for the 21 genes which were found to be significantly differentially expressed in the primary microarray study (Fig. 20, r=0.8827791, p=1.179e−07). Eight of the 21 genes which reached the adjusted significance threshold in the first microarray were also found to significantly differentially expressed in the second microarray following adjustment of the p-value threshold for multiple comparisons using the Benjamini-Hochberg method (Table 7).
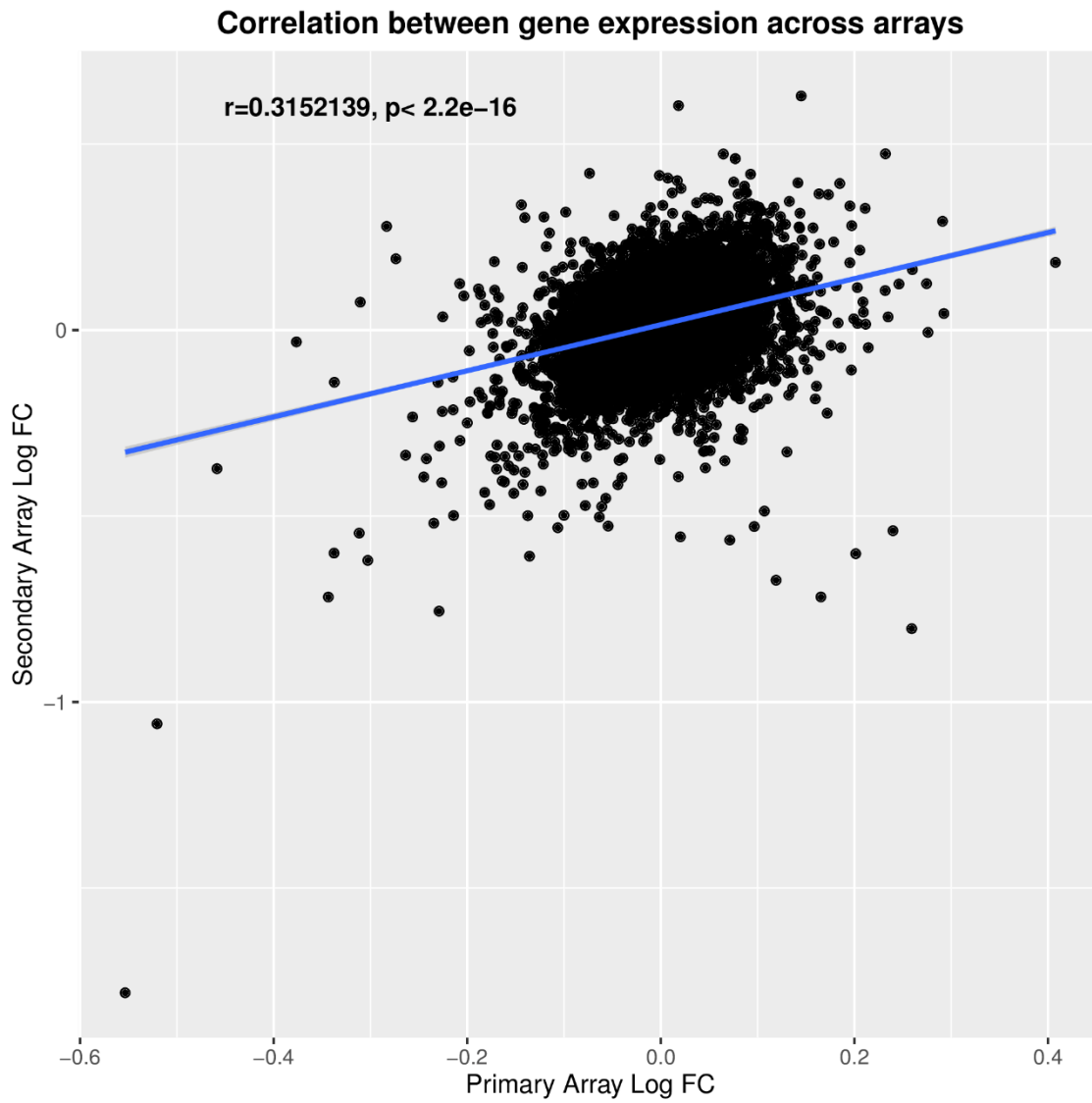
Figure 19. Significant correlation of probes' logFC across microarray studies. Analysis of all probes logFCs revealed statistically significant correlation across primary and secondary microarrays.

Figure 20. Correlation of significantly differentially expressed genes across microarray studies. Restriction of correlation analysis to the 21 genes found to be differentially expressed in the primary array found a statistically significant high degree of correlation with logFCs across primary and secondary arrays.

Replicated differentially expressed genes in secondary microarray				
Gene Symbol	**Name**	**logFC**	**P.Value**	**FDR**
KMT5A	lysine methyltransferase 5A	-0.20344685	0.000245633	0.00515829
IL7R	interleukin 7 receptor	0.25459685	0.001285199	0.01349459
RAB11FIP1	RAB11 family interacting protein 1	-0.29452447	0.002398875	0.01496377
LRRFIP1	LRR binding FLII interacting protein 1	-0.37253391	0.002850241	0.01496377
KLF4	Kruppel like factor 4	-0.49860849	0.004042922	0.01698027
SLA	Src like adaptor	0.19159416	0.008619182	0.03016714
EGR2	early growth response 2	-0.53134123	0.013252806	0.03975842
SNORA38	small nucleolar RNA, H/ACA box 38	-0.30254322	0.015201493	0.03990392

Table 9. Eight of the 21 differentially expressed genes in the primary microarray were also found to be differentially expressed in secondary microarray. Following threshold adjustment for multiple comparisons using the Benjamini-Hochberg method taking account of the 21 genes examined, eight genes were found to be significantly differentially expressed in the second microarray.

3.2.5 SNPs weighted with a combination of blood eQTL and differential gene expression data did not provide a polygenic risk score model which was significantly associated with ADHD status

PRS models encompassing all genes in the microarray and the PRS models restricted to the top 8 replicated genes were analysed separately. For both gene sets, individual PRS analyses of the three ADHD GWAS samples did not provide a significant predictive model at any of the p-value thresholds examined. Following a meta-analysis of the results, a meta p-value was determined for each p-value threshold. None of these meta p-values reached the threshold for statistical significance.

4. Discussion

This thesis describes the detailed investigation of *ADGRL3,* an ADHD-associated gene, using a genetically modified animal model and a case/control gene expression study in a human population which identified other novel genes which may be relevant to ADHD. ADGRL3 is a G-protein coupled receptor (GPCR) found at the post-synaptic membrane surface of neurons. It plays an important role in neuronal migration and synapse development, key processes in neurodevelopment and maintenance (Jackson et al., 2016; O'Sullivan et al., 2014, Sando 2019). The impairment of ADGRL3's synaptic functions is likely to be the cause of the numerous studies which have found associations between ADGRL3 and ADHD (Uhl et al. 2008a-b, Arcos-Burgos et al. 2010, Ribasés et al. 2011, Jain et al. 2012, Bruxel et al. 2015, Martinez et al. 2016, Gomez-Sanchez et al. 2017, Arcos-Burgos et al. 2019, Kappel et al. 2019). Alongside the genetic studies which have been previously described in the introduction to this thesis, there are a number of cross-species transgenic ADGRL3 models also described in the introduction which have shown anatomical, molecular and behavioural links to ADHD (Lange et al., 2012; van der Voet et al., 2016; Wallis et al., 2012). Evidence from both the animal models and human genetic studies make ADGRL3 one of the strongest ADHD-risk gene candidates.

