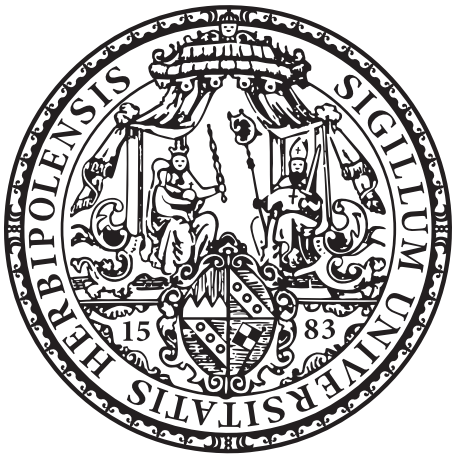
The role of nuclear architecture in the context of antigenic variation in Trypanosoma brucei

Über die Rolle der Zellkernarchitektur im Kontext von Antigenvariation in *Trypanosoma brucei*



Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

submitted by

Laura Müller-Hübner

from

Hildesheim, Germany

Würzburg, 2019

Laura Müller-Hübner: *The role of nuclear architecture in the context of antigenic variation in Trypanosoma brucei*

Submitted on:

Primary supervisor: Prof. Dr. T. Nicolai Siegel
Supervisor (second): Prof. Dr. Markus Engstler
Supervisor (third): Prof. Dr. Konrad U. Förstner
Supervisor (fourth): Dr. Susanne Fenz

Chairperson: Prof. Dr. Thomas Dandekar

Date of public defense:
Date of receipt of certificates:

Abstract

Antigenic variation of surface proteins is a commonly used strategy among pathogens to evade the host immune response [63]. The mechanism underlying antigenic variation relies on monoallelic exclusion of a single gene from a hypervariable multigene family combined with repeated, systematic changes in antigen expression. In many systems, these gene families are arranged in subtelomeric contingency loci that are subject to both transcriptional repression and enhanced mutagenesis and recombination [16].

Eviction of a selected gene from a repressed antigen repertoire can be achieved e.g. by recombination into a dedicated, transcriptionally permissive site or by local epigenetic alterations in chromatin composition of the selected gene.

Both processes are ultimately affected by genome architecture. Architectural proteins controlling antigenic variation have, however, remained elusive in any pathogen.

The unicellular protozoan parasite *Trypanosoma brucei* evades the host immune response by periodically changing expression of a single variant surface glycoprotein (VSG) from a repertoire of ~3000 *VSG* genes – the largest mutually exclusively expressed gene family described today. To activate a selected *VSG* gene, it needs to be located in a dedicated expression site that becomes subject to relocation into a distinct, transcriptionally active subnuclear compartment, the expression site body (ESB). Whereas this emphasizes the importance of nuclear architecture in regulating antigen expression in *T. brucei*, the mechanisms underlying spatial positioning of DNA in *T. brucei* are not well understood.

In this study I applied genome-wide chromosome conformation capture (Hi-C) to obtain a comprehensive picture of the *T. brucei* genome in three dimensions, both in procyclic and bloodstream form parasites. Hi-C revealed a highly structured nucleus with megabase chromosomes occupying distinct chromosome territories. Further, specific *trans* interactions between chromosomes, among which are clusters of centromeres, *rRNA* genes and *procyclins* became apparent. With respect to antigenic variation, Hi-C revealed a striking compaction of the subtelomeric *VSG* gene repertoire and a strong clustering of transcriptionally repressed *VSG*-containing expression sites. Further, Hi-C analyses confirmed the spatial separation of the actively transcribed from the silenced expression sites in three dimensions.

I further sought to characterize architectural proteins mediating nuclear architecture in *T. brucei*. Whereas CTCF is absent in non-metazoans, we found cohesin to be expressed throughout the cell cycle, emphasiz-

ing a function beyond sister chromatid cohesion in S-phase.
By Chromatin-Immunoprecipitation with sequencing (ChIPseq), I found cohesin enrichment to coincide with the presence of histone H3 variant (H3.V) and H4 variant (H4.V). Most importantly, cohesin and the histone variants were enriched towards the *VSG* gene at silent and active expression sites.
While the deletion of *H3.V* led to increased clustering of expression sites in three dimensions and increased chromatin accessibility at expression site promoters, the additional deletion of *H4.V* increased chromatin accessibility at expression sits even further.
RNAseq showed that mutually exclusive *VSG* expression was lost in *H3.V* and *H4.V* single and double deletion mutants. Immunofluorescence imaging of surface *VSGs*, flow cytometry and single-cell RNAseq revealed a progressive loss of *VSG-2* expression, indicative of an increase in *VSG* switching rate in the *H3.V/H4.V* double deletion mutants. Using long-read sequencing technology, we found that *VSG* switching occurred *via* recombination and concluded, that the concomitant increase in spatial proximity and accessibility among expression sites facilitated the recombination event.
I therefore identified the histone variants H3.V and H4.V to act at the interface of global nuclear architecture and chromatin accessibility and to represent a link between genome architecture and antigenic variation.

Zusammenfassung

Antigenvariation ist ein weit verbreiteter Mechanismus der Immunevasion von Pathogenen [63]. Sie beruht auf der transkriptionellen Selektion eines einzelnen Gens aus einer hypervariablen Multi-Gen Familie und dem wiederholten, systematischen Wechsel zwischen der Expression verschiedener Gene dieser Familie. In vielen Organismen sind diese Gene als Kontingenzgene in den Subtelomeren angeordnet, wo sind einerseits transkriptionell reprimiert werden, andererseits erhöhter Mutagenese und Rekombination unterliegen [16].
Monoallelische Exklusion eines Gens und die damit einhergehende Eviktion aus seinem reprimierten genomischen Umfeld beruht auf unterschiedlichen molekularen Mechanismen. Sie ist, zum Beispiel, das Resultat einer Rekombination des betreffenden Gens in einen dedizierten, transkriptionell permissiven Lokus oder wird durch epigenetische, bzw. räumliche Umstrukturierung des entsprechenden Gens oder zugrunde liegenden Chromatins erreicht.
Beide Prozesse sind letztendlich durch die Architektur des Genoms beeinflusst. Architekturelle Proteine, die ebenfalls Antigenvariation kontrollieren, sind in vielen Pathogenen unbekannt.

Der parasitäre Protozoe *Trypanosoma brucei* entkommt einer Elimination durch die Immunabwehr seines Wirtes durch den periodischen Wechsel in der Expression eines von fast 3000 variablen Oberflächenglykoproteinen (VSGs). *VSG*-Gene umfassen die größte, monoallelisch exprimierte Genfamilie, die bislang beschrieben wurde. Um exprimiert zu werden, muss das selektierte *VSG* Gen in eine Expressionsseite transloziert sein. Diese wiederum wird in einem dedizierten Kompartment des Zellkerns, dem Expressionsseiten-Zellkernkörper (ESB), transkribiert. Obgleich diese Gegebenheiten die zentrale Rolle der Zellkernarchitektur in der Antigenvariation in *T. brucei* verdeutlichen, so ist wenig über die ihr zugrundeliegenden Mechanismen bekannt.
Um ein umfassendes Bild der Zellkernarchitektur in Trypanosomen zu bekommen, habe ich in der hier vorliegenden Doktorarbeit Hi-C, eine Methode zur Feststellung chromosomaler Konformationen, in *T. brucei* Blutstromform und Prozyklen etabliert und angewendet. Die Applikation dieser Technik offenbarte einen hoch strukturierten Zellkern: Chromosome sind territorial angeordnet und gehen spezifische Interaktionen in *trans* untereinander ein. Dies sind beispielsweise Interaktionen zwischen Zentromeren, Genen für ribosomale RNA und Prozyklinen unterschiedlicher Chromosomen. Auch Interaktionen, die in funktionellem Zusammenhang mit Antigenvariation stehen, wurden gefunden. Dabei handelte es sich zum Einen um strukturelle Verdichtungen des subtelomerischen Chromatins transkriptionell reprimierter *VSG* Gene und zum Anderen um erhöhte Interaktionen zwischen reprimierten *VSG*-Expressionsseiten. Hi-C bestätigte außerdem die räumliche Separation der aktiv transkribierten Expressionsseite von den übrigen, stillen *VSG*-Expressionsseiten.
Des Weiteren suchte ich nach Proteinen, die in der Aufrechterhaltung der Zellkernarchitektur in *T. brucei* wirken. Anders als CTCF ist Cohesin nicht auf Metazoen beschränkt. Ich fand Cohesin über den gesamten Zellzyklus exprimiert, was eine architekturelle Rolle des Proteinkomplexes zuzüglich der Schwesterchromatidkohäsion suggerierte. Mittels Chromatin-Immunpräzipitation konnte ich feststellen, dass Cohesin mit den Histonvarianten H3.V und H4.V an vielen Stellen des Genoms kolokalisierte, insbesondere über dem *VSG* Gen der aktiven und reprimierten Expressionsseiten. Während eine Deletion von *H3.V* zu erhöhten Interaktionsfrequenzen zwischen Expressionsseiten führte, resultierte eine gleichzeitige Deletion von *H3.V* und *H4.V* zu einer additiven Öffnung des Chromatins an Expressionsseiten. RNA Sequenzierungen ergaben, dass in der *H3.V/H4.V* Doppeldeletionsmutante die Transcription von *VSG* Genen erhöht war, was auf einen funktionellen Verlust der monoallelischen Expression hindeutete.
Immunfluoreszenzaufnahmen der VSGs auf der Zelloberfläche, Durchflusszytometrie und RNA Sequenzierung einzelner Zellen zeigten einen fortschreitenden Verlust der Expression von *VSG-2*, was auf einen erhöhten Wechsel der *VSG*-Expression auf dem Einzelzellle-

vel hindeutete. Durch die Sequenzierung der genomischen DNA der *H3.V/H4.V* Doppeldeletionsmutante konnten wir feststellen, dass der primäre Mechanismus des Wechsels in der *VSG* Expression auf eine Rekombination zwischen Expressionsseiten zurückzuführen war. Diese Rekombination wurde vermutlich durch die gesteigerte räumliche Nähe und Öffnung des Chromatins der Expressionsseiten begünstigt. Zusammenfassend konnte ich feststellen, dass die Histonvarianten H3.V und H4.V auf der Schnittstelle zwischen globaler Zellkernarchitektur und lokaler Chromatinzugänglichkeit agieren und funktionell ein molekulares Verbindungsstück zwischen Genomarchitektur und Antigenvariation darstellen.

It is the pervading law of all things organic and inorganic,
of all things physical and metaphysical,
of all things human and all things superhuman,
of all true manifestations of the head,
of the heart,
of the soul,
that the life is recognizable in its expression,
that form ever follows function.
This is the law.

— Louis Sullivan, architect

Acknowledgements

Foremost, I would like to express my profound gratitude and great appreciation to my primary supervisor, Nicolai Siegel. I joined is laboratory for the first time as a Bachelor's student in 2012 and ever since, he's stood by my side as a true mentor, guided me to grow as a scientist and helped me through many scientific and personal decisions.
I would like to thank Konrad Förstner, not only for being on my committee board, but for spending endless hours and months with me setting up a pipeline for Hi-C data analysis. Working with you not only gave me huge insight into the world of bioinformatics, but also taught me 'the beauty of efficiency' and how to organize my data and work in that sense (literally starting from how to write dates correctly).
I sincerely would like to thank my thesis committee members Markus Engstler and Susanne Fenz, both for supporting me during the course of my thesis and beyond through the invaluable discussions and scientific input I gained in the biweekly TrypClub meetings.
I want to thank Emmanuel Saliba for letting me in on the secrets of single-cell RNAseq. Thank you for all the time and patience that it took – and also for being my 'franco-allemand language tandem' during that time.
Many thanks go to Julien Guizetti, who has not only started to teach me FISH already in 2012, but also continued to support me greatly in the course of this project when he was part of the Siegel lab in 2017.
I want to thank the members of the Janzen lab and especially Christian Janzen himself for his constant support and for sharing his ideas and opinions with me in the many joint lab meetings we had.
Many thanks go to all the current and previous members of the Siegel lab. Raúl, I have been so lucky that you (Wicked Smaht) decided to join the lab and this project. You have given our project a new dimension and it has been a great pleasure and an honor to fight by your side through the revision of our paper. This would have been impossible

without you and I am grateful for all the times you lifted me back up. I am happy to have you not only as a colleague but as a true friend.
Ramona, I still remember the day we first met when you were busy setting up the Siegel lab in Würzburg. Thank you for all your support and for always being a trusted friend, that I knew I could come to anytime.
My 'PhD peers' in the lab, Caro and Amelie, I sincerely want thank you for sharing the ups and downs of a PhD life with me. Amelie, I want to thank you especially for the effort you made during the organization of our lab move and beyond. Caro, I particularly want to thank you for 'jumping in' with the ATAC-seq on short notice, when you were busy yourself with wrapping up your PhD.
Vanessa, Benedikt and Ines, I want to thank you for the short but great time we've had and for all your support, when times were rough. Vanessa, thank you for sharing your enthusiasm and ideas in the manifold discussions we've had about our projects.
Charly, when I moved into your shared flat in 2009, I had no idea we would stick together throughout our biomedical studies and PhDs. I am grateful for our friendship and for all the guidance you gave me, especially when I got started in Würzburg.
Many thanks go to the colleagues from the IMIB, the GSLS and in particular the CoreUnit SysMed, who gave us great and quick support even after the lab had moved. I am also grateful to my friends and colleagues from the parasitology community and the BMC in Munich, Andreas Ladurner and Anton Eberharter, who gave us a warm welcome and provided a great infrastructure that made a smooth transferal and continuation of our projects possible.
I would like to thank the Dekker laboratory, Noam Kaplan and Maxime Imakaev for their advice during the setup of the Hi-C experiments and later in the data analysis and preparation of the manuscript.
I want to thank Christiane Ranke, for her precious friendship, her advice and encouragement even across the ocean. I am very fortunate to have you.
From all my heart, I want to thank my family, my sister Philine and my parents Ulli and Carola. You taught me to follow what moves my heart and my mind, to have faith and to also stick through stringent times. I am grateful for all I've learned from you up to now, for all the doors you've opened for me and for your unconditional support that I can always be certain of.
Finally, I want to thank my husband Kim, who went on this journey together with me and who has been by my side every day. I cannot put into words what your support, love, patience and care mean to me.

Contributions

Parts of the methodologies, results and data achieved and presented in the prevailing doctoral thesis were previously published in the manuscript **'Genome organization and DNA accessibility control antigenic variation in trypanosomes'** Müller LSM, Cosentino RO *et al.*, *Nature* **563**, 121-125 (2018) under a Creative Commons Attribution 4.0 International License.
In this respect, I would like to acknowledge and thank the collaborators and colleagues, that contributed to the generation of data sets and materials presented here.

Raúl O Cosentino analyzed the RNAseq and scRNAseq data sets and determined the site of recombination between expression sites.
Konrad Förstner helped in the establishment of a Hi-C analysis pipeline and wrote the custom python scripts used for Hi-C data analysis.
Antoine-Emmanuel Saliba established single-cell RNAseq in *T. brucei* and prepared the *T. brucei* single-cell RNAseq libraries with the support of Panagiota Arampazi and Tanja Achmedov.
Fluorescence *in situ* hybridization experiments against centromeres and telomeres in *T. brucei* were done by Julien Guizetti.
Electron microscopy images were taken with the support of Elisabeth Meyer-Natus.
Amelie J Kraus generated the plasmid for Ty1-tagging of H3.V.
Carolin Wedel generated the ATACseq libraries.
Christian J Janzen performed the flow cytometry experiments to determine cell cycle profiles and *VSG*-switching rates of *T. brucei* cell lines.

Contents

List of Figures

List of Tables

Acronyms

ANC active nuclear compartment

ASF1A anti-silencing factor 1A

ATACseq assay for transposase-accessible chromatin using sequencing

ATRX alpha thalassemia mental retardation X-linked protein

BAC bacterial artificial chromosome

BB Binding Buffer

BDF3 bromodomain factor 3

BIR break-induced replication

BES bloodstream-form expression site

BLAST Basic Local Alignment Search Tool

CAF-1 chromatin assembly factor 1

CAF-1b chromatin assembly factor 1b

CDS coding DNA sequence

CENP-A histone H3-like centromeric protein A

ChIA-PET chromatin interaction analysis by paired-end tag sequencing

ChIP chromatin immunoprecipitation

ChIPseq chromatin immunoprecipitation with sequencing

CID chromosomal interaction domain

cLAD constitutive lamina-associated domain

CTCF CCCTC-binding factor

CT chromosome territory

CT-IC chromosome territory - interchromatin compartment

ChIPseq Chromatin-Immunoprecipitation with sequencing

DamID DNA adenine methyltransferase identification

DAXX death-associated protein 6

DDD distance-dependent decay

D-loop displacement-loop

DLS domain-like structure

DOT1B disruptor of telomeric silencing-1B

DSB double-strand break

ER endoplasmatic reticulum

ESAG expression-site-associated-gene

ESB expression site body

FACT facilitates chromatin transactions

FECA first eukaryotic common ancestor

FISH fluorescence *in situ* hybridization

fLAD facultative lamina-associated domain

GC gene conversion

GFP green fluorescent protein

HAT Human African Trypanosomiasis

Hi-C genome-wide chromosome conformation capture

HIRA histone regulatory homolog A

HR homologous recombination

H3.V H3 variant

H4.V H4 variant

IC interchromatin compartment

ICE iterative correction and eigenvector decomposition

IF immuno-fluorescence

IGV Integrative Genomics Viewer

INC inactive nuclear compartment

ISWI imitation switch remodeling complex

LAD lamina-associated domain

LBR lamin B receptor

LECA last eukaryotic common ancestor

MC megabase chromosome

NHEJ non-homologous end-joining

MNase micrococcal nuclease

MNase-ChIPseq micrococcal nuclease chromatin immunoprecipitation with sequencing

mRNA messenger ribonucleic acid

mRNP messenger ribonucleoprotein

N50 size of the shortest contig at 50 % of the overall genome length

NAD nucleolus associated domain

NGS next generation sequencing

NOR nucleolar organizer region

NUP-1 nuclear pore complex protein-1

OLC overlay-layout-consensus

OR olfactory receptor

PAG procyclin-associated gene

PARP procyclic acidic repetitive protein

PCR polymerase chain reaction

PTM posttranslational modification

PTU polycistronic transcription unit

PFGE pulsed-field gel electrophoresis

Rad21 Double-strand-break repair protein

RNAseq ribonucleic acid sequencing

rRNA ribosomal ribonucleic acid

SDSA synthesis-dependent strand annealing

SMRT single molecule real-time

SMRTseq single molecule real-time sequencing

Smc1 Structural maintenance of chromosomes protein 1

Smc3 Structural maintenance of chromosomes protein 3

SL spliced-leader

Scc1 Sister-chromatid cohesion protein 1

Scc2 Sister-chromatid cohesion protein 2

Scc3 Sister-chromatid cohesion protein 3

Scc4 Sister-chromatid cohesion protein 4

scRNAseq single-cell ribonucleic acid sequencing

SNP single-nucleotide polymorphism

SSU small subunit

TAD topologically associating domain

TAR transformation-associated recombination

TLE Tris low EDTA

tRNA transfer ribonucleic acid

TSS transcription start site

TTS transcription termination site

TWB Tween wash buffer

VEX-1 VSG exclusion-1

VSG variant surface glycoprotein

3C chromosome conformation capture

4C chromosome conformation capture-on-chip

5C chromosome conformation capture carbon copy

Part I

Introduction

1 Chromosome conformation capture

Packaging of the human DNA inside the confinements of the nucleus requires a ~200,000-fold condensation of the genome [60]. Thereby, the genome is not folded at random, but its shape strongly relates to its functioning, thus representing the top level of the epigenome.
Studying nuclear organization was traditionally limited to the resolution of light and electron microscopy imaging and the analysis of single cells. The development of the chromosome conformation capture (3C) technology by Job Dekker [65] translated for the first time DNA physical contacts observed in individual cells into probabilities by systematically studying chromatin contacts across a great population of cells. Soon its derivative techniques entered the *'-omics'* field of genome-wide NGS -based analyses, which allowed comprehensive and more detailed studies of nuclear architecture at high resolution. This chapter shall give an overview on the principles of 3C-based methodologies with their applications, limitations and downstream computational analyses.