Our behavioural analysis of *Adgrl3^{-/-}* mice reveals a number of phenotypes including extreme hyperactivity, learning and memory impairments, increased impulsivity and highly reduced levels of aggression. The RNA-sequencing of the PFC, hippocampus and striatum provides links between ADGRL3 inactivation and multiple new molecular pathways which may bring about the phenotypic pathologies observed in the behaviour of *Adgrl3^{-/-}* mice. In contrast to the strong phenotypes seen in *Adgrl3^{-/-}* mice, we provide an analysis of heterozygous *Adgrl3* mice which finds these mice behave in a largely similar manner to *Adgrl3^{+/+}* mice. *ADGRL3* genetic variants identified in human studies have shown a link between decreased ADGRL3 expression and increased risk of ADHD, suggesting the association between *ADGRL3* and ADHD in humans is due to alterations in gene expression levels rather than a complete abolishment of ADGRL3 (Martinez et al., 2016). This finding may have provided a reasonable expectation of moderate to strong phenotypes in *Adgrl3^{+/-}* mice as well as *Adgrl3^{-/-}* mice. However, the absence of detected phenotypes is not overly surprising considering the robust genetic redundancy often seen in genetically altered mice, particularly in regard to neurobiological phenotypes (Barbaric et al., 2007). The risk allele effect size of the *ADGRL3* variants in human populations is also in accordance with the results from *Adgrl3^{+/-}* mice, as variants are not completely penetrant, meaning not all subjects with the expression reducing risk allele

developed ADHD (Martinez et al., 2016). The reduction in ADGRL3 levels alone is therefore insufficient to cause ADHD development by itself but can confer increased susceptibility, particularly in combination with other genetic variants and exposure to other environmental factors during developmentally sensitive timeframes.

Previous studies have reported hyperactivity in *Adgrl3*-deficient mice and other transgenic organisms. The results reported here reproduce these findings while providing a greater degree of detail. By analysing distance travelled across behaviour paradigms we observed significantly increased locomotion of *Adgrl3$^{-/-}$* mice in the OF, LDB, BM, OR and SI tests. Alongside this hyperactive behaviour we visually observed what appeared to be an impairment in the gait of *Adgrl3$^{-/-}$* mice. Gait impairments may be relevant to ADHD as gait dysfunction has been reported in ADHD patients and can be modulated by current treatments such as MPH administration (Naruse et al., 2017, Auriel et al., 2006; Möhring et al., 2018). In order to impartially and quantitatively test for the presence of gait impairments we tested mice using the Viewpoint GaitLab system. *Adgrl3$^{-/-}$* mice showed significantly reduced stance time, possibly as a result of hyperactivity. Altered fore and hind limb BOS was also observed, possibly recapitulating the gait dysfunction present in some ADHD patients (Naruse et al., 2017).

Our results also show *Adgrl3$^{-/-}$* mice have a definite impairment in various forms of memory. These results are in contrast to prior results of working memory tests on *Adgrl3$^{-/-}$* mice which did not show any phenotype in a lever-based paradigm (Orsini et al., 2016) The absence of a phenotype in this previous study may be explained by the reduced food motivation of *Adgrl3$^{-/-}$* mice seen during our CPT analysis, as the lever-based working memory task depends on food as a motivator for task completion. It is also possible that our finding of memory impairments in *Adgrl3$^{-/-}$* mice is the result of our utilisation of different, more classical behavioural tests of memory, the investigation of different memory domains and/or the use of larger mice sample sizes. A similar pattern of deficits exists across behavioural paradigms designed to investigate different memory domains. For example, *Adgrl3$^{-/-}$* mice have reduced recall ability in both the OR and SI tests. The reduction in social memory performance is in contrast to the increased sociability of *Adgrl3$^{-/-}$* mice further suggesting the results are truly the result of memory impairments rather than reduced motivation to explore new environments/setups. Similarly, motivation seemed to be unaffected in the Barnes Maze as *Adgrl3$^{-/-}$* mice were capable of learning the position of the escape hole during both the acquisition and reversal phases. The reduced performance of *Adgrl3$^{-/-}$* mice (as evidenced by the increased number of primary errors

and longer time taken to escape) was therefore most likely the result of spatial memory impairments. Taking the results from the OR, SI and Barnes Maze, $Adgrl3^{-/-}$ mice together seems to show a cross domain, general impairment in memory formation and retrieval. These memory deficits may be analogous to the commonly occurring learning and memory endophenotypes of ADHD patients (Kofler et al. 2018).

The results of the CPT do not suggest a general cognitive impairment of $Adgrl3^{-/-}$ mice but an increase in impulsivity represented by the higher FAR of these mice compared to wildtype controls. $Adgrl3^{-/-}$ mice may therefore recapitulate the increased impulsivity commonly seen in ADHD patients (American Psychiatric Association, 2013). CPT was chosen because it is a powerful method which provides a close comparison to human studies in which CPT is commonly used. For example, a study by Fallgatter and colleagues using the CPT in combination with an ongoing EEG in a sample of ADHD patients showed that a risk allele of *ADGRL3* gene affects the number of omission errors during the CPT as well as several measures involved in cognitive response control (Fallgatter et al., 2013). The final behavioural phenotype detected was the reduced aggression of $Adgrl3^{-/-}$ mice as seen in the resident-intruder paradigm. It is possible that this reduced aggression is linked to the increased sociability of $Adgrl3^{-/-}$ mice observed in the social interaction test.

Alongside the identification of several behavioural phenotypes which may recapitulate some of the endophenotypes seen in ADHD patients, we also provide a number of potential molecular pathways through which ADGRL3 inactivation may bring about these behavioural characteristics. We identified these novel molecular pathways through hypothesis-free transcriptomic analysis of gene expression. Specifically we used RNA-sequencing to identify dysregulated gene expression in three ADHD-linked brain regions; PFC, hippocampus and striatum. The genes identified may play a role in the aetiology of ADHD. Previous studies have explored the transcriptome of $Adgrl3^{-/-}$ mice and found interesting results such as the disruption of cell adhesion molecules, synaptic neurotransmission related genes and calcium signalling protein producing genes (Orsini et al., 2016). While these initial results were promising, our RNA-sequencing study involved a significantly more robust approach. The inclusion of more $Adgrl3^{+/+}$ and $Adgrl3^{-/-}$ mice and the independent analysis of each mouse sample rather than pooling across animals provides greater statistical power, thus providing more in-depth insights into differential gene expression in each brain region studied. Given the well-powered nature of our study it is interesting to see a relatively low total number of significantly differentially expressed genes. The low number of genes identified suggests the impairment of specific

pathway or pathways rather than a brain wide perturbation of neurobiology. It is unsurprising given the small effect size of *ADGRL3* variants in human populations and the large amount of complexity underlying ADHD aetiology that the effects of *Adgrl3* inactivation is subtle rather than pronounced. While each single genetic variant may cause a subtle abnormality, in combination with other genetic variants and environmental risk factors it can bring about a markedly increased risk of developing ADHD.

The PFC has the largest number of differentially expressed genes in *Adgrl3*[-/-] mice. The PFC has long been implicated in ADHD pathology due to its pivotal role in cognitive processes required for self-regulation such as impulse control, appropriate attention levels and cognitive flexibility (Kim et al., 2011; Rossi et al., 2009). The importance of the PFC to ADHD is highlighted by it being the primary target site for ADHD pharmacological substances. For example, MPH is known to target DAT, eliciting activation of DA transmission in the PFC. The finding that *Slc6a3,* the gene coding for DAT, is the most dysregulated gene in the PFC of *Adgrl3*[-/-] mice is therefore highly relevant and supportive of *Adgrl3* transgenic mice as a model of ADHD. Prior work had identified *Slc6a3* as a differentially expressed gene in *Adgrl3*[-/-] mice but this is the first study to link this dysregulation to a specific brain region (Wallis et al., 2012, Orsini et al., 2016). Findings from other *ADGRL3* transgenic model organisms are also supportive of the impact of *ADGRL3* inactivation on the DA system. For example anatomical deficits have been observed in the DA system of *adgrl3.1*-deficient zebrafish (Lange et al., 2012). The impact of *Slc6a3* dysregulation in the PFC may also be seen in the substantia nigra (SN) and ventral tegmental area (VTA) as *Slc6a3* mRNA found in the PFC arrives via axonal transport from the nuclei in these brain regions (D'Ardenne et al., 2012).