1.1 3C-based methodologies

Chromosome conformation capture approaches are based on inferring DNA folding principles by measuring the frequency of ligation events between different sites of the genome (see Figure 1.1). Therefore, the chromatin fold is captured by fixation of the DNA inside the nucleus, which is followed by a restriction digest or shearing of the DNA by sonication. Subsequently, the digested or sheared DNA is ligated back together. Ligation is often performed under dilute conditions to favor intra-molecular interactions over inter-molecular ones. Subsequent reverse cross-linking and fragmentation creates a library of interacting DNA fragments, connected by a ligation junction [65].
In a classical 3C experiment [65], relative interaction frequencies between different loci of interest are determined by (semi)quantitative polymerase chain reaction (PCR) amplification using primer combinations that bind to a ligation junction of choice. As the ligation frequency of two respective fragments is determined by the spatial proximity between them, most ligation events occur between genomic loci located in *cis*. However, rare interactions between DNA elements located in *trans* (e.g. spatial clustering of regulatory sequences) are also captured by 3C (Figure 1.1, bottom left).
A major improvement in contact map resolution was achieved with the development of chromosome conformation capture-on-chip (4C)

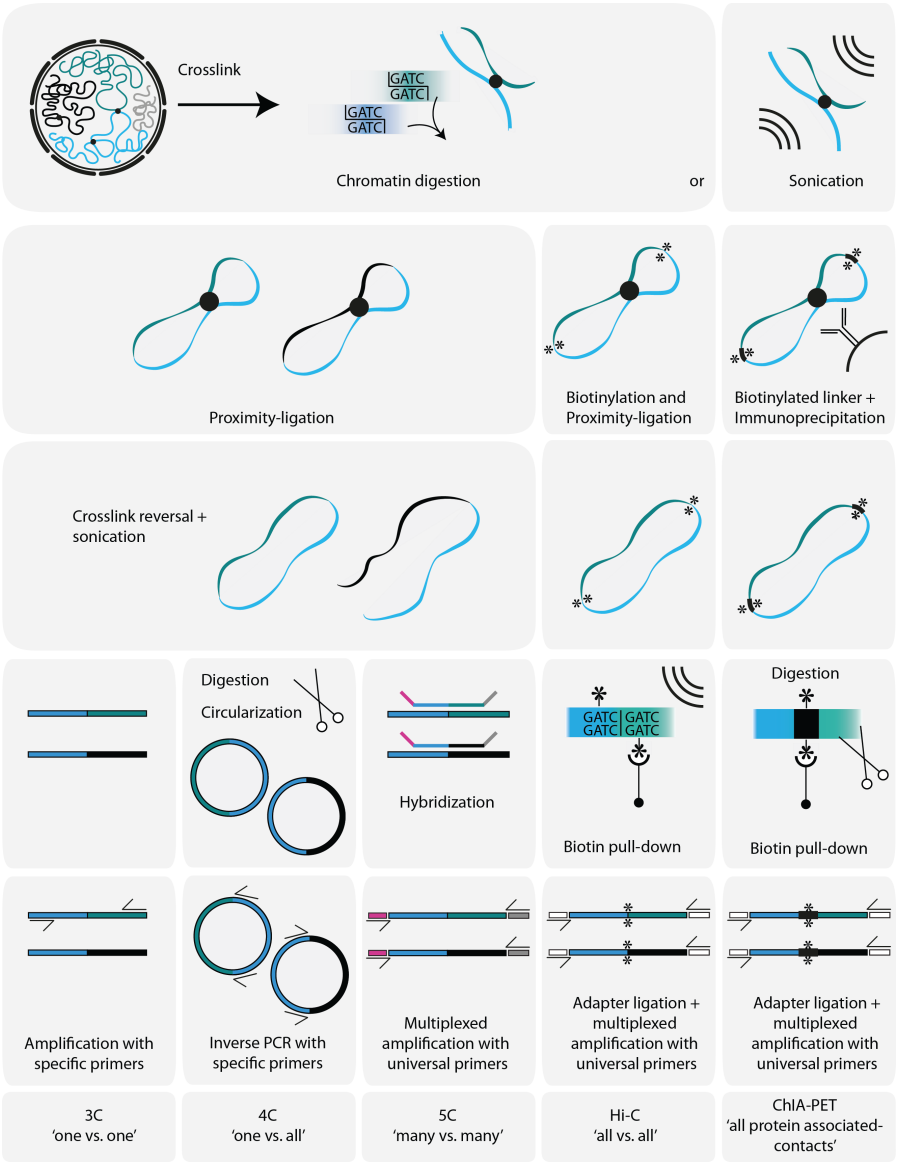
[248] [290], which probes the interactions of a distinct locus of interest (the 'bait') against a plethora of genomic loci or – ideally – the whole genome (Figure 1.1, second panel to the bottom left). In the process of 4C, the ligation products become circularized. Using inverse PCR with primers binding in the bait sequence, the corresponding interacting fragment is amplified and read out either by microarray [248] [290] or by NGS [249]. Whereas 4C is the approach of choice to investigate the interaction profile of a single locus across the genome, it is limited at giving information about local folding behavior [60].
In chromosome conformation capture carbon copy (5C) [71] the ligation product is used as a template for pairwise hybridization of oligonucleotides that are subsequently ligated together. The oligonucleotides carry a 5′ universal adapter that allows for amplification in a multiplexed PCR. Although limited in resolution, 5C has the advantage to circumvent PCR amplification bias. Also, 5C for the first time generated a contact matrix between a multitude of genomic loci simultaneously (Figure 1.1, bottom middle).
The first method to capture interactions between all possible loci of a genome was the Hi-C method developed by Lieberman-Aiden [156]. Hi-C is characterized by biotinylated ligation junctions, which are introduced during a blunting reaction prior to the ligation step. After shearing of the re-ligated DNA, the sample is enriched for fragments containing a valid ligation junction through a biotin pull-down with streptavidin-coated magnetic beads. An NGS library is created by adapter ligation to both ends of the Hi-C fragments. Each read pair resulting from paired-end sequencing is mapped back to the genome and represents a DNA-DNA contact (Figure 1.1, second panel to the bottom right). The resolution of this method is mostly determined by sequencing depth and therefore indirectly by genome size. Resolution is further limited by the frequency and distribution heterogeneity of the restriction site used. The recent usage of nucleases instead of a restriction digest allows the construction of interaction maps at nucleosome resolution (e.g. Micro-C [121] or DNAse Hi-C [168]).
Chromatin interaction analysis by paired-end tag sequencing ChIA-PET includes a chromatin immunoprecipitation (ChIP) step which is used to enrich for interactions mediated by a specific protein acting on DNA (Figure 1.1, bottom right) [89]. This limits this method to identifying solely interactions driven by the same protein and further is not compatible with protein knock down and -out experiments [60].

1.2 The Hi-C workflow

1.2.1 *Experimental considerations for Hi-C*

The careful experimental design of a Hi-C experiment is fundamental for the sequencing output and requires adjustments depending on

Figure 1.1: Chromosome conformation capture approaches



Scheme showing the experimental similarities and differences between 3C (first panel), 4C (second panel), 5C (third panel), Hi-C (fourth panel) and ChIA-PET (last panel) approaches.

cell-type and genome size.
Both, the analyses of large genomes and looping structures at the fine-scale require sufficient resolution. Due to the quadratic nature of Hi-C data, an increase in resolution by the factor X can only be achieved through an X^2 increase in sequencing depth. Deep sequencing, in turn, demands a sufficiently complex library. Complexity is determined by the number of unique chimeric DNA molecules and is therefore limiting for the amount of informative reads that can be obtained from a respective library [143].
Library complexity can be increased at several levels in the experimental Hi-C workflow. First, a sufficient amount of starting material is critical for library quality. Recent publications emphasize to enhance complexity by combining multiple libraries [222] [22], which allows to reduce library amplification steps.
As for all 3C-based experiments, the choice of the restriction enzyme primarily determines resolution. Assuming the statistical occurrence of a 6-bp restriction site every $4^6 = 4096$ nucleotides in the genome and every $4^4 = 256$ nucleotides for a 4-bp restriction site, resolution and library complexity improves by $\sqrt{16} = 4$-fold with the usage of a 4-bp cutter compared to the usage of a 6-bp cutter.
In the original Hi-C protocol, nuclei were lysed after the cross-linking step and the chromatin was subsequently digested and re-ligated in solution [156], yielding a resolution of ~1 Mb. Lately, *in situ* Hi-C, where digestion and ligation occur inside the permeabilized nucleus, were used [222]. This approach yielded more 'meaningful' contacts and thus denser interaction matrices with a resolution of up to 1 kb for the human genome.
The analysis of highly repetitive genomes is especially demanding. Repetitive genomic sites are often sparse in restriction sites, are difficult to sequence and provide few unique stretches, where reads can map unambiguously. Better representation of repetitive genomes may therefore require longer sequencing reads which improves unique alignment of the interacting fragments.

1.2.2 *Computational processing and analysis of Hi-C sequencing data*

1.2.2.1 *Mapping and filtering of Hi-C reads*

Paired-end sequencing of a Hi-C library yields mate pairs of reads stored in FASTQ format. Importantly, during the mapping step, the two mate reads are mapped independently from each other as they are not meant to represent two ends of a continuous fragment. Secondly, the reads need to be mapped uniquely to the genome to unambiguously determine the interaction between two loci. Any locus in the genome that is not unique is effectively 'invisible' in an Hi-C experiment. A read pair is removed even if one of the mates does not map uniquely to the genome ('singletons'). Any read pair representing an

exact sequence copy of another pair of reads ('duplicates') are not considered since they are likely the result of PCR over-amplification. In some of the reads, sequencing goes through the ligation junction. Due to their chimeric nature, these reads cannot be mapped in their full length to the genome. To discriminate between the two fragments of the chimeric molecule, some authors propose an iterative mapping process, in which the out-most 25 bp of each read are mapped uniquely. If unique mapping is not possible, the query is further extended stepwise [125]. In alternative approaches, all unmapped read fragments are searched for occurrence of the ligation junction motif in the sequence, truncated accordingly at these sites and are then re-aligned to the genome [240].
Due to the technical procedure of the Hi-C experiment, the digestion and re-ligation at restriction sites evokes, that the majority of reads can effectively be assigned to a restriction site in the genome. If the two mates of a read pair are located inside the same restriction fragment of the genome, these reads are usually uninformative as they represent a fragment that was not ligated (dangling end) or ligated to itself (self-circle) [23] and are therefore disregarded. A valid DNA-DNA interaction is represented by a chimeric DNA molecule, consisting of two genomic fragments separated by the motif of a filled-in and re-ligated restriction site. Importantly, the two joined DNA sequences origin from different restriction fragments in the genome.
To improve signal-to-noise ratio, the genome is usually binned into greater, regularly spaced genomic fragments ('bins') between 1 kb [222] 1 Mb [156] in size. In this way, interaction signal originating from different restriction sites is 'collected' in a given genomic bin, which strengthens meaningful interactions and reduces noise. Bins with overall low contact signal (e.g. repetitive regions) are usually removed from the matrix at this point, since they are difficult to interpret. To obtain meaningful matrices, the bin size should be chosen such that 80% of bins retrieve a minimum of 1,000 contacts [222]. For a flowchart listing the different steps of the data analysis, see Figure 5.1 of the results section.

1.2.2.2 *Hi-C data correction*

In an Hi-C experiment, contact probability between two loci correlates with the enrichment of the corresponding fragments in the sequencing pool. The occurrence of a given fragment in the sequencing data is, however, systematically biased by different parameters such as sequencing efficiency, fragment length and GC content [286]. Instead of *explicitly* correcting for all possible bias, Imakaev and colleagues developed an *implicit* correction approach based on iterative correction, referred to as ICEing [125]. Assuming all genomic loci (or bins) undergo a comparable number of interactions, the interaction matrix is 'balanced' such that the interaction values of every column and row

of the matrix add up to the same sum. In this manner, every locus of the genome is given the same 'visibility', regardless of experimentally introduced biases.

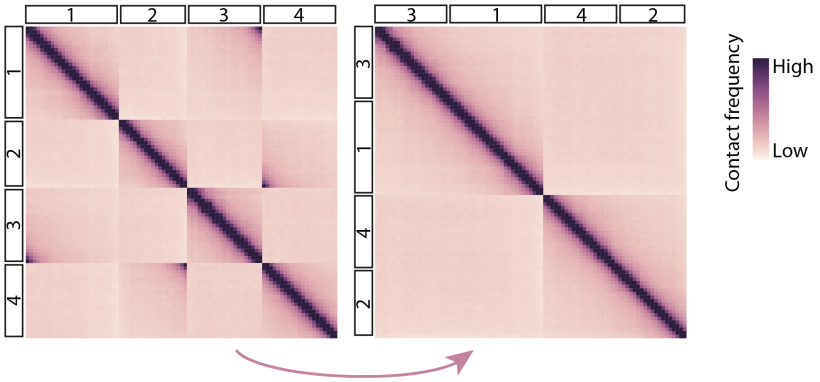
1.2.2.3 *Considerations for Hi-C data interpretation*

During the interpretation of Hi-C data, some biological and technical peculiarities inherent to the experimental setup need to be taken into account. First, 3C approaches typically capture the average interaction data across a great population of cells. Therefore, it is not possible to discern interactions, which are mutually exclusive at the level of single cells [143]. Further, one cannot discern, whether a weak Hi-C signal results from a widely distributed, but instable interaction that is short in time or represents a strong, mutual exclusive interaction that only occurs in a fraction of cells [143]. Recent efforts have been made to establish Hi-C also in single cells to study mutually exclusive regulatory interactions and cell-to-cell variability in genome conformation [195] [251].
Secondly, the frequency of interactions scales inversely with physical genomic distance between two loci (distance-dependent decay (DDD) of interaction frequencies). Therefore, when interrogating intra-chromosomal interactions, the linear distance between two respective loci needs to be taken into account (see 2.2.2). For inter-chromosomal interactions, indefinite distances are assumed.
As the DNA fold is of functional nature, folding patterns arising from Hi-C experiments can often not be generalized and need to be interpreted specifically with regards to species, cell-type and -condition. However, some interaction patterns repeatedly reoccurring are generally accepted to represent certain genomic folding behaviors. How these patterns define and are represented in Hi-C data will be explored further downstream in the according sections (see 2.2).

1.2.3 *Genome assemblies using Hi-C*

Besides the study of chromosome conformation, Hi-C interaction data has recently been exploited for scaffold validation or even *de novo* assembly of whole genomes [132].
Conventional genome assemblies rely on scaffolding short NGS reads through consensus overlap into greater contigs. Due to the repetitive nature of genomes, however, often large genomic regions cannot be bridged by short reads alone, thus leaving the assembly fragmented into many contigs [206].
Scaffolding of complex genomes has lately been greatly facilitated by combining long-read sequencing approaches with information on the three-dimensional genome structure as measured by Hi-C [132][40][21].
To scaffold sequence contigs relative to each other using Hi-C inter-

Figure 1.2: Triangulation of contigs



Hi-C heat maps before and after triangulation of contigs. Relative interaction frequencies are color-coded. **Left heat map** Hi-C contact data is mapped to exemplary unscaffolded contigs (numbered 1-4). High inter-contig interaction signal away from the heat map diagonal indicates that the respective contigs belong to the same chromosome. Their relative position to each other can be inferred from the decay of interaction frequencies along the length of the contigs (e.g. the 5′ end of contig 1 should be positioned downstream of the 3′ end of contig 3). **Right heat map** After scaffolding of contigs according to their relative interaction frequencies, the diagonal indicating intra-chromsomal interactions is continuous inside each chromosome. In this example, the four contigs were assembled into two separate chromosomes.

action data – also referred to as *DNA triangulation* [143] (see Figure 1.2) – it is taken advantage of two principles intrinsic to the three-dimensional structure of genomes. First, as contigs belonging to the same chromosome display significantly higher interaction frequencies among each other than contigs of different chromosomes, contigs can be grouped into chromosomes (see 2.2.3). Second, the DDD of interaction frequencies between contigs belonging to the same chromosome allows to position and orient them relative to each other [132]. If enough sequence variation (e.g. in form of SNPs) exists between two homologous chromosomes, DNA triangulation enables to phase the homologous chromosomes of diploid organisms resulting in haplotype-specific assemblies [239].

2

Evolution and structure of the nuclear compartment

2.1 Eukaryogenesis – the emergence of nuclear compartmentalization

The emergence of eukaryotes is dated back 1.7 billion years ago, presumably with both archaeal and proteobacterial ancestry [57]. Multiple events building on one another were involved in the process of eukaryogenesis and still a multitude of hypotheses regarding the emergence of the nuclear compartment are controversially discussed today [57]. The three-domains of life hypothesis proposes the Eukaryota to have evolved independently from a common ancestor shared with Archaea [96]. Adversely, recent lines of evidence place the emergence of the first eukaryotic common ancestor (FECA) within the archaeal lineage (two-domain hypothesis) [167] [288].
In both scenarios, these protoeukaryotes acquired genes of bacterial decent most likely through multiple events of horizontal gene transfer or endosymbiosis, leading to the development of primordial organelles present in the last eukaryotic common ancestor (LECA) of all present-day eukaryotes [57]. Still, the evolutionary origin of the nuclear envelope compartmentalizing the genome from the cytoplasm is unclear today (reviewed in [176]). Cavalier-Smith proposed that the nuclear envelope derived, like the endoplasmatic reticulum (ER), from autonomous invagination of the protoeukaryotic plasma membrane [43]. The first to suggest an endosymbiotic event as the origin of eukaryotic organelles was Mereschkowsky [182], which gave rise to many endosymbionic theories concerning the origin of the nucleus (reviewed in [176]). A strict endosymbiotic origin of the nucleus has, however, been challenged since the nuclear compartment itself does not mirror the physiology of any free-living unicellular organism [173]. To account for the similarities in i) genome organization between Archaea and Eukarya [177] and ii) lipid composition between Bacteria and Eukarya, syntrophic hypotheses as proposed by Martin [175] and Moreira [191] included continuous gene-transfer between the genomes of a bacterial symbiont and an archeal host, eventually leading to the evolution of the protoeukaryote genome. Martin and Koonin [174] proposed that the introduction of α-proteobacterial group II introns into the host DNA by seminal gene transfer made a separation of RNA splicing from translation necessary, which eventually drove the development of a nuclear envelope. With steadily increasing genome size

and organismal complexity, precise higher-order chromatin structures balancing genome compaction and accessibility evolved [177].

2.2 Organization of the eukaryotic nucleus

2.2.1 *The chromatin fiber*

Eukaryotic genomes are organized into nucleoprotein complexes, referred to as chromatin. At the basic level of folding, the DNA is bound by DNA-binding proteins called histones and packaged into nucleosomes [140] consisting of ~146 bp of DNA wrapped in 1.67 left-handed turns around an histone octamer [165]. The histone octamer is comprised of two H3-H4 and H2A-H2B core histone dimers. This 'particles-on-a-string' arrangement of nucleosomes [208] spaced by 40-60 bp of linker DNA associated with the linker histone H1 [231] is also referred to as the 10 nm fiber, that represents the first level of DNA compaction. The degree of DNA condensation can be altered through post-translational modification of histone N-termini as well as through the introduction of histone variant forms into the nucleosome (see 2.2.8.2).
The 10 nm fiber has been proposed to fold into a secondary structure, the 30 nm fiber [258] [234] [142], presumably through interactions between histones of different nucleosomes (e.g. interaction between H1 linker histones [258] or histone H4 tail interacting with the acidic patch of histone H2A [131]). The existence of the 30 nm fiber *in vivo* has, however, been challenged (reviewed in [260]).
Taking into account local heterogeneity in nucleosome charge, occupancy and fiber flexibility, the primary chromatin conformation is a heteropolymer [185] that is subject to spatial confinement by the lamina, as well as to bridging and compaction by structural proteins (see 2.2.8). Models on how the DNA heteropolymer forms higher-order structures inside the nucleus shall be discussed in the following.

2.2.2 *Models of topological constraints on the spatial organization of chromosomes*

The DNA folding behavior inside the nucleus is heavily determined by the presence or absence of topological constraints [126]. These forces drive the formation of a distinct polymer state of the DNA and thus determine the dynamic properties of chromosomes. In the past, several models predicting the folding behavior of the DNA polymer have been developed. Due to the range of cell-to-cell variability in the formation of chromatin even in cells of the same tissue, recent efforts went towards a general biophysical modeling of chromatin conformation [185]. Chromosome conformation capture (3C) approaches provide an averaged DNA folding profile of a cell population and are

thus suitable to infer general principles of chromatin architecture [156] [185]. Typically, the interaction probability P_c between two loci on the same chromosome decays with increasing genomic distance s between them (DDD). Calculating the power law by which contact probability P_c decays in relation to genomic distance s can serve as a proxy for the degree of chromosome condensation [156]:

$$P_c(s) = s^x.$$

A smaller scaling exponent x implies a faster decay over a given genomic distance, thus pointing towards a less compacted DNA polymer [156]. In the following, two DNA folding models, the *equilibrium* and *fractal globule* [99] model shall be explained in brief.

2.2.2.1 *The equilibrium model*

The equilibrium state of a DNA polymer (Figure 2.1 a) describes a dense, though relaxed, freely diffusing polymer solely confined by the barriers imposed by the globule, rather than by constraints intrinsic to the polymer itself. The DNA is highly knotted and concatenated throughout the globule and its density is constant (at equilibrium) [185]. Chromosomes do not occupy a distinct domain and due to their high degree of entanglement, chromosomal folding and unfolding is a laborious and thus infrequently observed process [137]. The contact probability of two loci on a subchain of the polymer scales with increasing genomic distance as

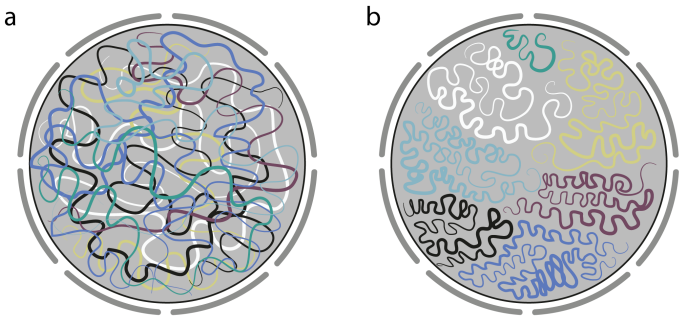
$P_c(s) = s^{-3/2}$, followed by a plateau at large s [164][185].

Such folding behavior has for example been described for the chromosomes of *Saccharomyzes cerevisiae* [74] [88], which exhibit a so-called *Rabl organization* of chromosomes, where centromeres and telomeres are juxtaposed in the nuclear periphery [253].

2.2.2.2 *The fractal globule model*

The concept of the fractal globule was first introduced by Grosberg and colleagues in 1988 [99] as a 'crumpled globule'. In a fractal configuration, the DNA is not at equilibrium, but constrained by interactions and confinements imposed by the polymer itself or by DNA-binding proteins [185]. The DNA folds hierarchically in a fractal-like fashion, ultimately resulting in chromosome territories, lacking any knots or entanglement and ideally not crossing itself (Figure 2.1 b). Due to its unknotted nature, the chromatin of the *fractal globule* can – in contrast to the *equilibrium globule* – easily and rapidly (de)condense [185]. This folding behavior predicts a power law scaling of:

Figure 2.1: Models of DNA folding behavior



Schemes displaying the DNA folding behavior **a** at equilibrium and **b** as a fractal of globules.

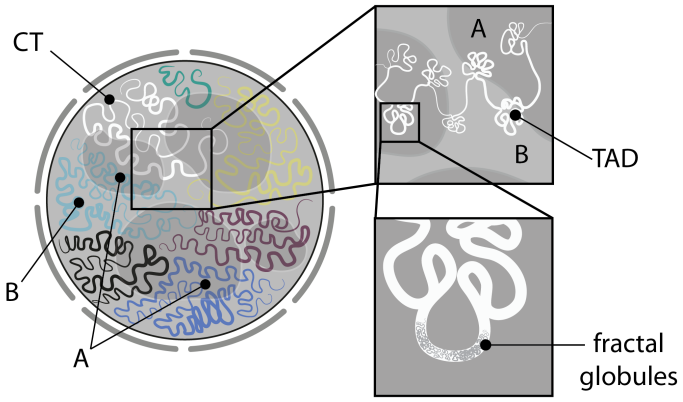
$P_c(s) = s^{-1}$ for s ranging from 0.1 to 10 Mb [185] [99].

In comparison to the equilibrium globule, the fractal globule is more compact, thus resulting in a slower decay of contact probability over distance. The scaling of s^{-1} further does not claim for a distinct length of chromatin loops and the fractal nature allows for topological segregation ('genomic territories') on all levels of folding [185]. These predictions were met in recent Hi-C experiments, including human cells [156] and *Drosophila* [241] and unicellular eukaryotes such as *Plasmodium falciparum* [13].
Based on the fractal globule module, several derivative (and partly species-specific) models have been developed (e.g. loop extrusion model [232], see 2.2.8.1).

2.2.3 *Chromosome territories*

The presence of fractal globules at the sub-chromosomal level may imply similar folding principles at the chromosomal level, ultimately resulting in the segregation into chromosome territories (CTs) [185] (see Figure 2.1 b and Figure 2.2). The occupancy of a certain nuclear domain by each chromosome entails that even distant loci of the same chromosome tend to be in spatial proximity. Following lines of evidence derived from nuclear micro-irradiation [292] and chromosome painting by fluorescence *in situ* hybridization experiments [32], the existence of CTs was also supported by chromosome conformation capture: In an Hi-C experiment, CTs are characterized by higher interaction frequencies inside a respective chromosome (*cis*- or intra-chromosomal interactions) than across chromosomes (*trans*- or inter-chromosomal interactions) [156].
A territorial arrangement of chromosomes during interphase has been found in all eukaryotes studied to date [51]. Besides the characteri-

Figure 2.2: Fractal globules fold into higher-order organizational domains



Chromosomes occupy distinct chromosome territories inside the nucleus. Each chromosome fiber is organized into units of topologically associated domains (TADs, **upper box**) that meander between active (dark patches, named 'A') and inactive ('B') domains inside the nuclear sphere. TAD formation results from the hierarchical folding of fractal globules (**lower box**).

zation of chromosome territories, the space arising between chromosome territories has been acknowledged as a functional entity, the *interchromatin compartment* (IC), which led to the development of the chromosome territory - interchromatin compartment (CT-IC) model [52].
The interchromatin compartment trenches beneath the chromatin domains like cisternae, possibly providing efficient routes e.g. for messenger ribonucleoprotein (mRNP) transport towards the nuclear pores in the periphery and – *vice versa* – facilitating noncoding RNAs and regulatory proteins (such as transcription factors) to reach their genomic destination [54]. Fakan and van Driel [81] propose decondensed 'perichromatin' at the interface between the IC and the CT which represents the subnuclear fraction of active chromatin that is transcribed and can evict from the CT and interact in *cis* or *trans* with other genomic regions inside the IC, e.g. by forming transcription factories or splicing speckles [85]. In the past, a substantial degree of chromosome intermingling between CTs has been shown [33], which appeared dependent on the presence of RNA Polymerase II [274] and was affected by blockage of transcription [33], indicating that these associations are at least in part dependent on transcription.

2.2.4 *Active and inactive nuclear compartments*

Differential nuclear positioning of genomic loci depending on their transcriptional status and gene density has been described in numerous cases, often involving eviction from the respective CT of a gene (reviewed in [26]). For example, induced expression of genes in the

Hoxb cluster resulted in the accumulation of active chromatin marks at those genes, evoked chromatin decondensation at these sites and looping out of the respective CT [44]. The authors proposed a model, in which alterations of chromatin structure (modifications and spatial positioning) precedes transcriptional activation.
In fetal liver cells, the active β-globin gene is – when activated – in association with other active genes, but remains in a silent cluster in tissues where it is not expressed [248]. This behavior was described with the help of 4C, which allowed Simonis and colleagues to determine genome wide interactions of the β-globin locus in different tissues. The observation of preferential interactions in *cis* and in *trans* between active and inactive genes, respectively, led to the hypothesis that the genome may fold into distinct domains of inactive and active chromatin [248].
The genomic partitioning into active and inactive compartments, typically referred to as A and B compartments (see Figure Figure 2.2), was also recapitulated in Hi-C experiments, where genomic blocks of associated DNA built a 'plaid pattern' across the interaction heat map and loci belonging to the same compartment revealed correlated interaction patterns across the genome [156] [222]. Whereas pairs of loci belonging to the active 'A' compartment showed lower interaction frequencies among each other, interactions inside the inactive 'B' compartment revealed higher interaction frequencies, pointing to a higher degree of compaction [156]. In line with this, the 'A' compartment was characterized by highly expressed genes, accessible chromatin and the presence of active chromatin marks (e.g. H3K26me3) in comparison to the 'B' compartment. A more recent study subdivided the two main compartments further (A1, A2 and B1-B4), taking into account gene density, transcriptional state, genomic content, chromatin marks, replication timing and association with nuclear structures such as the lamina [222].
To account for the finding of active and inactive chromatin compartments, the ANC-INC model [53] was developed as a refined derivative of the CT-IC model. The inactive nuclear compartment (INC) would classically describe the fraction of facultative heterochromatin. Chromatin belonging to the INC is highly condensed, transcriptionally inactive, typically rich in repressive chromatin marks [123] and located to the interior part of a chromatin domain with reduced accessibility. The active nuclear compartment (ANC), in contrast, circumscribes both the interchromatin compartment and peripheral chromatin that is more accessible and transcriptionally active [53]. The two compartments are, however, dynamically intertwined and movement of genes between the ANC and INC depending on their transcriptional status has been described (reviewed in [53]).

2.2.5 *Non-random radial arrangement of chromosomes and individual genes*

In human cells, small, gene-dense chromosomes have been observed to locate preferentially in the nuclear center, whereas larger and gene-poor chromosomes are rather positioned peripherally [156] [32]. Chromosomes are thus arranged in a radial gene-density gradient [51], which appears to establish during cell differentiation and gene activation [139]. Likewise, replication-timing has been found to be correlated with a radial arrangement of early replicating genes in the nuclear interior and late replicating genes in association with the nuclear periphery [98].
In some cases, the non-random radial arrangement has also been described to change for individual genes depending on the transcriptional state. A repositioning to the nuclear center has, for example, been observed for the *β-globin* locus during maturation of erythroid precursor cells [221] and the *Mash1* locus after neural commitment in stem-cells [279].

2.2.5.1 *Lamina-associated domains*

The constitutive association of heterochromatin with the nuclear periphery and the enrichment of active euchromatin in the nuclear center is a very generalized and long-standing view of the nucleus [141], which has been found relatively conserved across the animal kingdom [51].
Repressed chromatin domains that are in association with the nuclear lamina have been determined using a genome-wide DNA adenine methyltransferase identification (DamID) approach [215] and were termed lamina-associated domains (LADs) [101]. Typically, one-third of the genome is engaged into ~1,000-1,500 LADs of varying size (10 kb – 10 Mb) (reviewed in [267]). Identified in flies [215], *C. elegans* [124] and mammalian cells [101] LADs appear a metazoan characteristic.
LADs can consist of constitutive (cLADs), including pericentromeric and telomeric heterochromatin) or facultative heterochromatin (fLADs), respectively, and are enriched in the repressive chromatin marks di- and trimethylated H3K9 and trimethylated H3K27 [101]. Hi-C experiments revealed LADs to be part of the nuclear B compartment [135] as defined by Lieberman-Aiden and colleagues [156]. cLADs are greatly invariant and are supposed to anchor chromosomes to the nuclear lamina, thus providing structure to the nucleus (reviewed in [267]). As cLADs often comprise centromeric sequences, centromeres are often positioned as 'chromocenters' [91] in the nuclear periphery embedded among other LADs [101].
The mechanisms by which LADs are targeted to the nuclear lamina are multivalent [135] and not well understood today. Some lines of evidence point to the deposition of H3K9 and H3K27 methylation as an important players as knockdown of the corresponding methyltrans-

ferases causes reduced interactions between LADs and the nuclear lamina [107] [136]. Also, depletion of various lamin subtypes, the lamin B receptor (LBR) and emerin, a nuclear envelope transmembrane protein, resulted in LAD repositioning in fly, *C. elegans* and mammalian cells (reviewed in [267]).
LAD borders are sharply demarcated and are – in contrast to LAD-internal DNA – euchromatic and rich in active promoters and CCCTC-binding factor (CTCF) binding sites [101] (see 2.2.8), possibly restricting the spreading of heterochromatin marks [267].