The inactivation of DAT has been shown to result in remarkably similar phenotypes to those caused by *ADGRL3* inactivation in both mice and Drosophila. In Drosophila, the same light-sensitive hyperactivity phenotype is seen in response to RNAi mediated reduction of Adgrl or Dat (van der Voet et al., 2016). In addition, DAT transgenic mice show a pattern of spontaneous hyperactivity alongside memory and learning deficits and increased impulsivity similar to the phenotypes seen in *Adgrl3*[-/-] animals (Gainetdinov and Caron, 2001; 2000; Trinh et al., 2003; Wong et al., 2012, Yamashita et al., 2013). A study which involved the reduction of DAT levels to ~10% of typical expression levels showed that complete abolition of the DAT was not necessary to result in the hyperactivity phenotype (Kwiatkowski et al., 2017). While the knockdown and heterozygous DAT knockout mice show behavioural deficits with reduced severity to the complete Dat knockdown, equivalent phenotypes were not observed in our

behavioural analysis of *Adgrl3*[+/-] mice. The absence of clear behavioural phenotypes in *Adgrl3*[+/-] mice may be the result of the age at which the mice were tested, Dat heterozygous animals' phenotypes reduce with age and some phenotypes which are apparent during adolescence are not detected in adulthood (Ciampoli et al., 2017; França et al., 2016; Mereu et al., 2017). The failure to find novelty-induced locomotion in a knockdown mouse which had striatum Dat levels reduced to 40% is suggestive of region specific pathological thresholds over which Dat levels do not induce a behavioural phenotypes (Salahpour et al., 2007). It is possible that *Adgrl3*[+/-] mice do not cause Dat levels to slip below this threshold and thereby do not elicit a behavioural phenotype. Future studies could investigate *Adgrl3*[+/-] mice which also have other genetic variants or are exposed to environmental risk factors. In combination, these risk factors may result in the development of behavioural phenotypes. Taken together, the behavioural similarities seen in cross-species transgenic models of *ADGRL3* and DAT knockout/knockdown animals further strengthens the hypothesis that variants in *ADGRL3* may act through the DA pathway to confer an increased risk of ADHD.

The mechanism through which the DA system and *ADGRL3* is linked remains uncertain, but it seems likely that ADGRL3 may play an as yet undescribed role in DA synapse formation and function, similar to its known role in glutamatergic synapses. Mass spectrometry and affinity chromatography has been used to identify ADGRL3 interaction partners at glutamatergic synapses. These interaction partners include the fibronectin leucine-rich repeat transmembrane protein (FLRT) family of leucine-rich repeat proteins. In particular the olfactomedin (OLF) domain of ADGRL3 and ectodomain of FLRT3 are sufficient for protein association. The specific structure of this ADGRL3 OLF/FLRT3 interaction was determined using X-ray crystallography and found to contain a number of hydrophobic and charged residues (Ranaivoson et al., 2015). The reduction of FLRT3 levels *in vivo* has been shown to impact neural networks through abnormal afferent inputs and reduction in dendritic spine count. It is likely that the impairment of these processes which are vital for glutamatergic synapse development is the result of interference in the ADGRL3-FLRT3 complex (O'Sullivan et al. 2012). As well as interacting with FLRT3, ADGRL3 has been proposed to form a trimeric complex with FLRT3 and UNC5 at the synapse of neuronal cells. FLRT3 can form trans-synaptic complexes with both LPHN3 and UNC5 thereby mediating the development of intercellular contacts (Lu et al., 2015). Further studies have shown that ADGRL3 also forms a complex with FLRT2 which is mediated by the specific UNC5D protein homolog. This complex involves two ADGRL3 molecules therefore having a FLRT2:Unc5D:Lphn3

stoichiometry of 1:1:2 which can further hybridise to form an octamer super-complex with stoichiometry of 2:2:4 (Jackson et al. 2016).

ADGRL3-containing protein complexes are vital to the glutamatergic system (Ranaivoson et al., 2015). Impairments of the glutamatergic system are also known to play a role in the development of psychiatric disorders including ADHD (Belsham, 2001; Maltezos et al., 2014). Within the PFC, the glutamatergic and DA systems are tightly interlinked (Tseng and O'Donnell, 2004; Yuen et al., 2013). Impairments to both the DA and glutamatergic systems leading to dysregulation of excitatory and inhibitory signals is the leading hypothesis for the origin of psychotic disorders, emphasising the importance of these systems to normal brain development and function (Howes et al., 2015, Sohal & Rubenstein, 2019). The absence of evidence for differential gene expression of genes in the glutamatergic system of $Adgrl3^{-/-}$ mice may be due to compensatory effects from other genes or due to disturbances in glutamatergic gene expression being restricted to earlier developmental time periods. Future studies may be aimed at detecting the protein partners of ADGRL3 at DA synapses which may be different to the FLRT and teneurin proteins which ADGRL3 interacts with at glutamatergic synapses in the mouse cortex and hippocampus (O'Sullivan et al., 2012; O'Sullivan et al., 2014; Sando et al., 2019).

GSEA and pathway analysis performed in this study provided further insight into the neuropathology of the PFC caused by *Adgrl3* inactivation. Enricher analysis revealed a number of enriched KEGG pathways which were highly relevant to nervous system function and ADHD pathology in particular. The over representation of genes involved in the cocaine and amphetamine addiction pathways is interesting given the high rate of SUD in individuals with ADHD. In addition to a high rate of co-morbidity between ADHD and SUD, *ADGLR3* genetic variants in particular have been shown to increase the risk of developing SUD (Skoglund et al., 2017; Uhl et al., 2008a, 2008b, Arcos-Burgos M et al. 2019). Another KEGG pathway, the neuroactive ligand pathway, contains two genes *cholinergic receptor, nicotinic, alpha polypeptide 10 SUD (Chrna10)* and *cholinergic receptor, nicotinic, beta polypeptide 10 (Chrnb10)* which have also been associated with another form of addiction, specifically nicotine addiction (Keskitalo-Vuokko et al., 2011). It has been proposed that the cause of increased nicotine addiction in individuals with ADHD is the result of inhibited reinforcement processing as a result of impairments in the DA system (Kollins and Adcock, 2014). Mice models are supportive of this link, with *Dat* knockout animals showing dysregulation of nicotinic pathways and nicotine acetylcholine receptor β2-subunit deficient mice recapitulating

some of the core ADHD endophenotypes (Granon and Changeux, 2006, Weiss et al., 2007). Pharmacological testing of *Adgrl3-/-* mice is a promising avenue for future research given the disruption of genes involved in the MPH targeted DA system and the identification of dysregulation in addiction associated genes.

The number of genes exhibiting a logFC>|1| in the hippocampus and striatum (two and five respectively) was markedly lower than the number of genes in the PFC. Given the known role of ADGRL3 in the development and maintenance of glutamatergic synapses in the hippocampus as well as the cortex it might be considered surprising that only two hippocampal genes exhibited logFC>|1| dysregulation (O'Sullivan et al., 2012; Sando et al., 2019). Our behavioural findings of general learning and memory impairments in *Adgrl3-/-* mice is also suggestive of hippocampal abnormalities. The absence of large scale gene expression dysregulation may be the result of our RNA-sequencing study being conducted on samples from adult mice. The hippocampus has been suggested to play a reduced role in adults with ADHD in comparison to children with ADHD suggesting further studies of new born or adolescent *Adgrl3-/-* mice may find a larger degree of abnormal gene regulation in the hippocampus (Perlov et al., 2008; Plessen et al., 2006).