2.2.6 *Topologically associating domains and their boundaries*

By means of Hi-C, a sub-compartmentalization of chromosome territories into megabase-sized 'topologically associating domains' (TADs) (see Figure Figure 2.2) has been identified in several mammalian species [69] [202] [271] and in Drosophila [241]. Mammalian TADs are variable in size and range between 10 kb and 2 Mb [64]. Generally, they are conserved with respect to position and seem largely invariant across cell-types and species [69] [202]. Recent work identified larger TADs to be sub-structured into smaller looping domains of 100-200 kb in size [222].
TAD boundaries seem to spatially isolate adjacent interaction domains [69] [202]. Typically, these self-interacting domains appear as sub-chromosomal triangles along the diagonal of an Hi-C interaction heatmap with sharp transitions among them. *Cis*-regulatory elements (e.g. enhancer elements) and their targets, have been described to preferentially be located inside the same TAD, thus favoring their specific interaction[166] [106]. The frequency of such intra-TAD interactions may be further altered through smaller chromatin loops that insulate or facilitate interactions [72]. Directing interaction preferences along the genome and prohibiting interactions between adjacent TADs, TAD boundaries display classical insulator activity. These boundaries are typically transcriptionally active and characterized by the segregation of chromatin marks, the presence of binding sites for CTCF, cohesin, transcription start sites, tRNA genes and retro-transposable elements [69].
The close association of CTCF and cohesin at TAD boundaries suggests that they are involved in the establishment of looping interactions among TAD boundaries [293] (loop extrusion [232]), which will be discussed in 2.2.8. In fact, depletion of cohesin has shown to affect TAD domain borders but had only mild effects on transcription [223]. Deletion of DNA sequences encompassing a TAD boundary resulted in a 'fusion' of the two adjacent TADs with increased ectopic interactions among them [202] [166]. However, hormone-induced transcriptional changes did not alter TAD boundaries but influenced intra-TAD structures that correlated with changes in transcription and epigenetic

marks [147]. These observations led to the idea of TADs being 'epigenetic domains' and regulatory units of transcriptional activity [147] with invariant boundaries but dynamic chromatin state. Dekker and colleagues [64] proposed that TADs alternate between the two main A- and B- compartments (see 2.2.4) depending on their respective chromatin structure and transcriptional activity.
In line with a mechanistic link between TAD formation and transcriptional control, TAD formation could not be observed in transcriptionally inactivated chromosomes [202]. Also, the presence of TADs appears restricted to interphase nuclei, as in mitotic chromosomes, any hierarchical compartmentalization such as into TADs is lost [197]. Just like LADs, TAD compartments appear to be restricted to metazoans [64]. In *C. elegans* autosomes, TAD structures were probably secondarily lost or weakened but are, however, present at *C. elegans* gonosomes [50]. Comparable domain-like structures observed in non-metazoans such as *S. pombe* or the apicomplexan parasite *P. falciparum* seem to have established independently [1] and are not based on comparable molecular mechanisms such as the deposition of CTCF (see 2.2.8). This is consistent with the observation that TAD formation seems to direct promoter-enhancer interactions, as lower eukaryotes (non-metazoans) mostly do not regulate transcription through *cis*-regulatory elements.

2.2.7 *Genome architecture compartmentalizes transcription*

In eukaryotes, transcription of ribosomal (rRNA), messenger (mRNA) and transfer RNA (tRNA) is performed by the three RNA polymerase multi-subunit complexes RNA Pol I, Pol II and Pol III, respectively [49]. Generation of these three types of transcripts happens to some extent in dedicated spatial compartments.
The most prominent nuclear compartment in eukaryotes is the nucleolus, an RNA Pol I-rich compartment that is exclusively dedicated to *rRNA* transcription. Already in bacteria, where there is only one RNA polymerase, *rRNA* containing operons cluster together with RNA Pol foci, thus forming a nucleolus-like structure [129]. In eukaryotes, nucleoli form around nucleolar organizer regions (NORs) containing *rDNA* repeats that are embedded in constitutive heterochromatin. Nucleoli fuse subsequently to form the mature nucleolus (reviewed in [179]), which is encapsulated by perinucleolar heterochromatin, so-called nucleolus-associated domains (NADs) [205] that overlap in parts with LADs [205]. NADs are enriched in repressive chromatin marks and show generally reduced transcription, which is, however, not exclusive since highly expressed *5S RNA* and *tRNA* genes are also part of the perinucleolar heterochromatin [205].
Spatial clustering of *rDNA* in the nucleolus (and associations between *tRNA* and *rRNA* genes) have been recapitulated in chromosome con-

formation capture experiments e.g. in yeast [74] and *Plasmodium* [149]. In *Plasmodium*, clustering of *rRNA* genes even correlated with their respective transcriptional activation in different life-cycle stages [149]. Further, a detailed 3C study on *rRNA* genes in mouse liver cells revealed looping interactions between the active *rRNA* promoter and terminator sequences, thus pointing to an efficient 'reloading' mechanism of RNA Pol I onto the *rRNA* gene [204].
This principle of looping DNA surrounding a focal hub of active polymerases and transcription factors has likewise been described for RNA Pol II-transcribed genes. Such RNA pol II hubs have been termed 'transcription factories'. A transcription factory is composed of ~4-30 RNA Pol II complexes and their associated factors (reviewed in [227]). Multiple genes located in *cis* and in *trans* colocalize and share the same transcription factory and may reposition to or away from a respective transcription factory dependent on their transcriptional status [209]. Rather than Pol II sliding along a DNA loop, it is assumed that the transcribed sequence is pulled through the rather immobile RNA Pol transcription complex [85]. A functional relation between Pol II transcribed genes associated in the same transcription factory could be shown for some cases [236]. The existence of a transcriptional interactome is, however, generally questioned [225] and associations based on spatial proximity seem more likely. Also, whether spatial colocalization of genes is cause or consequence of their mutual transcription still remains unanswered. In support of pre-existing transcription factories, heat-shock experiments inhibiting Pol II transcription initiation did not disrupt transcription factory organization [187]. Further, gene relocations into pre-established transcription factories upon gene activation have been described in several publications [221] [210]. Razin and colleagues, however, propose a model, in which RNA Pol II complexes first binds to their promoters and those pre-initiation complexes subsequently assemble into a transcription factory [225].
A topological organization into clusters was also identified for RNA Pol III transcribed *tRNA* in human cells [217] [265] and yeast [74] by chromosome conformation capture. Acting as insulators at TAD boundaries (see 2.2.6), *tRNA* genes are connected *via* DNA loops across the genome [265]. Not only is clustering of *tRNAs* related to their function, but also to their transcriptional activity as Van Bortle and colleagues described an organization of *tRNA* genes into clusters dependent on their respective expression rates [265].
In addition to a spatial compartmentalization of transcription, actively transcribed loci are generally characterized by increased local chromatin accessibility. For example, as measured by assay for transposase-accessible chromatin using sequencing (ATACseq) [37], highly transcribed *tRNA* genes reveal a sharp local increase in chromatin accessibility that is depended on the underlying gene's transcriptional activity [265].

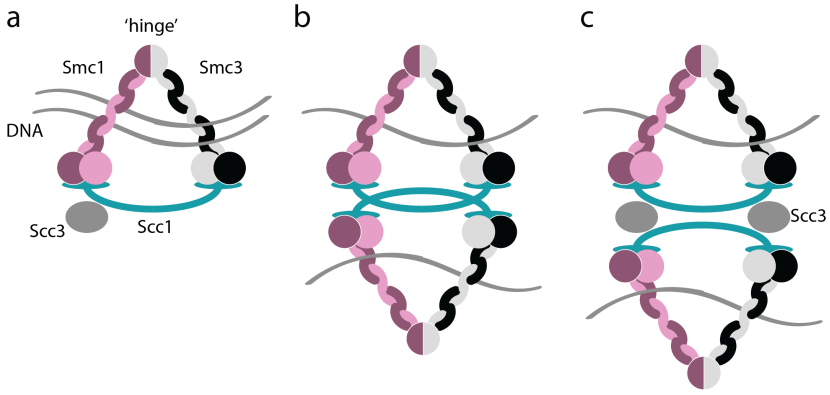
2.2.8 *Proteins establishing genome architecture*

Besides the physical properties of the DNA polymer and DNA-templated processes themselves that shape the 3D architecture, constraints on folding behavior are mostly exerted by proteins acting on the DNA. In this subsection, a brief overview will be given on the the architectural proteins cohesin and CTCF as well as on the capacities of histones, their modifications and variant forms to shape the 3D genome arrangement.

2.2.8.1 *Cohesin and CTCF*

Due to its function in sister chromatid cohesion during mitosis, the multiprotein complex consisting of the core subunits Smc1, Smc3, Scc1 (Rad21 homologue) and Scc3 was termed 'cohesin' [160]. Smc1 and Smc3 are connected at their globular hinge domain and bridged at their ATPase domains by the kleisin subunit Scc1, thus forming a tripartite circular complex that embraces the two sister chromatids of a replicated chromosome during S phase (strong ring model, Figure 2.3 a) [196]. Alternative (weak ring) models envisage two cohesin rings, each embracing a single DNA. The two cohesin rings are either connected by intercalation (Figure 2.3 b) or via the Scc3 subunit that is in association with both Scc1 subunits (Figure 2.3 c) [196].

Figure 2.3: Cohesin mediates DNA-DNA interactions



a Strong ring model. The cohesin complex embraces to DNA fibers. **b** and **c** Weak ring models. Two DNA fibers are individually bound by a cohesin ring. The cohesin rings interact by intercalation or via their Scc3 subunits.

Cohesins are loaded along the chromosomal DNA in G1 by the cohesin loading complexes consisting of Scc2 and Scc4 in yeast and humans (reviewed in [157]). Cohesin loading occurs presumably by transient opening of the Smc1-Smc3 hinge domain [100]. Cohesin binding occurs along the whole length of a chromosome and is not biased by DNA sequence. Rather, accumulation of cohesin loading

at distinct sites (e.g. at pericentromeric regions) is favored by DNA binding proteins or facilitated by chromatin remodelers: For example, in humans, the ATRX chromatin remodeler, that is also involved in H3.3 incorporation, is associated with cohesin deposition, but the mechanism employed is currently unclear ([228] [77], also reviewed in [157]). In humans, most of the cohesin complexes are already removed from the DNA in prophase by wapl. As wapl opens the ring at the Scc1-Smc3 interface, the cohesin rings are left intact and can be reloaded quickly in telophase [90]. Pericentromeric cohesion is only resolved by proteolytic cleavage of the kleisin subunit Scc1 by separase in late metaphase, which releases the two chromatids for separation in anaphase (reviewed in [213] and [196]).
In yeast, the loaded cohesin complexes are presumably shifted to sites of convergent transcription through the action of RNA polymerases [150]. In vertebrates, cohesin is preferentially located at DNAseI-hypersensitive sites [211], in non-coding intergenic regions and introns [277]. Here, cohesin appears to be positioned by the zinc-finger DNA binding protein CTCF, both in G1 and G2 phase [277]. The presence of CTCF is, however, restricted to bilateria [109] and is not required for cohesin loading *per se* or sister chromatid cohesion during mitosis [277]. Rather, cohesin binding at CTCF sites has implications for transcription regulation and the establishment of a cell type-specific functional nuclear architecture in G1 [102] [196]. As mentioned before (2.2.6), cohesin and CTCF constitutively insulate TADs by occupying their boundaries in an invariant manner, but may also structure intra-TAD domains at finer scales [69] [214]. Investigation of CTCF binding motifs at topological boundaries revealed a convergent orientation at the loop origin as requirement for loop formation [222]. This observation lead to the development of a 'loop extrusion model' for TAD formation [232]: Here, CTCF proteins bind the motifs in an oriented manner. Cohesin complexes loaded onto the DNA scan the DNA for CTCF proteins, thereby extruding a growing DNA loop off their ring structure. Loop extrusion is stopped once two opposing CTCF proteins are reached.
CTCF is thus understood as a global insulator across the genome, that segregates specific chromatin interactions and exerts its function ultimately through binding and positioning of cohesin [277]. However, Hi-C experiments performed by Zuin and colleagues in cohesin- and CTCF-depleted cells, respectively, indicated that CTCF alone is sufficient and necessary to maintain TAD boundaries and insulate interaction between them [293]. Loss of cohesin was rather associated with disrupted (promoter-enhancer) interactions formed inside a TAD [293]. In a more recent study, Rao and colleagues [223] claimed a complete loss of looped domains upon cohesin depletion. However, cohesin depletion also resulted in a stronger segregation of genome compartments, thus indicating that genome compartmentalization is

itself not driven, but counter-acted by cohesin-dependent DNA-DNA interactions [223]. Also, in lower eukaryotes such as *S. pombe*, which lack CTCF and TADs, the integrity of chromatin topology is dependent on the presence of cohesin [188].

2.2.8.2 *Histone marks and histone variants shape chromatin architecture*

The nucleosome represents the fundamental unit of chromatin compaction. Therefore, changes in chromatin topology are often accompanied by epigenomic changes at the histone level, e.g. modification of histone tails or deposition of histone variants.
As mentioned before (see 2.2.4), the human nucleus is divided into six nuclear subcompartments and loci within a respective compartment show correlated patterns of histone modifications and variants [222]. Subcompartments of the 'A'-type are enriched with activating chromatin marks (e.g. H3K36me3, H3K79me2, H3K27ac and H3K4me1) and histone variant H2A.Z. Subcompartments of the 'B'-type show varying patterns of chromatin marks, e.g. B1 is rich in H3K27me3 and depleted for H3K36me3 thus pointing to properties of facultative heterochromatin.
In addition to their presence across these large compartments, histone modifications can mediate distinct interactions. In *S. pombe*, cohesin loading is supposedly facilitated by histone methylation. Deletion of the sole methyltransferase Clr4 in *S. pombe* resulted in decreased chromosome territoriality and diminished inter-centromere and inter-telomere interactions due to reduced heterochromatin compaction [188].
Besides histone modifications, histones themselves and particularly their variant forms have been found to shape nuclear organization. Histone variants represent paralogues of their canonical counterparts H3, H4, H2A and H2B [255]. While canonical histones are expressed in a transcriptional burst during DNA replication and deposited at replication foci only, histone variants are expressed throughout the cell cycle [169]. A histone variant form may therefore be incorporated into a nucleosome in exchange of a synthesis-coupled canonical histone at any given time. The deposition of a histone variant may profoundly change the properties of the according nucleosome e.g. with regards to DNA binding, accessibility and interaction with neighboring nucleosomes [255]. It further holds the potential to alter or remove post-translational modification patterns [111]. The presence of histone variants at distinct genomic sites also serves their functional definition and stable inheritance. For example, the presence of a variant form of H3 (cenH3 or CENP-A) along the chromosome represents the basis for centromere identity and function in most eukaryotic genomes (reviewed in [111]).
Whereas there are only few examples for variant forms of H4 (e.g. in wheat [252], *Tetrahymena* [108] and *Trypanosoma* [245]), most eukaryotes

express at least two H3 variant forms [169]. Human genomes contain two canonical H3 histones (H3.1 and H3.2), a centromeric CENP-A [28] and five non-centromeric histone H3 variants (H3.3 [3], H3.X, H3.Y [278] and testis-specific H3.1t [281] and H3.5 [235]). The H3 variants are differentially associated with post-translational modifications and are employed in different biological contexts (e.g. gene activation or silencing) and tissues [104].
Histone variants of the 'H3.3-type' have conserved functions in different eukaryotic lineages and are widely distributed, however, H3.3 itself arose several times independently during evolution [169]. In mammals, H3.3 differs from canonical H3 only at four amino acid residue positions, possibly resulting in reduced packaging strength of the variant histone-containing nucleosome [169]. Different from canonical H3, which is deposited by the histone chaperone CAF-1 in a replication-coupled manner, H3.3 deposition at active and repressed genes is mediated replication-independently by the HIRA histone chaperone complex [254] [97].
Due to its accumulation outside of S-phase, histone variant H3.3, can serve as a replacement for canonical nucleosomes that were evicted in DNA-templated processes such as transcription or DNA repair. Therefore, H3.3 is predominantly found at sites of active chromatin and high nucleosome turnover, e.g. at promoters, actively transcribed coding regions and transcription start and termination sites [84] [111]. H3.3 is not essential in *Drosophila* and loss of H3.3 has revealed only mild effects on transcription. Here, H3.3 loss was found to be compensated by upregulation of canonical H3 (reviewed in [111]).
Despite a primary association of H3.3 with active chromatin, H3.3. is also present at repetitive and heterochromatic sites. In line with that, depletion of H3.3 in pluripotent mouse embryonic stem cells resulted in a reduction of repressive H3K27me3 at bivalent promoters [15]. At telomeres, H3.3 is deposited by the ATRX/DAXX chaperone complex in a HIRA-independent manner [97]. The presence of H3.3 at telomeres has been described to provide repressive H3K9me3 marks and H3.3 deficiency in mouse embryonic stem cells resulted in telomeric damage and transcriptional repression at telomeres [261]. In these lines, Elsässer and colleagues found that deletion of H3.3 led to decreased deposition of repressive H3K9me3 marks at retroviral elements and a derepression of adjacent genes [78]. These findings suggest that H3.3 also plays a role in transcriptional silencing, e.g. by maintaining the repressed state of telomeres [282], pericentric heterochromatin and retrotransposable elements [272]. The presence of H3.3 is therefore not exclusively associated with either active or silent chromatin. Rather, H3 variant deposition designates sites characterized by high nucleosome-turnover and is thus important in the maintenance of genome integrity [111].
In higher eukaryotes, H3.3 seems directly involved in nuclear organi-

zation since it was found to co-localize with the histone variant H2A.Z at CTCF binding sites. It has been proposed that H3.3 deposition at these sites facilitates binding of CTCF to linker DNA by excluding histone H1, which is known to compete for linker DNA occupancy (reviewed in [184]).
Non-allelic variants of histone H2A are highly diversified, with the variant forms H2A.Z and H2A.X being present throughout the eukaryotic tree of life [42]. The variant form H2A.Z shares only 60% sequence similarity with the corresponding canonical H2A histone [111]. The C-terminus of H2A.Z exhibits an extended acidic patch that mediates interaction with other histones and establishment of higher-order chromatin [42]. H2A.Z can fulfill diverse functions (e.g. transcriptional activation and repression, heterochromatin establishment, repair, replication [111]) dependent on different post-translational modifications of the histone variant [256]. In this respect, acetylated forms of H2A.Z are associated with instable nucleosomes at actively transcribed regions.
As mentioned before, H2A.Z co-localizes with H3.3 at CTCF binding sites. Here, H2A.Z is thought to aide CTCF deposition by impeding DNA methylation, since CTCF has been shown to only recognize un-methylated DNA [184]. Together with H3.3, H2A.Z builds a particularly fragile nucleosome that is found at insulator regions and at promoters [128], possibly conferring access of CTCF and cohesin to DNA [184]. Apparently, H2A.Z and H3.3 establish the chromatin context that is required for CTCF-mediated DNA interactions [184].
The same histone variant can exhibit very different, sometimes contradictory nuclear functions in different biological contexts, including roles in eu- and heterochromatin establishment, DNA accessibility and CTCF deposition. Thus, a single and precise contribution to nuclear architecture cannot be assigned to an individual histone variant form today.

2.2.8.3 *Genome architecture contributes to genome integrity*

Not only exogenous agents but also processes intrinsic to the cell continuously cause DNA lesions, thereby challenging genome integrity. A cytotoxic double-strand DNA break (DSB) can either be repaired by non-homologous end-joining (NHEJ) or homologous recombination (HR). The latter requires an intact homologous sequence as repair template. Therefore, in the process of an HR, the DNA at the site of the DSB is resected, resulting in protruding 3′-DNA overhangs. After binding of the Rad51 recombinase, one of the single-stranded 3′-ends probes the surrounding donor genome for regions of homology and eventually invades a fitting locus. The DNA at this site is displaced in a so-called D-loop structure, which serves as a template to elongate the 3′-strand. The 3′-strand then provides enough sequence to reanneal with its broken counterpart and to repair the lesion by synthesis-dependent strand annealing (SDSA). Alternatively, both protruding 3′-ends invade

the same region of homology, thereby forming a double Holliday junction. The break is then repaired in a gene conversion (GC) event (reviewed in [20]).
The success of a Rad51-driven homology search is not only depended on target size of the donor sequence [148], but also on chromosomal organization constraints. As measured by nuclease-induced DSB repair in budding yeast, the efficiency of a DSB repair by recombination is correlated with spatial proximity between the site of the lesion and the donor sequence as measured by Hi-C [148]. A donor sequence can be templated from the sister chromatid in mitotic cells, the homologous chromosome or ectopic sequences derived from non-homologous chromosomes (reviewed in [157]). In both, human and yeast, checkpoint proteins trigger cohesin recruitment and loading in proximity to a DSB [243]. By limiting the mobility of the damaged DNA site, cohesin promotes the usage of sister chromatids as repair template over a non-allelic replacement by ectopic sequences, which holds the risk of chromosomal rearrangements [68]. Agmon and colleagues investigated the effect of nuclear architecture on DSB repair efficiency [2]. They found that in yeast nuclei, high recombinational efficiency was restricted to regions with overlapping chromosomal territories, thus, again, preventing recombination between non-allelic loci. In line with these observations Lee and colleagues found DSB repair to be most efficient when a repair template was provided intra-chromosomally [148], stressing the importance of a territorial organization in genome maintenance.
Further, Agmon and colleagues showed that tethering of centromeres and telomeres in yeast promotes preferential homology search inside the same chromosomal arm [2]. Telomeric, subtelomeric, and centromeric regions are characterized by their repressive chromatin structure and their spatial association in heterochromatic foci in *trans* [180] [74]. Batté and colleagues found that in yeast the heterochromatic compaction of DNA at subtelomeres limits extensive resection at the site of the DSB by nucleases. This, in turn, led to reduced loss of genetic information and facilitated homologous recombination [20]. Interestingly, subtelomeric DSBs were preferentially repaired by break-induced replication (BIR). BIR involves a large sequence exchange between subtelomeres, thus leading to loss of heterozygosity. However, BIR is also highly mutagenic and may therefore drive genetic diversity at these genomic sites [20]. Recombinational repair events in repetitive heterochromatic domains (e.g. at pericentromeric regions) are error-prone due to the close spatial proximity and over-representation of repeated, non-homologous sequences. Recent studies have found, that at the site of the DSB, heterochromatin expands and the DSB is relocated outside of the heterochromatic domain. Only then Rad51 is recruited and homologous recombination repair is initiated [46] [127] [8].

These findings exemplify the central role of nuclear architecture in the context of DNA damage repair and genome integrity maintenance. Further, these examples show that nuclear architecture compartmentalizes diversification and mutagenesis to sites, where it is beneficial.

3

Genome architecture in the context of mutually exclusive expression

3.1 Mutually exclusive gene expression – principles and mechanisms

Mutually exclusive expression refers to the expression of a single gene from a large gene family with concomitant repression of all other gene family members. Exclusive expression of a singular gene is achieved either through interaction of the respective gene with a distinct regulatory element (e.g. enhancer) [171] [130] or through a recombinational event into a dedicated transcriptionally permissive environment or nuclear body [120] [199]. Both types of events are, however, subject to changes in nuclear architecture. In olfactory neurons, for example, one out of ~2800 olfactory receptor (OR) alleles is activated through the convergence of multiple *trans*-interacting enhancers building a transcriptional hub around the randomly selected gene [171] [189] [190]. The accumulation of activating transcription factors bound to the enhancer elements displaces heterochromatin from this transcription-competent hub [189]. The remaining *OR* alleles are silenced in heterochromatic foci from which the active *OR* gene escapes [11], thus emphasizing the importance of nuclear subcompartmentalization of transcription in this process.
During differentiation of B- and T-lymphocytes, diversity of antigen receptors is generated through the genomic recombination of *V*, *D* and *J* elements. The recombination of these distant gene segments requires locus contraction mediated by DNA looping [113].
Once the monoallelic choice is made in the above-mentioned examples, it remains for the rest of a cell's lifetime. In the context of evasion of an infecting pathogen from its host immune response, the mutually exclusive choice of an expressed immunogenic surface antigen is repeatedly altered, a process referred to as antigenic variation. Although evolutionary distant, antigenic variation is commonly used among different bacterial and eukaryotic microbial pathogens [63].
Antigenic variation requires the presence of a hypervariable antigen gene family, that is subject to enhanced mutagenesis or ectopic recombination driving diversification of the antigen repertoire [17]. These repertoires are often located in subtelomeric environments which are thought to provide both, reversible repression and homology for recombination through the close association to the telomere [17].
Immune evasion through antigenic variation maintains a persistent infection of the host and potentially allows for re- or superinfection of

the same organism [63]. The mechanisms for antigenic variation are diverse and may be of epigenetic nature (e.g. in *Trypanosoma brucei*, *P. falciparum*), involve recombination (e.g. in *Treponema pallidum*, *T. brucei*) or interfere with translation (e.g. in *Giardia lamblia*) (reviewed in [63]). A prerequisite is, however, to follow a certain hierarchy in antigen expression at the single-cell and population level to not exhaust the antigen repertoire too quickly. These tendencies have been observed in different organisms, among which is *T. brucei*. The mechanisms of antigenic variation in *T. brucei* will be introduced in the following section.

3.2 Antigenic variation in *Trypanosoma brucei*

The protozoan parasite *T. brucei* is a major pathogen in sub-Saharan Africa and the causative agent of Human African Trypanosomiasis (HAT) (also referred to as sleeping sickness) and nagana in livestock. Grouped within the order of Kinetoplastida, the monophyletic group of trypanosomatids branched early from the eukaryotic tree of life [250]. From their single origin, trypanosomes adapted to infect a range of vertebrates through adaptive radiation [250]. Their parasitism includes a complex life cycle between their vertebrate host and (blood-sucking) insect vector.
African trypanosomes are transmitted through the bite of the tsetse fly (*Glossina spp.*) and circulate extracellularly in the mammalian host bloodstream and tissue-spaces. The parasites are therefore constantly visible for the hostile immune system which they evade through antigenic variation of their dense VSG coat [117]. A repertoire of more than 2000 *VSG* genes [56] is distributed across the subtelomeres of its 11 megabase chromosomes (MCs, 1-6 Mb in size). The *VSG* archive is further diversified to a small number of intermediate- (ICs, 100-800 kb) and ~100 mini-chromosomes (25-100 kb) [181].
A single *VSG* gene is, however, only expressed when located in one out of ~15 subtelomeric, RNA Pol I transcribed polycistronic bloodstream-form expression sites (BESs), which are located upstream of some telomeres. Importantly, only one expression site is active at any given time, thus ensuring monoallelic expression of a single *VSG*. Monoallelic expression combined with periodical switching between immunologically distinct *VSGs* allows repeated population of the bloodstream upon elimination of a large proportion of the parasite population by the host, thus permitting the establishment of a persistent infection.
The genome of *T. brucei* is of an unusual structure and a great proportion is dedicated to immunologically different surface antigens. The subtelomeres harboring the vast *VSG* repertoire had not been assembled at the start of my thesis work [194] (see 3.3). In the following sections, the structure of the linear genome of *T. brucei* shall be introduced, including its unusual modes of transcription. Subsequently,

the epigenetic mechanisms, including changes in spatial arrangement of the genome that are known to be involved in the regulation of antigenic variation in *T. brucei*, will be described.

3.3 The *T. brucei* Lister 427 genome

3.3.1 *De novo assembly of the T. brucei using long-read sequencing and genome contact data*

Previous attempts by Melville and colleagues to karyotype the megabase chromosomes of the monomorphic *T. brucei* Lister 427 laboratory isolate using pulsed-field gel electrophoresis (PFGE) revealed a high degree of size-variation between homologous chromosomes and larger chromosome sizes than previously estimated [181]. Based on these experiments, the diploid genome content was estimated to be ~70 Mb [181], compared to previous estimates of ~35-40 Mb [29]. A recent *de novo* assembly of the Lister 427, generated with Pacbio and Hi-C data I produced as part of my thesis work, confirmed the estimates made by Melville and colleagues [194]. As the genome assembly is based on data acquired during my thesis work, parts of the assembly strategy and the resulting genome shall be described in detail in this section.
As described in 1.2.3, conformational data acquired by Hi-C can be used for the scaffolding of genomes. Our recent Lister 427 *de novo* genome assembly based on PacBio SMRT sequencing [76] and Hi-C data resulted in an almost complete and partially phased assembly of the diploid *T. brucei* megabase chromosomes (see Figure 5.2.3.3 and Table 3.1 [194]). PacBio SMRT sequencing of genomic DNA from the *T. brucei* Lister 427 isolate resulted in a ~100-fold coverage of the parasite genome (642,583 reads, N50: 11.94 kb). The reads were assembled into contigs by overlay-layout-consensus (OLC), error-corrected, extended and further joined. The contigs were subsequently scaffolded based on interaction frequencies as obtained by Hi-C and further joined, which resulted in 91 contigs representing the *de novo* karyotype of the 11 megabase chromosomes. For the two homologous chromosomes, the diploid 'core' part (22.71 Mb) containing the housekeeping genes appeared highly conserved between the two respective chromosomes and the assembly showed a high degree of synteny compared to the TREU 927 *T. brucei* isolate [194].

3.3.2 *Subtelomeres are haploid-like and harbor the VSG repertoire*

The combination of long-read sequencing with 3D contact information allowed to assemble for the first time the subtelomeric fraction of the megabase chromosomes. Despite the presence of repetitive DNA and the great similarity between the *VSG* genes, the high degree of heterozygosity between the *VSG* arrays located at the respective sub-

telomeric regions allowed a haplotype-specific assembly [239] of most 5′ and 3′ chromosomal ends. In the following, the two alternative, heterologous subtelomeres are designated as 'A' and 'B' subtelomeres (19.54 Mb of total subtelomeric sequence) [194].
A BLAST search of the 2582 *VSG* genes of the previously published *T. brucei* VSGnome [56] in the *de novo* assembly revealed that 2486 of them were represented (more than 90% query coverage and at least 95% sequence identity).
A substantial portion (11.08 Mb) of sequence still remains unassembled and presumably contains sequences belonging to the sub-megabase portion of the genome, such as intermediate- and minichromosomes.

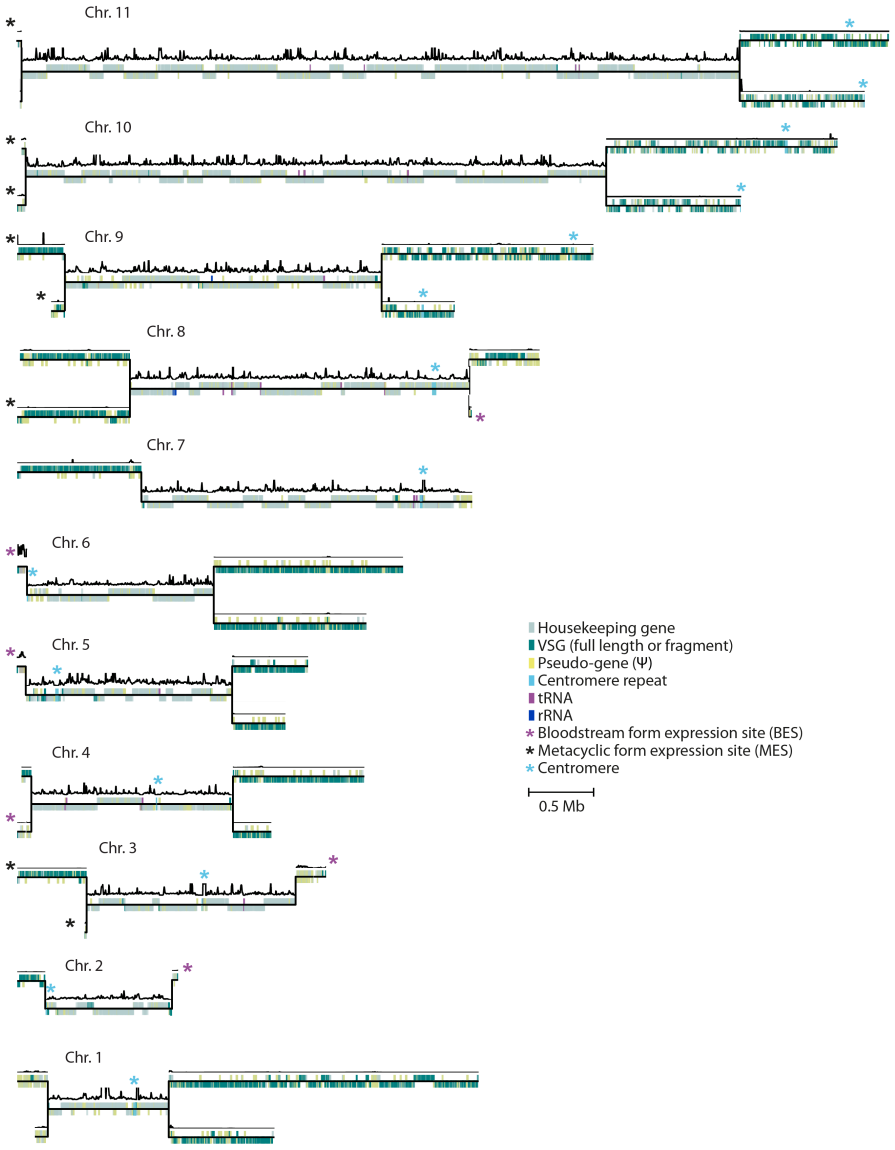
Table 3.1: Parameters of the Tb427v9 genome assembly

The Tb427v9 genome	
Total megabase genome size	42.25 Mb
Core regions (diploid)	22.71 Mb
Subtelomeric regions (haploid)	19.54 Mb
Megabase chromosomes (MCs)	11
Contigs	96
Expression sites assigned to MCs	14
Sequence assigned to sub-megabase chromosomes	0.8 Mb
Unscaffolded contigs	11.08 Mb

3.3.3 *Assignment of bloodstream form expression sites*

In the past, the *VSG* containing bloodstream form expression sites (BESs) have been sequenced in their full length using BAC [25] and TAR clones [112]. Efforts to assign BESs to their respective chromosomes were previously made by dissecting chromosomes *via* electrophoresis [199] [55] [181] [112].
In our genome assembly, BES8 was the only BES directly assembled into a megabase chromosome (3′-end of chr8). However, analysis of the relative interaction frequencies between the remaining BESs and chromosomal endings allowed to place all BESs located on megabase chromosomes accordingly. Thereby, previous estimates about BES positions could be confirmed (BES1 as part of chr6, BES3 as part of chr4, [199] [55] and new assignments could be made (BES12 to chr2, BES15 to chr3 and BES5 to chr5). Further, megabase BESs could be assigned to the 3′ or 5′ end of a chromosome, respectively. In three cases (BES3, BES8, BES15) even the according subtelomeric allele (A or B) could be identified. Table 3.2 summarizes all BESs and *VSG*s

Figure 3.1: The *T. brucei* Lister 427 genome



The homozygous core of each diploid megabase chromosome (Chr.1 -11) is depicted once together with the according heterozygous subtelomeres at the 3′ or 5′ endings. Relative transcript levels are depicted in black on top of each chromosome (window size 5,001 bp, step size 101 bp). The assignment of bloodstream form and metacyclic expression sites are depicted by green and black asterisks, respectively. Centromeric regions are depicted by blue asterisks.

together with their previously predicted and newly assigned genomic locations.

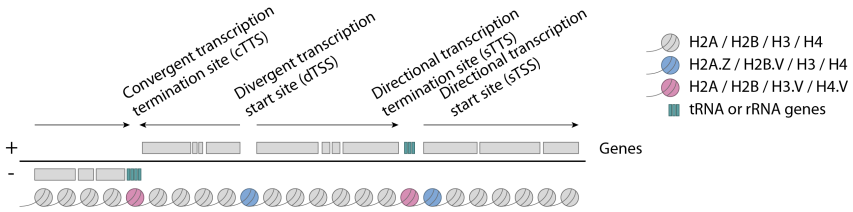
Table 3.2: Overview of BES assignments to megabase (MC) and intermediate chromosomes (IC).