A recent study of mice which were selectively bred to be hyperactive found *Adgrl3* expression to be lower in the striatum of these mice in comparison to control mice (Sorokina et al., 2018). This finding coupled with the known importance of the striatum to ADHD pathophysiology makes the impact of *Adgrl3* inactivation on striatum gene expression highly interesting. In our transcriptome analysis of *Adgrl3-/-* mice, four *(arginine vasopressin, oxytocin, hypocretin and pmch)* of the five downregulated genes have previously been shown to regulate neuronal activity. *Arginine Vasopressin* and *Oxytocin* may play a role in bringing about the increased sociability and reduced aggression seen in *Adgrl3-/-* mice as they both code for neuro-hormone products involved in the regulation of social behaviour (Cataldo et al., 2018). These genes have also been implicated in the development of ASD which often appears comorbid with ADHD (Cataldo et al. 2018, Ghirardi et al., 2018). Alongside their impact on sociability, the identified neuro-hormones/peptides coding genes play an integral part in both motor and reward systems. The impulsivity and lack of motivation induced by food seen in *Adgrl3-/-* mice may be the result of disruptions to the mouse reward system while the gait impairments may be associated with a deficit in the motor system. The striatum's function as a modulator of the reward system is highly relevant to ADHD (van Hulst et al., 2017). The finding of differential neuro-

hormones/peptides gene expression was surprising given these genes are not normally expressed in the striatum of mice.

It is important to note the brain regions investigated do not function independently and disruptions in the PFC may lead to dysregulation in the striatum or vice versa. For example, disruption of the DA system in other brain regions such as the PFC can lead to dysregulated striatal neuropeptide levels (Engber et al., 1992). The DA system of the striatum is also relevant to ADHD. In particular, the level of striatal dopamine transporter protein has been shown to modulate MPH response in ADHD patients (Krause et al., 2005). Neuropeptides have also been shown to impact the DA system through modulation of dopamine release (Sulzer et al., 2016). Both *Hypocretin* and *Pmch* were dysregulated in the *Adgrl3*$^{-/-}$ mouse striatum and code for neuropeptides. The dysregulation of these genes could therefore be linked to the DA system impairments seen in the PFC of *Adgrl3*$^{-/-}$ mice. The PFC and striatum are interlinked brain regions with both cognitive and executive processes being dependent on functional connectivity between the regions (Antzoulatos and Miller, 2014). The importance of this connectivity is highlighted by the link between impairments in their connectivity and both psychosis and ASD (Padmanabhan et al., 2013; Simpson et al., 2010). Dysregulation of neuropeptides may impact the PFC's DA system and bring about deficits in learning, memory and impulsivity (Puig et al., 2014; Yates et al., 2016). The dysregulation of *Hypocretin* and *Pmch* may therefore have contributed to the memory endophenotypes seen in *Adgrl3* knockout mice.

Our findings from both behavioural and transcriptome studies are strongly supportive of *Adgrl3* knockout mice as a transgenic animal model of ADHD. The behavioural phenotypes including hyperactivity, impulsivity, learning and memory deficits recapitulate a number of the core endophenotypes seen in ADHD patients. Our transcriptomic results reveal impairments in the DA system, particularly in the PFC. The known link between the DA system and MPH coupled with our findings makes *Adgrl3* knockout mice a promising model for future pharmacological testing. The PFC and striatum specific annotated pathways provide insight into how *Adgrl3* inactivation impacts neurobiology and the relevance of these pathways to ADHD supports these brain regions as targets for further research in this and other *ADGRL3* model organisms. CRISPR-Cas9 is making the development of transgenic animal models increasingly simple and allowing specific mutations to be introduced into the model's genome. The creation of a transgenic mouse model which harbours a specific *ADGRL3* risk allele identified in human

studies would be particularly useful in deciphering the genetic link between non-coding *ADGRL3* variants and ADHD.

ADGRL3 is one of the main genes which has been repeatedly linked to ADHD through genetic and animal studies. Genetics has been shown to be the primary driver behind the development of ADHD with a meta-analysis of multiple large scale twin studies estimating the heritability of childhood and adolescent ADHD at between 0.7 and 0.8 (Nikolas & Burt, 2010). In order to identify the genes and specific genetic variants which lead to this high degree of heritability a large number of human genetic studies have been conducted. The largest Genome Wide Association Study (GWAS) have successfully identified 12 independent loci which confer a slightly increased risk of developing ADHD; however the majority of factors underlying ADHD´s genetic heritability remain undiscovered (Demontis et al. 2019). While the latest GWAS represents a major breakthrough, having reported the first genome wide significant associations for ADHD, the pathway between genetic variation and ADHD development remains unclear.

One method through which genetic variants can increase the risk of ADHD is through modulation of gene expression. Our microarray experiment identified a set of novel genes which showed significant differences in gene expression levels in the blood of ADHD patients in comparison to controls. These differences in gene expression may be the consequence of genetic variants known as expression quantitative trait loci (eQTL). eQTLs are often broken into two categories, cis-eQTLs which impact local gene expression and trans-eQTLs which impact the expression of distant genes. There is strong evidence that at least some of the ADHD associated risk alleles in *ADGRL3* are non-coding cis-eQTLs which impact the risk of developing ADHD through altering ADGRL3 protein levels in neural tissues rather than altering protein function directly (Martinez et al. 2017). Interestingly, our hypothesis free human transcriptome study found *ADGRL3* to be nominally differentially expressed in ADHD patients relative to control subjects, although this result was not significant following testing for multiple comparisons.

Environmental factors such as premature birth and maternal smoking during pregnancy have also been linked with increased risk of developing ADHD and this increased risk is likely to be at least partially mediated through modulation of gene expression (Halmøy et al., 2012; J. L. Zhu et al., 2014). Genetic and environmental factors are not independent of one another: the interplay between these factors, known as gene by environment interactions, is recognized as

playing an important role in the development of many psychiatric disorders including ADHD (Sanchez-Mora et al. 2015). Genetic variants including common polymorphisms and rare variations such as CNVs may interact with environmental stimuli resulting in altered gene expression. In order to remove one environmental stimulus which is likely to have a high impact on gene expression we restricted our microarray analysis to ADHD-medication naïve patients.

This study uses microarray technology to measure gene expression in Peripheral Mononuclear Blood Cells (PMBC) of ADHD cases and controls. Microarray assays are a relatively low cost, hypothesis free method of detecting differences in gene expression. Our microarray analysis examined 18227 genes and found 1793 of these to be nominally differentially expressed between ADHD cases and controls. Enrichment analysis for these nominally expressed genes showed they were over-represented in a previously published database of ADHD and MPH response associated genes (Pagerols et al. 2018). In addition, pathway analysis of these genes using IPA revealed a top gene network titled Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Developmental Disorder. Taken together these findings show strong support for the utility of PMBC gene expression analysis in examining ADHD pathophysiology.

Given the inherent variability of gene expression and the technical variability of microarray assays, we assessed the generalisability of our findings by performing a replication study looking at the correspondence between our large primary microarray sample and a secondary smaller sample. When our correlation analysis included data from all gene expression probes, we found a highly significant ($p < 2.2e{-}16$) but relatively low level of correlation across studies ($r=0.3152139$). Restricting the correlation analysis to the 21 genes which were significantly differentially expressed in the primary array following multiple correction comparison found a much higher rate of correlation ($r=0.8827791$, $p=1.179e{-}07$) across studies. The high degree of correlation for differentially expressed genes is highly supportive of the validity of this analysis to uncover genes which are relevant to ADHD pathophysiology or serve as potential biomarkers. Eight (*KMT5A, IL7R, RAB11FIP1, LRRFIP1, KLF4, SLA, EGR2* and *SNORA38*) of the 21 genes were found to be significantly differentially expressed in the secondary array following correction for multiple comparisons.