Expression site	*VSG*	Chromosome	Arm	Allele
BES1	*VSG-2*	MC 6	5′	
BES2	*VSG-9*	IC		
BES3	*VSG-6*	MC 4	5′	B
BES4	*VSG-21*	IC		
BES5	*VSG-18*	MC 5	5′	
BES7	*VSG-3*	MC 6	5′	
BES8	*VSG-14*	MC 8	3′	B
BES10	*VSG-15*	IC		
BES11	*VSG-16*	IC		
BES12	*VSG-8*	MC 2	3′	
BES13	*VSG-17*	IC		
BES14	*VSG-8*	MC 7	3′	
BES15	*VSG-11*	MC 3	3′	A
BES17	*VSG-13*	IC		

3.4 The epigenome controls transcription in *T. brucei*

Besides transcription of the active *VSG*, only a minority of genes is promoter-driven in *T. brucei*, among which are the spliced-leader (SL) RNAs (RNA Pol II), *procyclins* and *procyclin-associated genes* (PAGs) (RNA Pol I), rRNA (RNA Pol I) and tRNA genes (RNA Pol III). These gene families have distinct distributions across the genome. For example, the SL genes are arranged in tandem on chr9 and arrays of rRNA genes are mainly located at the borders of centromeric regions of different chromosomes [4]. tRNA genes are specifically found in the small region between the start and the end of RNA Pol II transcription units ([245], see below), where they are thought to act as insulators and contribute to RNA Pol II transcription termination as observed in higher eukaryotes ([264] [245]).
RNA Pol II-transcribed housekeeping genes are arranged in polycistronic transcription units (PTUs, see Figure 3.2) and are unlikely to be regulated at the transcriptional level [47] [276]. Two adjacent PTUs can be transcribed in a convergent, divergent or non-divergent (head-to-tail) orientation [245]. The transcripts of a polycistron are co-transcriptionally separated in a *trans*-splicing reaction that adds

Figure 3.2: Generic structure of a polycistronic transcription unit.



Polycistronic transcription units (PTUs) are arranged in a convergent (head-to-head), divergent (tail-to-tail) or unidirectional (head-to-tail) fashion. Transcription start- and termination sites of RNA Pol II transcription are demarked by histone variants H2A.Z/H2B.V (blue) and H3.V/H4.V (red), respectively. Canonical histones are depicted in gray.

a capping structure, the 'spliced-leader' and a poly-A-tail to every mRNA (reviewed in [155]).
Transcription is not dependent on the presence of well-conserved core promoters but initiates at dispersed, unregulated promoters inside the TSS sequence [276]. Initiation seems rather driven by the deposition of histone variants destabilizing TSSs nucleosomes [42], thus mediating chromatin accessibility for the transcription machinery [245] [276]. In fact, the boundaries of each transcription unit are demarked by the presence of four *T. brucei* histone variants, namely H2A.Z and H2B.V at transcription start sites (TSSs) and H3.V and H4.V at transcription termination sites (TTSs) [163] [245]. TSS nucleosomes are also characterized by PTMs at canonical histone tails (e.g. H4K10ac [245], H3K4me3 [284]) and the DNA flanking a PTU exhibits a modified thymine base (β-D-glucosyl-hydroxymethyluracil, or 'base-J') (reviewed in [30]). In agreement with the lack of focused promoter motifs, no *trans*-acting elements regulating transcription have been identified until now.
These observations attribute a central role to histone variants and -modifications in RNA Pol II transcription initiation and termination and link these DNA-templated processes to chromatin architecture. Two of the *T. brucei* histone variants, H3.V and H4.V, are particularly interesting since homologs present in other eukaryotes remain elusive [162] [245]. These two variants shall be described in more detail in the following.

3.4.1 *T. brucei histone variants H3.V and H4.V*

Unlike its canonical counterpart, the sole H3 variant (H3.V) that is conserved among kinetoplastids, is encoded as a single-copy gene in *T. brucei* [162]. H3.V shares ~58% identity in amino acid sequence with the canonical form of H3 in *T. brucei* (see Figure 3.3, upper panel) and up to 62% sequence identity with the canonical H3 of other eukaryotes,

but sequence conservation is largely limited to histone C-termini [162]. It does, however, neither represent a homolog to any other eukaryotic H3.V nor does it meet the structural requirements of a conventional CenH3 [162].
Previous ChIPseq, fluorescence *in situ* hybridization (FISH) and immunofluorescence (IF) based microscopy experiments in *T. brucei* have localized H3.V to the 177-bp repeats of mini-chromosomes, telomeric repeats, subtelomeres and RNA Pol II transcription termination sites [245] [162]. It does, expectedly, not localize to centromeric regions and is therefore not considered as the trypanosomal CenH3 [162]. In agreement with that, *Tb*H3.V appears to be not essential for viability or chromosome segregation under laboratory conditions [162].
Despite the prominent localization at telomeres, H3.V is not required for telomere length maintenance [162] [238] and – accordingly – Lowell and colleagues did not observe *H3.V* knockout to impair silencing of inactive *VSGs* located in BESs [162]. In contrast, other laboratories observed a mild (~4-fold) derepression of *VSGs* located in BESs, but no increased switching rates upon *H3.V* deletion [226] [238]. Reynolds and colleagues concluded H3.V to function in attenuation of RNA Pol I driven transcription specifically at the *VSG* gene, not along the whole BES [226].
At TTSs, H3.V acts independently from but synergistically with base-J in bloodstream form *T. brucei* [226] [238]. Here, H3.V seems to function in the inhibition of RNA Pol II transcription elongation as loss of H3.V leads to 'read-through' of transcription and increased generation of siRNA transcripts at convergent TTSs [226].
H4 is a slowly evolving histone and variant forms are therefore rare [257]. The *T. brucei* H4 variant, which was first described by Siegel and colleagues, shares 85% sequence identity with its canonical counterpart (see Figure 3.3, lower panel) [245]. The genomic distribution of H4.V mirrors that of H3.V its enrichment at telomeric or subtelomeric sites is much less pronounced than that of H3.V [245]. Just like H3.V, H4.V is not essential in *T. brucei* [162] [245].

3.5 Molecular mechanisms of antigenic variation in *T. brucei*

The generic *VSG* expression site active in the bloodstream form consists of a *VSG* promoter that recruits RNA Pol I [291], largely syntenic expression site associated genes (ESAGs), often related to host-parasite interactions [27] [285] and the telomere-proximal *VSG* gene [112]. Both, the *VSG* genes located in an expression site and in subtelomeric arrays are mostly flanked by common 70-bp repeats upstream [158]. Thus, the ES *VSG* is literally embedded between the repetitive sequences of the 70-bp repeats and the telomere.
Antigenic variation occurs e.g. through duplicative transposition of

Figure 3.3: H3 and H4 canonical and variant forms in *T. brucei*

```
H3.V    RWRPGTVALREIRRLQSSTDFLIQRAPFRRFLREVVSNLKDSYRMSAACVDAIQEATETY  103
        RWRPGTVALREIR+ Q STD L+Q+APF+R +REV    K+  R  ++ + A QEATE+Y
H3      RWRPGTVALREIRQFQRSTDLLLQKAPFQRLVREVSGAQKEGLRFQSSAILAAQEATESY  97

H3.V    ITSVFMDANLCTLHANRVTLFPKDIQLALKLRGER  138      58% identity
        I S+  D N   +H+ RVT+ PKDI LAL LRGER          72% positives
H3      IVSLLADTNRACIHSGRVTIQPKDIHLALCLRGER  132

H4.V    MAKGKRVGESKGAQKRQKKVLRDNVRGITRGSIRRLARRAGVKRISGVIYDEVRGVLKTF  60
        MAKGK+ GE+KG+QKRQKKVLR+NVRGITRGSIRRLARR GVKRISGVIYDEVRGVLK+F
H4      MAKGKKSGEAKGSQKRQKKVLRENVRGITRGSIRRLARRGGVKRISGVIYDEVRGVLKSF  60

H4.V    VESIVRDAGAYTEYSRKKTVTAAHVVFALRKRGKVLYGY  99       86% identity
        VE +VRDA AYTEYSRKKTVTA  VV ALRKRGK+LYGY          92% positives
H4      VEGVVRDATAYTEYSRKKTVTAVDVVNALRKRGKILYGY  99
```

Amino acid sequence alignments of H3.V and H3 (top) and H4.V and H4 (bottom). Conserved amino acid residues are written in blue, similar substitutive residues are marked with a plus symbol.

a novel *VSG* gene originating from the subtelomeric archive or from a silenced ES into the currently transcribed ES. This *recombinational switch* is thought to be triggered by spontaneous double-strand breaks around the actively transcribed *VSG* [93] and facilitated by DNA sequence homology provided by the *VSG* flanking sequences or the *VSG* itself [158] [118]. Partial conversion of a *VSG* gene results in the formation of mosaic genes [105] [193], thus expanding the *VSG* even further.
Switches in *VSG* expression can also occur through transcriptional inactivation of the active expression site with simultaneous activation of an alternative ES, also referred to as *in situ* switching (reviewed in [116]). This form of allelic exclusion is thought to be epigenetically controlled as it does not involve alterations at the level of DNA sequence.
Several factors associated with epigenetic control and regulation of *VSG* expression are discussed in the following sections.

3.5.1 *Epigenetic factors involved in antigenic variation in T. brucei*

3.5.1.1 *Spatial arrangement of the T. brucei genome*

Positioning and dynamics of silent and active expression sites appear critical in ensuring monoallelic expression of *VSGs*. As mentioned before, in bloodstream form trypanosomes, transcription of the active expression site is thought to occur in a dedicated, membrane-less nuclear compartment, the ESB [45] [199]. The ESB is an extranucleolar RNA Pol I-rich focus with high transcriptional activity. In contrast to the actively transcribed expression site, imaging of GFP-tagged inactive expression sites [199] and *in situ* hybridization against the 50-bp repeats located upstream of each BES [45] [199] have shown a preferen-

tial peripheral location of silent expression sites in bloodstream form parasites. However, no strict association with the nuclear envelope was observed. In line with that, silenced BESs are not entirely repressed but show slight transcriptional activity at the promoter region [133]. These findings point to a 'poised' state of the inactive BESs in bloodstream form *T. brucei* cells.
In contrast, imaging of the 50-bp repeats in procyclic cells, where *VSG* activation is not required, indicated a more peripheral location of expression sites as compared to bloodstream form parasites [219]. In the process of differentiation from bloodstream form to procyclic stage, the active BES promoter tagged with green fluorescent protein (GFP) was observed to rapidly locate to the nuclear periphery whilst the ESB dissociates [144]. These findings suggest a silencing mechanism in the surrounding heterochromatic nuclear periphery. The importance of the lamina in BES silencing was also shown in knockdown experiments of the trypanosomal lamin-orthologue nuclear pore complex protein-1 (NUP-1), which led to disruption of nuclear organization, derepression of silent BESs and increased *VSG* switching rates. Notably, NUP-1 knockdown also inhibited repositioning of the active BES during differentiation [73].
The importance of the lamina as a repressive compartment became apparent in Br-UTP labeling of nascent RNAs, which revealed a transcriptionally competent nuclear center with foci of Br-UTP and a transcriptionally silent nuclear periphery [9] [200]. These findings point towards a radial arrangement of active and inactive genes in *T. brucei* with the nucleolus and the ESB being the most transcriptionally active compartments [199].
Importantly, in bloodstream-form parasites, the ESB separates late in mitosis, possibly mediating the stable inheritance of the *VSG* expression status [145]. This bipartite division of the main genome and the active expression site is driven by cohesin. Upon cohesin knockdown, stable inheritance of the *VSG* expression status failed and switching frequency between BESs was increased [145].
How the ESB is established at the actively transcribed site or whether it is a pre-defined subcompartment recruiting a single expression site is currently not understood. Regulatory DNA-elements such as an enhancer that would stochastically engage with a single expression site has remained elusive. However, proteins such as VSG exclusion-1 (VEX-1), that are involved in the maintenance of allelic exclusion have been found in close association with the ESB [94]. VEX-1 sequestration to the ESB ensures monoallelic exclusion of *VSGs* as ectopic knockdown or overexpression resulted in activation of multiple *VSGs* located in expression sites.
Previous work on antigenic variation and differentiation in trypanosomes has revealed a close link between developmentally regulated genes (e.g. procyclins and *VSGs*) and genome architecture [45] [199]

[144] [145]. In an environment of mostly unregulated RNA Pol II transcription, these genes are precisely regulated and transcribed from privileged spatial compartments. The epigenetic memory of *VSG* expression is – at least in parts – mediated by cohesin [145] and is therefore preserved and transmitted to the progeny by spatial DNA conformation.

3.5.1.2 *Differential chromatin remodeling of the active and silent expression sites*

Besides their differential nuclear positioning, the active and silent expression sites substantially differ in their chromatin landscape.
In *Plasmodium*, silencing of *var* genes has been shown to be dependent on histone-deacetylases of the sirtuin family [86]. In *T. brucei*, Sir-2 has been found to repress expression of a subtelomeric reporter, however, loss of Sir-2rp1 retained *VSG* silencing [7].
Still, proximity to the telomere appears important for maintenance of the heterochromatic state at silent ESs since disruption of telomere integrity by depletion of the telomere-binding protein RAP1 leads to a derepression of silent ESs and loss of mutual exclusive *VSG* expression [287].
The active expression site shows a less condensed chromatin landscape permissive for transcription. Using MNase digestion and ChIP against H3, Figueiredo and Cross found the active expression site to be nucleosome depleted [82]. They further showed that chromatin at the active BES is more accessible than at silent BESs. Furthermore, artificial depletion of nucleosomes at silent expression sites (e.g. through knockdown of H1 [218] or H3 [6]) or remodeling of their chromatin towards increased accessibility (e.g. via depletion of ISWI [122], of FACT [66], of histone chaperones ASF1A and CAF-1b [6], of the histone-deacetylase DAC3 [273], inhibition of the bromodomain protein BDF3 [237]; also reviewed in [92]) led to a derepression of silent BESs. For example, the histone-methyltransferase DOT1B, conferring H3K76 trimethylation, is required to maintain repression specifically at silent BESs [83] and to attenuate a previously active BES upon activation of a novel one [19]. Loss of DOT1B led to a ~10-fold up regulation of transcription from silent ESs and slower *in situ* switching [83].
In line with these observations, a modified thymidine (β-D-glucosyl-hydroxymethyluracil or also referred to as base-J), was found to commonly associate with transcriptionally silent regions [238]. In addition, it is located at repressed BESs, but not at the actively transcribed BES [266]. However, its role in antigenic variation remains elusive since deletion of base-J alone did not result in *VSG* derepression from inactive sites [238].
These and other experiments clearly assign a central role to the epigenetic remodeling of active and silent expression sites in monoallelic exclusion of *VSGs* besides the spatial arrangement of active and

silent sites in the nucleus. However, none of the abovementioned experiments have yielded a complete abrogation of allelic exclusion, suggesting that it is mechanistically controlled at multiple levels and controlled in a complex epigenetic network in *T. brucei*.

4

Aim of the study

Genome architecture has been proven key in several examples of monoallelic exclusion [171] [120] [199] [113] [130]. Studies on antigenic variation in *Trypanosoma brucei* have shown that mutually exclusive expression of *VSGs* is controlled at multiple levels and that both, global 3D genome architecture and local chromatin folding are driving forces for allelic exclusion in trypanosomes (reviewed in [92]). While different studies have investigated the chromatin landscape at active and silent expression sites, little research has focused on the identification of factors determining spatial positioning of the DNA in *T. brucei*.
The aim of this study was to gain a comprehensive picture of the *T. brucei* genome in three dimensions. As part of my thesis work, I sought to understand the relative spatial arrangement of chromosomes in *T. brucei* by applying chromosome-conformation capture (Hi-C) to this unicellular parasite. I aimed at analyzing the spatial conformation of chromatin with respect to transcription, differentiation and antigenic variation.
Specifically, I sought to answer the following questions:

- How are chromosomes structured and folded in *T. brucei*?
- How does genome architecture developmentally change?
- Do chromosomes exhibit levels of subcompartmentalization?
- How does genome architecture contribute to transcriptional regulation of housekeeping and *VSG* genes?
- How are the expression sites and the vast *VSG* reservoir, that comprises approximately one-third of the *T. brucei* genome, arranged in the nucleus?
- Are there any *trans*-acting elements involved in the activation of a single *VSG*?

Further, I aimed at finding ways to perturb the particular DNA architecture of *T. brucei* to address the following questions:

- Which factors and architectural proteins establish an maintain genome architecture in *T. brucei*?
- What is the effect of disrupting nuclear architecture on antigenic variation?

Answers to these questions will help to understand how an early diverging single-cellular organism such as *T. brucei* – with a presumably simpler genome architecture and interaction network compared to as found in higher eukaryotes – is able to tightly control monoallelic expression of antigens.

Part II

Results

5

The *T. brucei* Lister 427 genome in three dimensions

The following chapter describes the development of a Hi-C pipeline in *T. brucei*, including both the experimental setup (see Figure 5.1 a) and the computational analysis (see Figure 5.1 b). Further, characteristics of the three-dimensional *T. brucei* genome fold shall be described.

5.1 Establishment of Hi-C in *Trypanosoma brucei* cells

5.1.1 *Experimental setup*

To study the spatial arrangement of the megabase chromosomes in *T. brucei*, I established Hi-C in *T. brucei* nuclei based on previously published protocols [22][222].
Hi-C protocols conducted with human cells [156] used 100 million nuclei as starting material, yielding approximately 600 μg of genomic DNA. To account for the ~80-fold difference in genome size between human (~6 pg DNA per nucleus [41], ~3200 Mb [14]) and *T. brucei* nuclei (~0.04 pg [31]), ~40 Mb [194]), I subjected $2x10^8$ cells to each Hi-C experiment to obtain a comparable read coverage and resolution across the genome (~0.2 μg/Mb).
The 'visibility' of each genomic locus in a Hi-C experiment is highly dependent on the abundance of the targeted restriction site in a respective area. To obtain high-resolution Hi-C matrices for genome-wide interaction analyses, the four-base pair cutter MboI, recognizing 5′-GATC-3′-sites, was used (Figure 5.1 a). In an Hi-C experiment, only read pairs with both mates uniquely aligned can be considered, the remainder is discarded. Due to the high degree of repetitiveness in the *T. brucei* genome, only ~70% of the genome is unique. Additionally, a considerable proportion of the genome remains unassembled (~21%) and accounts e.g. for mini-chromosomal DNA and intermediate chromosomes. Taking this into consideration, only in approximately for half of the reads both read mates can be mapped uniquely to the genome. We therefore aimed for ~$1x10^8$ reads for each Hi-C library.
First attempts, however, yielded only low complexity libraries, which were particularly characterized by a high fraction of PCR duplicates. These libraries, were quickly sequenced beyond saturation (see Figure 5.1 c). Subsequently, higher library complexity was achieved by i) conducting digestion and ligation *in situ* [222], ii) performing the end-repair of chimeric molecules prior to the biotin-pull down step [22] and iii) splitting library amplification into multiple reactions, thereby

reducing amplification cycles [22] (Figure a.2). These adjustments led to a reduction of duplicates from ~71% to ~1% and increased valid interactions from 22.5% to 91% in the uniquely mapped reads (see Figure 5.1 c and Table 5.1).

Table 5.1: Hi-C mapping statistics in single-marker triplicates

	SM 1	**SM 2**	**SM 3**
Valid pairs	39621230	29246517	34224540
Duplicates removed	39118877	28965016	33821714
cis	18077783	14129415	15873681
trans	21041094	14835601	17948033
Dangling-end pairs	3267860	1973862	2183522
Religation pairs	512445	333170	383430
Self-cycle pairs	67045	41226	46734
Single-end pairs	0	0	0
Dumped pairs	3668	3350	3340

Mapping statistics for three Hi-C libraries on single-marker *T. brucei* as determined by *Hi-C Pro* [240]. Values for *cis* and *trans* interactions are under- and overestimated, respectively, as only one of the two subtelomeric alleles can be joined to the respective core region. Interactions between core regions and the remaining subtelomere therefore appear as *trans* interaction.

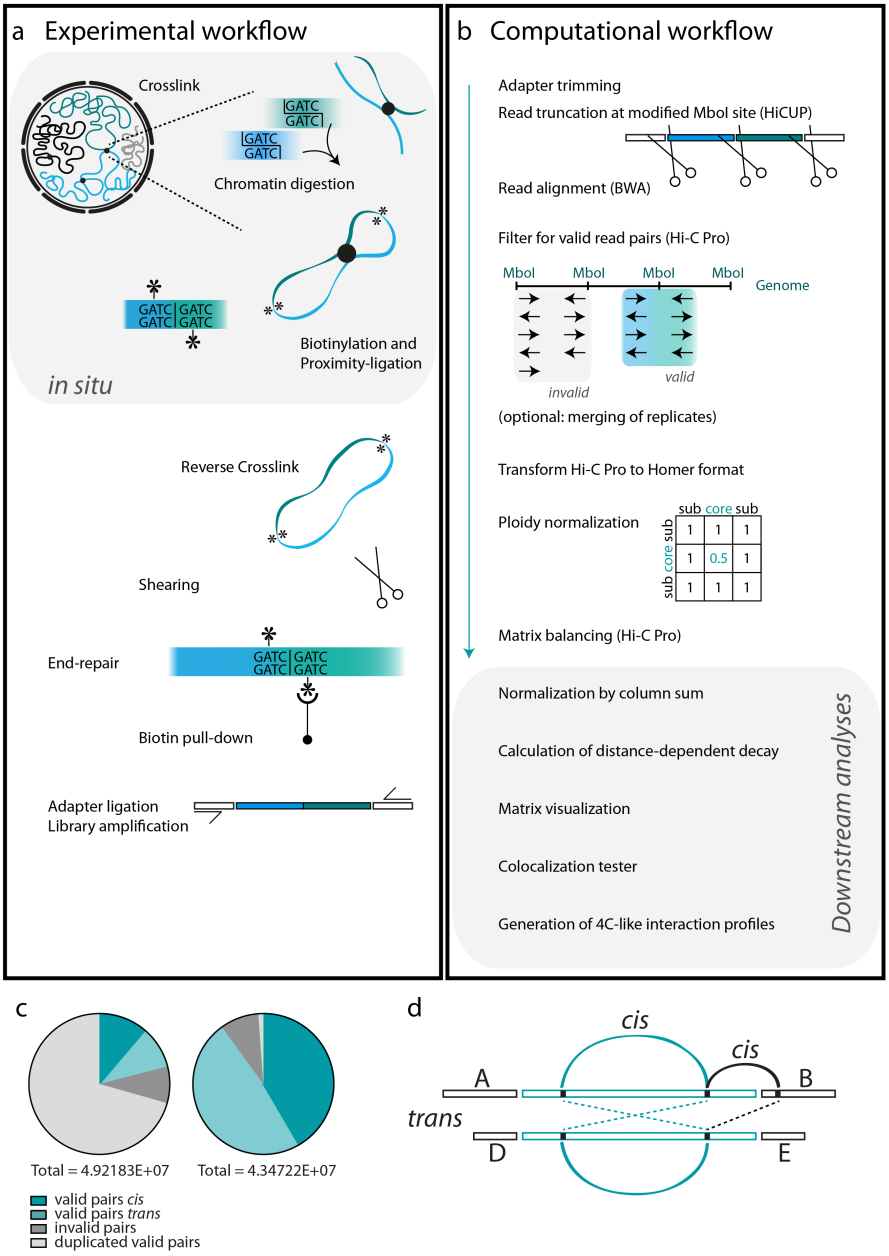
5.1.2 *Processing of Hi-C data to study 3D genome architecture in Trypanosoma brucei cells*

To filter, normalize and analyze the Hi-C data, we applied publicly available computational packages and complemented these with custom python scripts [194] to account for biological peculiarities of the *T. brucei* genome.

5.1.2.1 *Mapping and filtering of reads*

During the mapping process, Hi-C forward and reverse read pairs are aligned independently of each other as they do not represent a continuous DNA stretch. Secondly, in case a sequencing read covers the ligation junction, a respective aligner needs to be capable of discerning the two parts of the chimeric molecule.
To meet these requirements, previous publications relied on iterative mapping [125] or on re-mapping 5′-ends of chimera that could not be aligned [240]. Here, we used HiCUP [280] to examine the reads for the presence of the modified MboI site and eventually truncate the sequence at this site (Figure 5.1 b). Otherwise, the presence of a 'hybrid' sequence would hinder read mapping by conventional aligners.

Figure 5.1: Generation and analysis of Hi-C data.



a Experimental workflow. Restriction digest with *MboI*, biotin-labeling and proximity-ligation were performed *in situ*. Prior to library generation, the chimeric molecules were fragmented and those molecules with a biotinylated ligation junction were captured in a pulldown step with streptavidin. **b** Computational workflow. Sequencing reads are truncated at the ligation junction using HiCUP and read pairs are aligned independently using BWA, retaining only uniquely mapping reads. Filtering for valid read pairs is conducted with Hi-C Pro. Only after the removal of PCR duplicates, replicates are merged. Read counts are normalized according to their genomic ploidy before balancing of matrices (Hi-C Pro). All downstream analyses are custom python scripts (see methods section).

c Proportions of valid and invalid read pairs before and after adjustments in library preparation. **d** Ploidy normalization. Due to the heterozygous core-subtelomere junctions between the two homologous chromosomes, *cis*-interactions in the core region (green bow) occur twice as often as *cis*-interactions between a core and a respective subtelomere of the same chromosome (black bow). Interactions in *cis* (dashed lines) are rare in relation to *cis*-interactions and are therefore not considered in the ploidy normalization calculation.

We tested different short read mappers to seek for the best recovery of truncated reads mapping uniquely to the *T. brucei* genome. A direct comparison of Bowtie2 [146], STAR [70] and BWA-mem [152] was carried out by subjecting the same set of truncated Hi-C reads to the respective aligners. The amount of uniquely mapped reads was computed using SAMtools [153]. Whereas STAR did not align a great proportion of reads, Bowtie2 and BWA-mem performed comparably robust in mapping percentage, even when filtering for high-quality, uniquely mapped reads (Table 5.2). We implemented BWA-mem in our pipeline due to its high mapping accuracy [134] and capacity to also recognize read chimera.
Genomic sequences ('unitigs') that could not be assembled into the scaffold of the 11 megabase chromosomes of *T. brucei* were included in the reference genome during the mapping process to ensure the most unambiguous alignment of every single read whilst reducing 'false positive' alignments.
Subsequent to read mapping, we used Hi-C Pro [240] to filter out invalid read pairs, representing duplicates, dangling ends and (circularized) self-ligation products (see Figure 5.1 b and supplementary data Figures a.3-a.9).

Table 5.2: Comparison of aligners for Hi-C reads.

	BWA-mem	**Bowtie2**	**STAR**
MAPQ>1	78.55%	99.77%	38.74%
MAPQ>10	76.73%	77.77%	26.62%
MAPQ>30	65.22%	67.53%	26.62%

Percentage of reads aligned when filtered for different read mapping quality scores.

5.1.2.2 *Adjustments for chromosomal ploidy*

The homologous *T. brucei* chromosomes share a homozygous core region that is flanked by heterozygous subtelomeres [194] (Figure 5.2.3.3). Raw interaction matrices were generated by mapping the Hi-C reads onto a genome where core regions and subtelomeres were

represented only once. However, these differences in ploidy cause a bias when calculating (intra-chromosomal) interaction frequencies. Even under the assumption that inter-chromosomal interactions are rare and their contribution to the overall interactions of a respective chromosome is almost negligible, the presence of two chromosomal core regions increases intra-core interactions by factor two over intra-chromosomal interactions involving the subtelomere (see Figure 5.1 d). Thus, assuming that mostly *cis* interactions are affected by changes in ploidy, interaction values obtained for intra-chromosomal core interactions were multiplied by 0.5 in the interaction matrix (Figure 5.1 b). These ploidy-normalized matrices were subsequently subjected to matrix balancing by iterative correction and eigenvector decomposition (ICEing) [125] using *Hi-C Pro* [240].

5.1.2.3 *Development of computational tools to analyze interaction matrices*

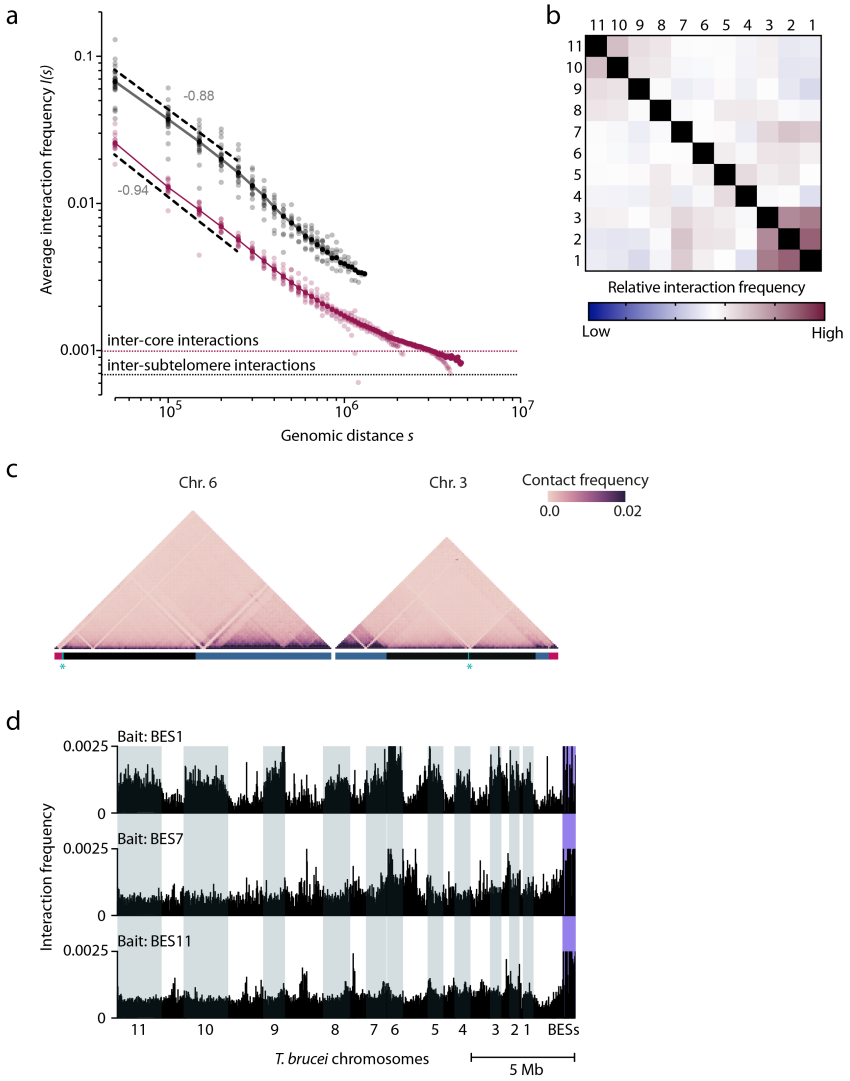
To extract information from the interaction matrices obtained, we developed a computational pipeline that shall be described in brief but can be found in detail in the methods section of this thesis (8.3). In summary, the pipeline comprises an analysis of the distance-dependent folding behavior of the DNA, the (differential) visualization of interaction matrices and submatrices, the extraction of 4C-like interaction profiles and the interrogation of point-to-point interactions in *trans* across chromosomes.

5.2 Hi-C in *T. brucei* reveals a structured nucleus with DNA conformations functionally linked to antigenic variation

5.2.1 *T. brucei chromosome folding behavior is consistent with a fractal globule*

To analyze the folding behavior of the *T. brucei* chromosomes, I investigated the relationship between the change in interaction frequencies across increasing genomic distances. Therefore, the average interaction frequencies *I(s)* for subtelomeric regions and core regions were plotted chromosome arm-wise as a function of genomic distance *s*. I calculated a linear regression of the log-log values to estimate the slope of the distance-dependent decay as a proxy for the folding behavior of the DNA polymer [156] (Figure 5.1 a). Both, the subtelomeric regions and the core regions exhibited a power-law scaling of ~-1 for distances *s* smaller than 0.25 Mb (subtelomeric regions: -0.88, core regions: -0.93, dashed lines). This folding conformation was consistent with a polymer-like behavior of a fractal globule, rather than a freely diffusing polymer at equilibrium which exhibits a log-linear fit of ~-1.5 [99][126].

Figure 5.1: Chromosomal conformation of in *T. brucei* bloodstream form nuclei.



a Distance-dependent decay of interaction frequencies. For each chromosome, intra-chromosomal interaction frequencies (interaction data from three wild-type replicates merged) were plotted as a function of genomic distance (50 kb step size) for core regions (red) and subtelomeric regions (black), respectively. Average inter-chromosomal interactions were calculated from a 50 kb interaction matrix. To determine the power-law scaling of the distance-dependent decay, a linear regression across the log10 values of interaction frequencies for distances below 0.25 Mb was calculated. **b** Preferential inter-chromosomal interactions. Shown are the relative average interaction frequencies among the core regions of the 11 megabase chromosomes (interaction data from three wild-type replicates merged). **c** Intra-chromosomal interactions across chromosomes 6 and 3 (20 kb bin size) are depicted as heat maps. Core regions in black, subtelomeric regions in blue, expression sites in red. Centromeres are demarked by blue asterisks. **d** 4C-like interaction profiles of expression sites. Genome-wide interactions of BES1 (active), BES7 and BES11 (inactive) are shown as 4C-like profiles (20 kb bin size) across all megabase chromosomes. Core regions are shaded in grey, according subtelomeres are with white background. Expression sites are shaded in blue.

5.2.2 *The T. brucei chromosomes occupy distinct chromosome territories*

A territorial arrangement of interphase chromosomes has been described for all eukaryotes for which the nuclear architecture has been studied to date [51]. For the megabase-chromosomes of *T. brucei*, I investigated the presence of chromosome territories by calculating the average interaction frequencies for all pairs of loci inside a chromosome (intra-chromosomal interactions) and comparing these to the average interaction frequencies with other megabase-chromosomes of *T. brucei* (inter-chromosomal interactions): Up to distances of 3 Mb, intra-core interactions (red line with circles) were favored over inter-core interactions between different chromosomes (dashed red line), pointing to a high degree of chromosome territoriality [156]. Only at distances greater than 3 Mb, the average intra-core interaction frequency (red line) dropped below the average inter-chromosomal interaction frequency (I_c=0.00066) (red dashed line), indicative of intermingling between chromosome territories (Figure 5.1 a).
Hi-C experiments in a human lymphoblastoid cell line have shown, that small, gene-rich chromosomes preferentially associate in three dimensions [156]. To test preferences for interactions between megabase chromosomes in *T. brucei*, the average inter-chromosomal interaction frequency between any two chromosomal core regions was calculated from a *T. brucei* wild-type interaction matrix binned to 50 kb. The heat map (Figure 5.1 b) showed substantially higher interactions between smaller chromosomes (Chr. 1-3) and – to a lesser degree – also between the largest megabase-chromosomes of *T. brucei*. These results suggest a territorial and non-random organization of megabase chromosomes in *T. brucei*.

5.2.3 *Subtelomeres and chromosomal core regions are differentially organized in T. brucei*

5.2.3.1 *Subtelomeres build a compacted domain inside each chromosome with elevated levels of intra-chromosomal interactions*

Analysis of the DNA-DNA interactions frequencies across the length of a chromosome arm revealed a slower decay of intra-chromosomal interactions below 0.25 Mb for the subtelomeres as compared to the core region (Figure 5.1 a), which points to a more tightly compacted chromatin at subtelomeres for short distances. Additionally, subtelomeres exposed higher interaction frequencies than core regions at any given genomic distance, suggesting a higher degree of compaction of the underlying DNA (Figure 5.1 a). This became also apparent when looking at intra-chromosomal interaction heat maps, where interaction frequencies were higher for intra-subtelomeric regions compared to intra-core regions (Figure 5.1 c and supplementary Figure a.10).
Further, the two chromosome arms appeared spatially separated at

the site of the centromeric repeats (blue asterisk in Figure 5.1 c) which seemed to insulate interactions between the regions up- and downstream.
However, a finer sub-compartmentalization of chromatin into smaller structures such as topologically associated domains (TADs) was not apparent from the interaction heat maps for *T. brucei* at 10 kb resolution. These findings suggested the presence of large chromosomal domains spatially dividing i) the two chromosome arms and ii) the chromosome itself into the core housekeeping portion and the *VSG*-containing subtelomeres which exhibited differential degrees of chromatin compaction.