We performed a further replication study which examined gene expression using qPCR rather than microarray technology. Due to restrictions on sample availability we had to confine our

study to the top four genes from the primary microarray study; *KMT5A*, *NXF1*, *KLF4* and *LRRFIP1*. Only *KLF4* differential expression was replicated in this qPCR analysis. *KLF4* codes for a zinc-finger-containing protein which functions as a transcription factor impacting cellular proliferation and differentiation as well as apoptosis (McConnell & Yang 2010). KLF4 is expressed in neural stem cells and has been shown to impact axonal regeneration and radial neuronal migration through modulation of the JAK-STAT pathway (Qin & Zhang 2012). KLF4 has also been shown to play a vital role in cortical neurogenesis by promoting the self-renewal of neural progenitor cells (NPCs) via Stau1-mediated decay of neurogenesis-associated mRNAs (Moon et al. 2018). Another function of KLF4 is in the activation of microglia which results in neuro-inflammation and has been linked with amyloid plaque formation and Alzheimer's disease (Kaushik et al. 2010, Li et al. 2017). Our findings regarding *KLF4* differential expression which is replicated across three gene expression studies, coupled with its known role in neurogenesis and neuro-inflammation, make it a promising future target for ADHD research. This future research may include the creation and investigation of an animal model such as the *Adgrl3* constitutive knockout mouse model described in this thesis.

While the concordance of gene expression across tissues is a subject of continuing research, the clear advantage of peripheral tissues over central nervous system tissue is the ability to easily collect samples from living participants. It is important to note the use of blood tissue prohibits the assumption that any genes found to be differentially expressed in our study are also differentially expressed in the brain. Indeed any differential expression in the brain would likely be confined to specific brain regions and cell types rather than the whole of the central nervous system (Tylee et al. 2013). While taking account of these limitations, previous studies have shown that overall there is a significant degree of correlation ($r^2 \approx 0.5$) between gene expression in blood and neural tissues (Cai et al. 2010). The detection of differential gene expression in blood samples can therefore serve a dual role in the study of ADHD. Firstly, as differential gene expression between ADHD cases and controls may be an initial factor contributing to the development of ADHD or a downstream effect of ADHD pathophysiology, the detection of differentially expressed genes may serve as an entry point into an ADHD-relevant biological pathway. Secondly, differential gene expression in combination with other biomarkers may act as a method of identifying liability markers for the development of ADHD or isolating homogenous subtypes of ADHD (Thome et al. 2012).

The PRS analysis described in this thesis is one method of trying to combine gene expression and genotype data to develop a biomarker for ADHD risk. We found no evidence of association

between the expression-based PRSs and ADHD for any of the p-value thresholds examined in the individual ADHD/control datasets. A meta-analysis of the three PRS datasets also failed to return any models which reached the level of statistical significance. A future study using both cis-eQTL and differential gene expression data from ADHD relevant brain region(s) may produce more promising results.

The development of PRSs which can encompass data on thousands of genes may identify individuals who are at risk of developing ADHD prior to symptom development or provide risk estimates for specific ADHD subtypes or persistence into adulthood. PRS could then inform clinical treatment plans at the individual patient level. Currently the range of treatments for ADHD is rather limited, with a large degree of variation in patient response. The development of new effective treatments will require increased understanding of the aetiology of ADHD. In order to fully understand ADHD biology it is necessary to incorporate multiple different types of research. As highlighted by this thesis, each of these strands of research are not performed in isolation but inform one another. For example, the investigation of a single gene such as *ADGRL3* using an animal model requires the gene to be first identified using human studies. These studies may be genome based (such as GWAS) or transcriptome based (such as the microarray study described in this thesis). In turn, the animal models provide biological insight into the targets identified in human studies and may themselves provide further relevant gene targets. Only by combining research from disparate sources can we develop the thorough understanding on ADHD biology required for treatment development, which is the ultimate goal of translational science research.

5. Bibliography

Acosta, M. T., J. I. Velez, M. L. Bustamante, J. Z. Balog, M. Arcos-Burgos, and Maximilian Muenke. "A two-locus genetic interaction between LPHN3 and 11q predicts ADHD severity and long-term outcome." *Translational psychiatry* 1, no. 7 (2011): e17.

Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science*, 322(5903), 881-888.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.)

Arcos-Burgos, á., Jain, M., Acosta, M., Shively, S., Stanescu, H., Wallis, D., . . . Balog, J. (2010). A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Molecular psychiatry, 15*(11), 1053.

Arcos-Burgos, M., Vélez, J. I., Martinez, A. F., Ribasés, M., Ramos-Quiroga, J. A., Sánchez-Mora, C., ... & Casas, M. (2019). ADGRL3 (LPHN3) variants predict substance use disorder. *Translational psychiatry*, *9*(1), 42.

Bailey-Wilson, Joan. (2018). Parametric and Nonparametric Linkage Analysis. 10.1002/9780470015902.a0005403.pub2.

Balazs, J., & Kereszteny, A. (2017). Attention-deficit/hyperactivity disorder and suicide: A systematic review. *World journal of psychiatry*, *7*(1), 44.

Barrett, S. P., & Salzman, J. (2016). Circular RNAs: analysis, expression and potential functions. *Development*, *143*(11), 1838-1847.

Bier, E. (2005). Drosophila, the golden bug, emerges as a tool for human genetics. *Nature Reviews Genetics*, *6*(1), 9.

Brikell, I., Kuja-Halkola, R., & Larsson, H. (2015). Heritability of attention-deficit hyperactivity disorder in adults. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(6), 406-413.

Bruxel, E. M., Salatino-Oliveira, A., Akutagava-Martins, G. C., Tovo-Rodrigues, L., Genro, J. P., Zeni, C. P., ... & Rohde, L. A. (2015). LPHN 3 and attention-deficit/hyperactivity disorder: a susceptibility and pharmacogenetic study. *Genes, Brain and Behavior*, *14*(5), 419-427.

Bumgarner, R. (2013). Overview of DNA microarrays: types, applications, and their future. *Current protocols in molecular biology*, *101*(1), 22-1.

Cai, C., Langfelder, P., Fuller, T. F., Oldham, M. C., Luo, R., van den Berg, L. H., ... & Horvath, S. (2010). Is human blood a good surrogate for brain tissue in transcriptional studies?. *BMC genomics*, *11*(1), 589.

Catalanotto, C., Cogoni, C., & Zardo, G. (2016). MicroRNA in control of gene expression: an overview of nuclear functions. *International journal of molecular sciences*, *17*(10), 1712.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.

Chang, Z., Lichtenstein, P., & Larsson, H. (2012). The effects of childhood ADHD symptoms on early-onset substance use: a Swedish twin study. *Journal of abnormal child psychology*, *40*(3), 425-435.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, *14*(1), 128.

Chhangawala, S., Rudy, G., Mason, C. E., & Rosenfeld, J. A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome biology*, *16*(1), 131.

Corominas, J., Klein, M., Zayats, T., Rivero, O., Ziegler, G., Pauper. M., Neveling, K., Poelmans, G., Jansch, C., Svirin, E., Geissler, J., Weber, H., Reif, A., Arias-Vásquez, A., Galesloot, T., Kiemeney, L., & Buitelaar, J., Ramos-Quiroga, J. A., Cormand, B., Lesch, K. P. (2018). Identification of ADHD risk genes in extended pedigrees by combining linkage analysis and whole-exome sequencing. Molecular Psychiatry. 10.1038/s41380-018-0210-6.

Davletov, B. A., Shamotienko, O. G., Lelianova, V. G., Grishin, E. V., & Ushkaryov, Y. A. (1996). Isolation and biochemical characterization of a Ca2+-independent α-latrotoxin-binding protein. *Journal of Biological Chemistry*, *271*(38), 23239-23245.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., ... & Cerrato, F. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics*, *51*(1), 63.

Depledge, D. P., Srinivas, K. P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D. G., ... & Wilson, A. C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nature communications*, *10*(1), 754.

Doshi, J. A., Hodgkins, P., Kahle, J., Sikirica, V., Cangelosi, M. J., Setyawan, J., ... & Neumann, P. J. (2012). Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(10), 990-1002.

Dupont, C., Armant, D. R., & Brenner, C. A. (2009, September). Epigenetics: definition, mechanisms and clinical perspective. In *Seminars in reproductive medicine* (Vol. 27, No. 05, pp. 351-357). © Thieme Medical Publishers.

Epstein, C. B., & Butow, R. A. (2000). Microarray technology—enhanced versatility, persistent challenge. *Current opinion in biotechnology*, *11*(1), 36-41.

Faraone, S. V., & Larsson, H. (2019). Genetics of attention deficit hyperactivity disorder. *Molecular psychiatry*, *24*(4), 562-575.

Franke, B., Faraone, S. V., Asherson, P., Buitelaar, J., Bau, C. H. D., Ramos-Quiroga, J. A., ... & Lesch, K. P. (2012). The genetics of attention deficit/hyperactivity disorder in adults, a review. *Molecular psychiatry*, *17*(10), 960.

Franke, B., Michelini, G., Asherson, P., Banaschewski, T., Bilbow, A., Buitelaar, J. K., ... & Kuntsi, J. (2018). Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *European Neuropsychopharmacology*, *28*(10), 1059-1088.

Frantzi, M., Latosinska, A., & Mischak, H. (2019). Proteomics in Drug Development: The Dawn of a New Era?. *PROTEOMICS–Clinical Applications*, *13*(2), 1800087.

Gomez-Sanchez, C. I., Carballo, J. J., Riveiro-Alvarez, R., Soto-Insuga, V., Rodrigo, M., Mahillo-Fernandez, I., ... & Ayuso, C. (2017). Pharmacogenetics of methylphenidate in childhood attention-deficit/hyperactivity disorder: long-term effects. *Scientific reports*, *7*(1), 10391.

Gudmundsson, O. O., Walters, G. B., Ingason, A., Johansson, S., Zayats, T., Athanasiu, L., ... & Jonsson, L. (2019). Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Translational psychiatry*, *9*(1), 1-9.

Haavik, J., Halmøy, A., Lundervold, A. J., & Fasmer, O. B. (2010). Clinical assessment and diagnosis of adults with attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics, 10*(10), 1569-1580.

Halmøy, A., Klungsøyr, K., Skjærven, R., & Haavik, J. (2012). Pre-and perinatal risk factors in adults with attention-deficit/hyperactivity disorder. *Biological psychiatry*, *71*(5), 474-481.

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., & Gloriam, D. E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. *Nature reviews Drug discovery*, *16*(12), 829.

Hawi, Z., Cummins, T. D. R., Tong, J., Johnson, B., Lau, R., Samarrai, W., & Bellgrove, M. A. (2015). The molecular genetic architecture of attention deficit hyperactivity disorder. *Molecular psychiatry*, *20*(3), 289.

Holstein, D. H., Vollenweider, F. X., Geyer, M. A., Csomor, P. A., Belser, N., & Eich, D. (2013). Sensory and sensorimotor gating in adult attention-deficit/hyperactivity disorder (ADHD). *Psychiatry research*, *205*(1-2), 117-126.

Iwamoto, N., & Shimada, T. (2018). Recent advances in mass spectrometry-based approaches for proteomics and biologics: Great contribution for developing therapeutic antibodies. *Pharmacology & therapeutics*, *185*, 147-154.

Jackson, VA., del Toro, D., Carrasquero, M., Roversi, P., Harlos, K., Klein, R., & Seiradake, E. (2015). Structural Basis of Latrophilin-FLRT Interaction. *Structure, 23*(4), 774-781.

Jain, M., Vélez, J. I., Acosta, M. T., Palacio, L. G., Balog, J., Roessler, E., ... & Lopera, F. (2012). A cooperative interaction between LPHN3 and 11q doubles the risk for ADHD. *Molecular psychiatry*, *17*(7), 741-747.

Kappel, D. B., Schuch, J. B., Rovaris, D. L., da Silva, B. S., Müller, D., Breda, V., ... & Grevet, E. H. (2019). ADGRL3 rs6551665 as a Common Vulnerability Factor Underlying Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder. *Neuromolecular medicine*, *21*(1), 60-67.

Koch, M., & Schnitzler, H. U. (1997). The acoustic startle response in rats—circuits mediating evocation, inhibition and potentiation. *Behavioural brain research*, *89*(1-2), 35-49.

Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I. H., ... & Saada, A. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PloS one*, *7*(12), e49475.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... & McDermott, M. G. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, *44*(W1), W90-W97.

Lamsal, R., & Zwicker, J. D. (2017). Economic evaluation of interventions for children with neurodevelopmental disorders: opportunities and challenges. *Applied health economics and health policy*, *15*(6), 763-772.

Lange, M., Norton, W., Coolen, M., Chaminade, M., Merker, S., Proft, F., . . . Bally-Cuif, L. (2012). The ADHD-susceptibility gene lphn3. 1 modulates dopaminergic neuron formation and locomotor activity during zebrafish development. *Molecular psychiatry, 17*(9), 946-954.

Li, D., Sham, P. C., Owen, M. J., & He, L. (2006). Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). *Human molecular genetics*, *15*(14), 2276-2284.

Li, Z., Chang, S. H., Zhang, L. Y., Gao, L., & Wang, J. (2014). Molecular genetic studies of ADHD and its candidate genes: a review. *Psychiatry research*, *219*(1), 10-24.

Lionel, A. C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., ... & Wei, J. (2011). Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Science translational medicine*, *3*(95), 95ra75-95ra75.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature News*, *456*(7218), 18-21.

Márquez-Luna, Carla, et al. "Multiethnic polygenic risk scores improve risk prediction in diverse populations." *Genetic epidemiology* 41.8 (2017): 811-823.

Martinez, A. F., Abe, Y., Hong, S., Molyneux, K., Yarnell, D., Löhr, H., . . . Muenke, M. (2016). An Ultraconserved Brain-Specific Enhancer Within ADGRL3 (LPHN3) Underpins

Attention-Deficit/Hyperactivity Disorder Susceptibility. *Biological psychiatry, 80*(12), 943-954.

Meijer, D., & Burbach, J. P. H. (2019). Latrophilin's social protein network. Frontiers in neuroscience, 13, 643.

Meng, S., Zhou, H., Feng, Z., Xu, Z., Tang, Y., & Wu, M. (2019). Epigenetics in Neurodevelopment: Emerging Role of Circular RNA. *Frontiers in cellular neuroscience*, *13*.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American journal of human genetics*, *7*(3), 277.

Nikolas, M. A., & Burt, S. A. (2010). Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: a meta-analysis. *Journal of abnormal psychology*, *119*(1), 1.

O'Sullivan, M. L., Martini, F., von Daake, S., Comoletti, D., & Ghosh, A. (2014). LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5. *Neural Development, 9*(1), 7.

Oerlemans, A. M., Burmanje, M. J., Franke, B., Buitelaar, J. K., Hartman, C. A., & Rommelse, N. N. (2016). Identifying unique versus shared pre-and perinatal risk factors for ASD and ADHD using a simplex-multiplex stratification. *Journal of abnormal child psychology*, *44*(5), 923-935.

Orsini, C. A., Setlow, B., DeJesus, M., Galaviz, S., Loesch, K., Ioerger, T., & Wallis, D. (2016). Behavioral and transcriptomic profiling of mice null forLphn3, a gene implicated in ADHD and addiction. *Molecular Genetics & Genomic Medicine, 4*(3), 322-343.

Pagerols, M., Richarte, V., Sánchez-Mora, C., Rovira, P., Artigas, M. S., Garcia-Martínez, I., ... & Victor, M. M. (2018). Integrative genomic analysis of methylphenidate response in attention-deficit/hyperactivity disorder. *Scientific reports*, *8*(1), 1-11.

Palacio, J. D., Castellanos, F. X., Pineda, D. A., Lopera, F., Arcos-Burgos, M., Quiroz, Y. T., ... & Bailey-Wilson, J. (2004). Attention-deficit/hyperactivity disorder and comorbidities in 18 Paisa Colombian multigenerational families. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(12), 1506.

Perona, M. T., Waters, S., Hall, F. S., Sora, I., Lesch, K. P., Murphy, D. L., ... & Uhl, G. R. (2008). Animal models of depression in dopamine, serotonin and norepinephrine transporter knockout mice: prominent effects of dopamine transporter deletions. *Behavioural pharmacology*, *19*(5-6), 566.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256), 748-752.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., ... & Cho, E. K. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444.

Regan, S. L., Hufgard, J. R., Pitzer, E. M., Sugimoto, C., Hu, Y. C., Williams, M. T., & Vorhees, C. V. (2019). Knockout of latrophilin-3 in Sprague-Dawley rats causes hyperactivity, hyper-reactivity, under-response to amphetamine, and disrupted dopamine markers. *Neurobiology of disease*, 104494.

Reimherr, F. W., Marchant, B. K., Gift, T. E., & Steans, T. A. (2017). ADHD and Anxiety: Clinical Significance and Treatment Implications. *Current Psychiatry Reports, 19*(12).

Reuter, I., Knaup, S., Romanos, M., Lesch, K. P., Drepper, C., & Lillesaar, C. (2016). Developmental exposure to acetaminophen does not induce hyperactivity in zebrafish larvae. *Journal of Neural Transmission*, *123*(8), 841-848.

Ribasés, M., Ramos-Quiroga, J. A., Sánchez-Mora, C., Bosch, R., Richarte, V., Palomar, G., . . . Casas, M. (2010). Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study. *Genes, Brain and Behavior, 10*(2), 149-157.

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., ... & Pers, T. H. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421.

Rivero, O., Selten, M. M., Sich, S., Popp, S., Bacmeister, L., Amendola, E., ... & Schmitt, A. G. (2015). Cadherin-13, a risk gene for ADHD and comorbid disorders, impacts GABAergic function in hippocampus and cognition. *Translational psychiatry*, *5*(10), e655.

Rovira, P., Demontis, D., Sánchez-Mora, C., Zayats, T., Klein, M., Mota, N. R., ... & Arribas, L. (2019). Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *bioRxiv*, 589614.

Sando, R., Jiang, X., & Südhof, T. C. (2019). Latrophilin GPCRs direct synapse specificity by coincident binding of FLRTs and teneurins. Science, 363(6429), eaav7969.

Schbath, S., Martin, V., Zytnicki, M., Fayolle, J., Loux, V., & Gibrat, J. F. (2012). Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. *Journal of Computational Biology*, *19*(6), 796-813.

Sciberras, E., Mulraney, M., Silva, D., & Coghill, D. (2017). Prenatal risk factors and the etiology of ADHD—review of existing evidence. *Current psychiatry reports*, *19*(1), 1.

Shishido, E., Aleksic, B., & Ozaki, N. (2014). Copy-number variation in the pathogenesis of autism spectrum disorder. *Psychiatry and clinical neurosciences*, *68*(2), 85-95.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121-132.

Skoglund, C., Brandt, L., D'Onofrio, B., Larsson, H., & Franck, J. (2017). Methylphenidate doses in Attention Deficit/Hyperactivity Disorder and comorbid substance use disorders. *European Neuropsychopharmacology, 27*(11), 1144-1152.

Sohal, V. S., & Rubenstein, J. L. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Molecular psychiatry*, *24*(9), 1248-1257.

Soleimani, R., Salehi, Z., Soltanipour, S., Hasandokht, T., & Jalali, M. M. (2018). SLC6A3 polymorphism and response to methylphenidate in children with ADHD: A systematic review and meta-analysis. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *177*(3), 287-300.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631-656.

Südhof, T. C. (2001). α-Latrotoxin and its receptors: neurexins and CIRL/latrophilins. *Annual review of neuroscience*, *24*(1), 933-962.

Sugrue, L. P., & Desikan, R. S. (2019). What Are Polygenic Scores and Why Are They Important?. *Jama*, *321*(18), 1820-1821.

Sujan, A. C., Rickert, M. E., Öberg, A. S., Quinn, P. D., Hernández-Díaz, S., Almqvist, C., ... & D'Onofrio, B. M. (2017). Associations of maternal antidepressant use during the first trimester of pregnancy with preterm birth, small for gestational age, autism spectrum disorder, and attention-deficit/hyperactivity disorder in offspring. *Jama*, *317*(15), 1553-1562.

Sullivan, P. F., Fan, C., & Perou, C. M. (2006). Evaluating the comparability of gene expression in blood and brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *141*(3), 261-268.

Teare, M. D., & Barrett, J. H. (2005). Genetic linkage studies. *The Lancet*, *366*(9490), 1036-1044.

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, *19*(9), 581.

Tylee, D. S., Kawaguchi, D. M., & Glatt, S. J. (2013). On the outside, looking in: A review and evaluation of the comparability of blood and brain "-omes". *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *162*(7), 595-603.

Uhl G. R., Drgon T., Johnson C., Fatusin O. O., Liu Q. R., Contoreggi C., et al. 2008b. "Higher order" addiction molecular genetics: convergent data from genome-wide association in humans and mice. Biochem. Pharmacol. 75:98–111.

Uhl G. R., Drgon T., Liu Q. R., Johnson C., Walther D., Komiyama T., et al. 2008a. Genome-wide association for methamphetamine dependence: convergent results from 2 samples. Arch. Gen. Psychiatry 65:345–355.

van der Voet, M., Harich, B., Franke, B., & Schenck, A. (2015). ADHD-associated dopamine transporter, latrophilin and neurofibromin share a dopamine-related locomotor signature in Drosophila. *Molecular psychiatry, 21*(4), 565-573.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7-24.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), 5-22.

Vizzini, L., Popovic, M., Zugna, D., Vitiello, B., Trevisan, M., Pizzi, C., ... & Richiardi, L. (2019). Maternal anxiety, depression and sleep disorders before and during pregnancy, and preschool ADHD symptoms in the NINFEA birth cohort study. *Epidemiology and psychiatric sciences*, *28*(5), 521-531.

Wainschtein, P., Jain, D. P., Yengo, L., Zheng, Z., Cupples, L. A., Shadyab, A. H., ... & Kooperberg, C. (2019). Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020.

Wallis, D., Hill, D. S., Mendez, I. A., Abbott, L. C., Finnell, R. H., Wellman, P. J., & Setlow, B. (2012). Initial characterization of mice null for Lphn3, a gene implicated in ADHD and addiction. *Brain Research, 1463*, 85-92.

Weinhold, B. (2006). Epigenetics: the science of change.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., ... & Parkinson, H. (2013). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, *42*(D1), D1001-D1006.

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., ... & Zhernakova, A. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics*, *45*(10), 1238-1243.

Wilens, T. E., Martelon, M., Joshi, G., Bateman, C., Fried, R., Petty, C., & Biederman, J. (2011). Does ADHD predict substance-use disorders? A 10-year follow-up study of young adults with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(6), 543-553.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., ... & Amin, N. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, *46*(11), 1173.

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., ... & Sadowski, N. (2019). Nanopore native RNA sequencing of a human poly (A) transcriptome. *Nature methods*, 1-9.

Zhao, S., Zhang, Y., Gamini, R., Zhang, B., & von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific reports*, *8*(1), 4781.

Zhu, J. L., Olsen, J., Liew, Z., Li, J., Niclasen, J., & Obel, C. (2014). Parental smoking during pregnancy and ADHD in children: the Danish national birth cohort. *Pediatrics*, *134*(2), e382-e388.

Zulauf, C. A., Sprich, S. E., Safren, S. A., & Wilens, T. E. (2014). The complicated relationship between attention deficit/hyperactivity disorder and substance use disorders. *Current psychiatry reports*, *16*(3), 436.

6. Publications
Peer Review Publications Linked to Doctoral Thesis Research

- **Mortimer N**, Ganster T, O'Leary A, Popp S, Freudenberg F, Reif A, Soler Artigas M, Ribasés M, Ramos-Quiroga JA, Lesch KP, Rivero O. Dissociation of impulsivity and aggression in mice deficient for the ADHD risk gene Adgrl3: Evidence for dopamine transporter dysregulation. Neuropharmacology. 2019 Sep 15;156:107557. doi: 10.1016/j.neuropharm.2019.02.039. Epub 2019 Mar 6.
- Dalla Vecchia E*, **Mortimer N***, Palladino VS*, Kittel-Schneider S, Lesch KP, Reif A, Schenck A, Norton WHJ. Cross-species models of attention-deficit/hyperactivity disorder and autism spectrum disorder: lessons from CNTNAP2, ADGRL3, and PARK2. Psychiatr Genet. 2019 Feb;29(1):1-17. *Review.* *Joint lead author

Separate Peer Review Publications

- Fazzari P, **Mortimer N**, Yabut O, Vogt D, Pla R. Cortical distribution of GABAergic interneurons is determined by migration time and brain size [published online ahead of print, 2020 Jun 25]. *Development*. 2020;dev.185033.
- Pla R, Stanco A, Howard MA, Rubin AN, Vogt D, **Mortimer N**, Cobos I, Potter GB, Lindtner S, Price JD, Nord AS, Visel A, Schreiner CE, Baraban SC, Rowitch DH, Rubenstein JLR. Dlx1 and Dlx2 Promote Interneuron GABA Synthesis, Synaptogenesis, and Dendritogenesis. Cereb Cortex. 2018 Nov 1;28(11):3797-3815. doi:10.1093/cercor/bhx241.
- Trofimov A, Strekalova T, **Mortimer N**, Zubareva O, Schwarz A, Svirin E, Umriukhin A, Svistunov A, Lesch KP, Klimenko V. Postnatal LPS Challenge Impacts Escape Learning and Expression of Plasticity Factors Mmp9 and Timp1 in Rats: Effects of Repeated Training. Neurotox Res. 2017 Aug;32(2):175-186. Epub 2017 Apr 18. PubMed PMID: 28421528;
- Tropea D, **Mortimer N**, Bellini S, Molinos I, Sanfeliu A, Shovlin S, McAllister D, Gill M, Mitchell K, Corvin A. Expression of nuclear Methyl-CpG binding protein 2 (Mecp2) is dependent on neuronal stimulation and application of Insulin-like growth factor 1. Neurosci Lett. 2016 May 16;621:111-116. Epub 2016 Apr 11.
- Pini G, Congiu L, Benincasa A, DiMarco P, Bigoni S, Dyer AH, **Mortimer N**, Della-Chiesa A, O'Leary S, McNamara R, Mitchell KJ, Gill M, Tropea D. Illness Severity, Social and Cognitive Ability, and EEG Analysis of Ten Patients with Rett Syndrome Treated with Mecasermin (Recombinant Human IGF-1). Autism Res Treat. 2016;2016:5073078. doi: 10.1155/2016/5073078. Epub 2016 Jan 26.

7. Curriculum Vitae

EDUCATION & QUALIFICATIONS

Sep. 2015 - July 2020 University of Würzburg, Germany and Vall d'Hebron Research Institute, Spain: Completing jointly supervised double degree program of the Julius-Maximilians- Universität Würzburg and Universitat de Barcelona. Awarded Marie Sklodowska-Curie Early Stage Researcher Fellowship

Sep. 2018 - May 2019 National Institute for Bioprocessing Research and Training, Ireland: L9 Postgraduate Certificate in Biopharmaceutical Science

Sep. 2010- Sep. 2015   Trinity College Dublin, Ireland: Research Masters in Science, Genetics. Bachelor's Degree in Natural Science, Moderatorship in Genetics. Recipient of Government of Ireland Postgraduate Scholarship

Jun. 2013 – Sep. 2013 University of California, San Francisco: American-Ireland Fund Vincent Scholarship for Summer Internship in Prof. John Rubenstein's lab at UCSF

SCIENTIFIC EXPERIENCE

Post-Doctoral Scientist in Genomics and Drug Discovery, Data Science Department, Genetics Guided Dementia Discovery (G2D2), Eisai. March 2020-Current Role
Currently completing Eisai Transitional Postdoctoral Fellowship in Genomics and Drug Discovery. Part of the maiden cohort of postdocs employed by G2D2 Eisai, my role involves exposure to all aspects of the drug development process across research departments. Specific activities include an analysis of genetically defined Alzheimer's disease targets as candidates for entry in the therapeutic development pipeline and the functional characterisation of genetic variants in current drug targets.

Health Analyst, S3 Connected Health Dec. 2018-April. 2019
Researched clinical and scientific literature on disease areas to inform the development of novel digital solutions aimed at gathering data from patients and health care professionals. Co-developed data strategy for digital solutions to provide real world evidence.

Research & Development Scientist, Serosep Ltd. Sept. 2018-Dec. 2018
Developed and managed an environmentally friendly fixative. Established a product development framework in line with ISO13485 requirements. Built partnerships with clinical sites for testing and managed intellectual property applications.

PhD Thesis Title: ADHD Genetics in Mouse and Man. Supervisors: Prof. Klaus-Peter Lesch, Univeristy of Würzburg & Dr. Marta Ribases, Vall d'Hebron Research Institute

Awarded a Marie-Curie Early Stage Researcher Fellowship to pursue a joint-PhD in two world-class academic research labs. This fellowship is recognised as among the most prestigious in Europe. Research was conducted within the European wide MiND group which aims to greatly increase knowledge of ADHD and Autism by bringing together the most gifted young researchers from disparate fields. Research involved the development and utilisation of both dry and wet lab skills. Project was focused on the role of genetics and transcriptomics in the neuropathology of ADHD. Human study began with extraction of nucleic acids from blood samples followed by microarray analysis of RNA and identification

of differential gene expression. Data analysis required development of an R-based data analysis pipeline for high-throughput transcriptomic data. Advanced statistical models were used to detect significant differences in gene expression between cases and controls followed by pathway analysis of these genes using specialised bioinformatic software. ADHD-associated genes were further investigated in genetically modified animal models. Models were examined using a combination of behavioural testing and molecular techniques including RNA-sequencing and immunohistochemistry.

Master's Project: Received the only Government of Ireland Postgraduate Scholarship from the Irish Research Council to pursue an unstructured research masters. This project investigated the potential of IGF1 as a treatment for the rare-disease Rett Syndrome. IGF1's effect on Rett Syndrome patients was analysed in a small scale single arm clinical study. Primary neuronal cell culture was used to investigate the effects of IGF1 administration on gene and protein expression using quantitative polymerase chain reaction and western blotting, respectively.

_____

Niall Mortimer