5.2.3.2 *Subtelomeres and core regions represent different nuclear compartments*

To test whether subtelomeric regions not only form regions of higher interaction frequencies within each chromosome, but also engage in three dimensions, I analyzed the genome-wide interaction profile of the active expression site BES1 in comparison to a variety of silent expression sites. This analysis was based on the assumption, that silent expression sites would – as part of the heterochromatin fraction of the nucleus – engage frequently with repressed subtelomeric regions. In contrast, the active expression site, which is highly transcribed and described to be associated with a dedicated nuclear body, the ESB, should rather be located in the nuclear center [199].
A 4C-like interaction profile of BES1, as well as BES7 and BES11 as representatives of silent expression sites across the whole length of the *T. brucei* genome showed distinct interaction profiles for the active and silent expression sites, respectively (Figure 5.1 d): The 4C profile of the active expression site showed high interaction frequencies with each respective chromosomal core region (shaded in grey) and was depleted for interactions with the subtelomeric regions (white background). In contrast, for BES7 and BES11, inter-chromosomal interactions were more evenly distributed between core and subtelomeric regions with spurious peaks of high interactions with subtelomeres.
The subtelomeric regions and silenced expression sites therefore seemed to be located in a nuclear compartment that exhibited a different DNA interaction and folding behavior as compared to the transcriptionally active nuclear center. However, as displayed in Figure 5.1 a, the average inter-subtelomere interaction frequencies are low, thus indicating that the high degree of compaction prevents different subtelomeres from intermingling in *trans* and leads to the formation of compacted entities on their own.

5.2.3.3 *Genomic compartments correlate with transcriptional activity*

As mentioned above, our genome assembly revealed for each homologous chromosome distinct homozygous core parts containing the housekeeping portion of genes and heterozygous subtelomeres and expression sites, containing (pseudo-)*VSG*s, *ESAG*s and retrotransposable elements [194]. Hi-C revealed a differential compaction for these two chromosomal compartments.
Chromatin conformation has been shown to be linked with transcriptional activity [242]. To investigate the relation between chromatin compaction and transcriptional activity across the *T. brucei* megabase chromosomes, I performed a genome-wide analysis of the Lister 427 wild-type transcriptome including non-polyadenylated transcripts. Whole genome RNAseq revealed a transcriptional partitioning of the genome that coincides with the haploid-diploid boundaries of the chromosomes: Whereas the chromosomal core was transcribed throughout, the subtelomeres and expression sites – with exception of the active expression site BES1 on chromosome 6 – appeared transcriptionally repressed (see). This transcriptional compartmentalization of the linear genome anti-correlated with the degree of compaction of core regions and subtelomeres as measured by Hi-C.

5.2.4 *Inter-chromosomal interactions in the T. brucei bloodstream form genome*

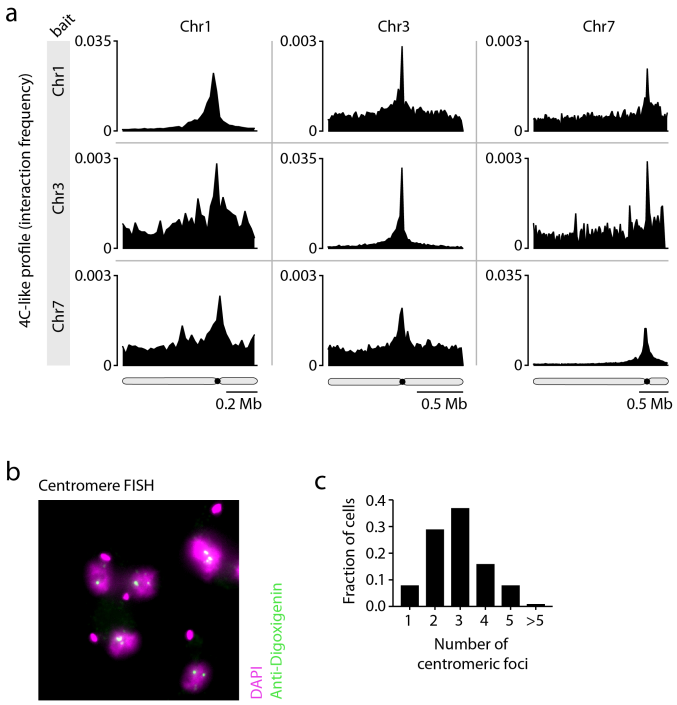
5.2.4.1 *Centromeres interact in three dimensions*

Hi-C revealed elevated inter-chromosomal interaction frequencies between centromeric regions as compared to a randomized background sample that mirrored the query in size and distribution (1.32-fold, p= 0.0002) in *T. brucei*. A 4C-like interaction profile of centromeres from chromosomes 1, 3 and 7 as bait showed a clear association between centromeres of different chromosomes with a smooth distance-dependent decay of interaction frequencies towards the chromosome endings (Figure 5.2 a and Figure 5.3 a). These interactions were further supported by performing FISH against the centromeric repeats using a mixture of nine labeled oligonucleotides (Figure 5.2 b). Whilst one would expect a maximum signal of 28 spots originating from totally dispersed centromeric regions, analysis of the acquired images revealed on average three centromeric clusters per nucleus (Figure 5.2 c).

5.2.4.2 *Transcription is compartmentalized in T. brucei*

In higher eukaryotes, transcriptional hubs ('transcription factories') containing active RNA polymerases and their associated transcription factors can act themselves as architectural organizers [48].
To test for co-localization of genes transcribed by RNA polymerase

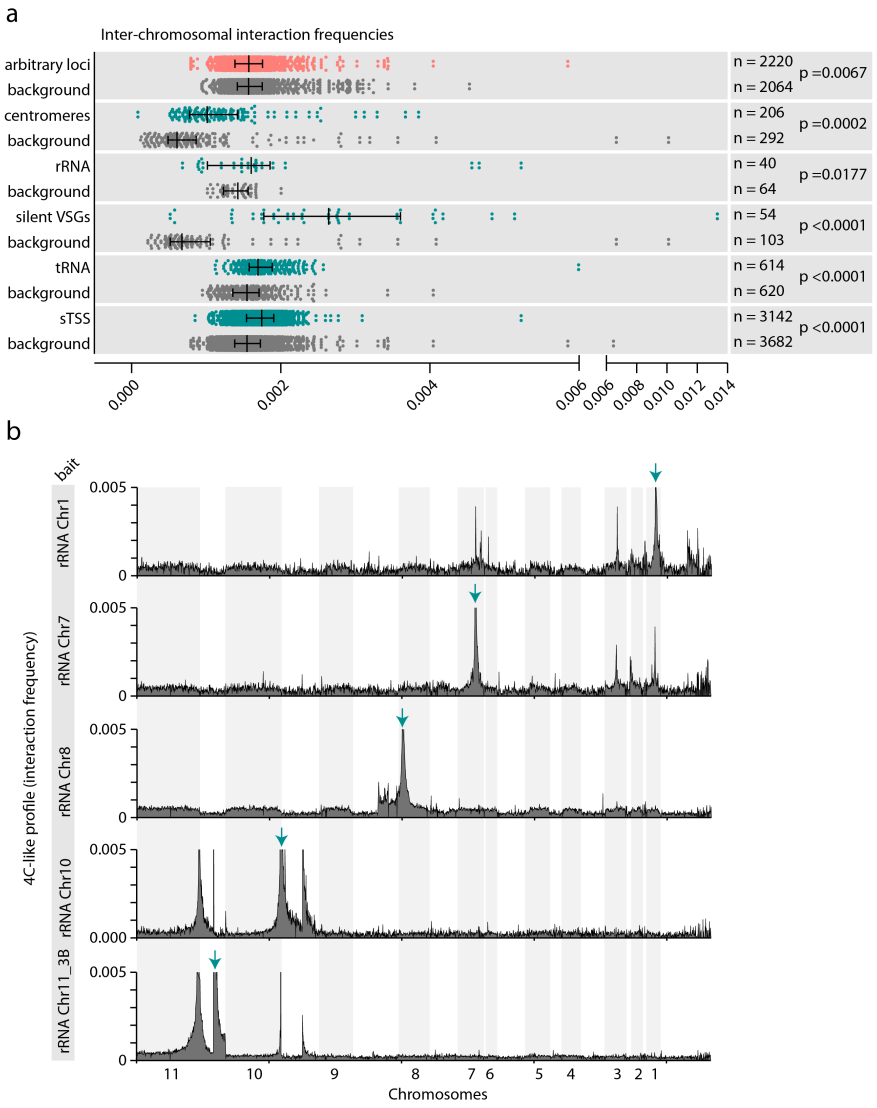
Figure 5.2: Centromeres interact in *trans* in *T. brucei*



a Matrix of 4C-like interaction profiles (20 kb bin size) of centromeres located on chromosomes 1, 3 and 7 across the core regions of the same chromosomes. Chromosomal location is demarked in black on a schematic chromosome below. **b** Centromere fluorescence *in situ* hybridization (FISH). DNA is stained with DAPI (magenta), centromere-probes are digoxigenin-labelled and detected with an Alexa-488-conjugated anti-digoxigenin antibody (green). **c** Quantification of centromere FISH signal using Imaris 8 to determine fraction of cells with 1, 2, 3, 4, 5 or more centromeric foci. Analysis is based on n=167 wild-type nuclei.

I, II or III, I tested rRNA, tRNA and RNA Pol II transcription start sites for colocalization in *trans* by comparing interaction frequencies with background interaction models (8.3.2.3) and by generation and examination of 4C-like profiles (8.3.2.2).
In *T. brucei*, rRNA genes occur as arrays in the genome, often in linear proximity to a centromere [4]. Testing the inter-chromosomal colocalization of these *rRNA* arrays by Hi-C showed significantly increased inter-chromosomal interactions between these loci compared to the background sample (Figure 5.3 a). A representation of inter-chromosomal interactions of individual *rRNA* loci by virtual 4C (Figure 5.3 b) revealed correlated interaction profiles between different *rRNA* gene arrays (e.g. genome-wide interaction profiles of *rRNA* genes located on Chr1 correspond to those of *rRNA* genes on Chr7; *rRNA* located on Chr10 seemed to correlate with *rRNA* genes on Chr11 3B). These correlated interactions suggested that *rRNA* genes located

Figure 5.3: Distinct *trans*-interactions between in *T. brucei* bloodstream form nuclei



a Scatter plot of inter-chromosomal interaction frequencies (50 kb bin size) among different genomic loci. Each query sample (green) is tested against a matching, randomly selected background sample (grey) mirroring the query sample in size and distribution of the tested loci. Bins with zero values were removed from the query and background sample, respectively. P-values (given on the right) are based on Welch's *t*-test (two-sided). Error bars represent mean values and standard deviation. **b** Genome-wide 4C-like interaction profiles of rRNA genes located on chromosomes 1, 7, 8, 10 and 11_3B. Core regions are shaded in grey, subtelomeres in white. The respective 'bait' rRNA gene is designated with a green arrow.

on different chromosomes assemble in multiple clusters in the nucleolar periphery. Notably, the *5S rRNA* gene located on chr8, which is transcribed by RNA Pol III [103], did not colocalize with any of the other RNA Pol I transcribed *rRNA genes* (Figure 5.3 b).
Despite the close linear association between *rRNA* genes and centromeres, *trans* interactions among *rRNA* genes and centromeres, respectively, are not necessarily interdependent. For example, whereas the centromeres of the homologous chromosomes 10 are located in the subtelomeric regions, *rRNA* genes are located in the chromosomal core regions. The subtelomeric centromeres on chr10_3A and chr11_3A interacted independently from the presence of *rRNA* genes. *Vice versa*, *rRNA* located on chr11_3B interacts independently from the presence a centromere with *rRNA* located on chr10 core (see also Figure 6.6 b and c). *rRNA* interactions in *trans* and show therefore centromere-independent interactions with other *rRNA* genes in *trans*.
The boundaries of RNA Pol II transcribed PTUs represent both RNA Pol II transcription start and termination sites and are often associated with the presence of RNA Pol III transcribed *tRNA* genes [245]. Both, the 4C-like profiles (not shown) and a colocalization test against a randomized background-model (Figure 5.3 a) showed comparably few *trans*-interactions between *tRNA* arrays (1.11-fold, p=8x10e-21). Further, unidirectional PTU borders (see section 3.4) containing both, sites of transcription start and termination, gave only very weak signal of colocalization in *trans* (1.09-fold, p=6.5x10e-91) (Figure 5.3 a).
In summary, these findings reflect the spatial assembly of ribosomal genes in the nucleolus, which represents a membrane-less, RNA Pol I rich nuclear body and reveal for the first time the presence of distinct *rRNA* gene clusters in *T. brucei*. RNA Pol II and III transcribed genes, however, showed only few associations across chromosomes, which is in agreement with the presence of chromosome territories. This, however, does not preclude looping of PTU borders in *cis*, which remains to be determined.

5.2.4.3 *The actively transcribed VSG-2 strongly interacts in trans with the spliced-leader array*

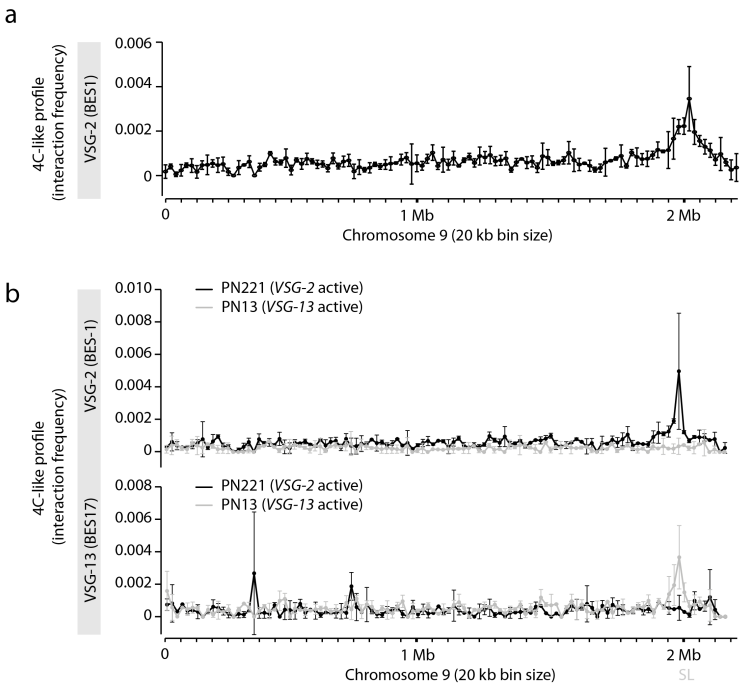
As mentioned before, the genome-wide 4C-profiles of silent expression sites suggested a location away from the compartment of transcribed chromosomal core regions. To test if this implied a co-localization of silent expression sites, I measured interactions among silent *VSG* expression sites and compared them to a fitted background model of randomly picked interactions. This revealed a pronounced spatial association of different silent expression sites (2.52-fold, p<0.0001); Figure 5.3 a).
The actively transcribed expression site had previously been described to be located in a dedicated extranucleolar RNA polymerase I enriched focus, the ESB [199]. To assess interactions of the active expression site

BES1, I generated a 4C-like profile from a 20-kb Hi-C matrix, using the actively transcribed *VSG-2* gene as a bait. As depicted in Figure 5.3 a, the active expression site showed a single distinct and specific *trans*-interaction with a genomic locus close to the 3′ end of chr9, which contains the spliced-leader repeat sequence. This *trans*-interaction could not be observed for any other of the silent BESs.
To test whether the expression site – spliced-leader interaction was exclusive for the active expression site and functionally linked to antigenic variation, I repeated the Hi-C experiment in two previously published cell lines PN221 and PN13 [83] (Table 7.3). In PN221 the actively transcribed expression site BES1 (containing *VSG-2*) and the silent expression site BES17 (containing *VSG-13*) were both tagged with a resistance marker downstream of the ES promoter. By changing antibiotic selection, cells that stochastically switched towards transcription of BES17 were selected and named PN13. 4C-like profiles using *VSG-2* and *VSG-13* as bait revealed that whereas in PN221 cells *VSG-2* interacted with the spliced-leader, this interaction was lost in the PN13 cells (Figure 5.4 b). Instead, the in PN13 newly activated *VSG-13* showed high interaction with the spliced-leader locus on chr9 (Figure 5.4 c).
Together, these results indicated a spatial association of the silenced ESs and confirmed a privileged localization of the active expression site and described for the first time a direct interaction of the active expression site with another DNA-element in *trans*.

5.3 Developmental folding of the *T. brucei* genome across bloodstream form and procyclic life-cycle stages

An analysis of intra-chromosomal interaction frequencies in procyclic *T. brucei* nuclei, revealed overall higher interaction frequencies inside subtelomeric regions (Figure 5.5 a, black line) as compared to core regions of each chromosome (Figure 5.5 a, red line). This was also apparent from intra-chromosomal interaction heat maps (Figure 5.5 b). The distance-dependent decay of interaction frequencies of core and subtelomeric regions in procyclic cells exhibited a power-law scaling of -0.95 and -0.82, respectively (Figure 5.5 a, dashed lines). This corresponded to the scaling observed in a bloodstream form nucleus (-0.93 and -0.88), with a slightly slower decay of interaction frequencies for subtelomeric regions.
In the process of differentiation from bloodstream to procyclic form *T. brucei*, transcription of *VSGs* is halted and as the ESB dissociates, the previously active expression site is repositioned from the nuclear center towards the heterochromatic nuclear periphery [144]. This developmental silencing of *VSG-2* was mirrored in genome-wide 4C-like profiles using *VSG-2* as virtual bait (Figure 5.5 c). Whereas in bloodstream form parasites, the expected wave-like pattern of high

Figure 5.4: *Trans*-interactions of the active *VSG* in bloodstream form nuclei

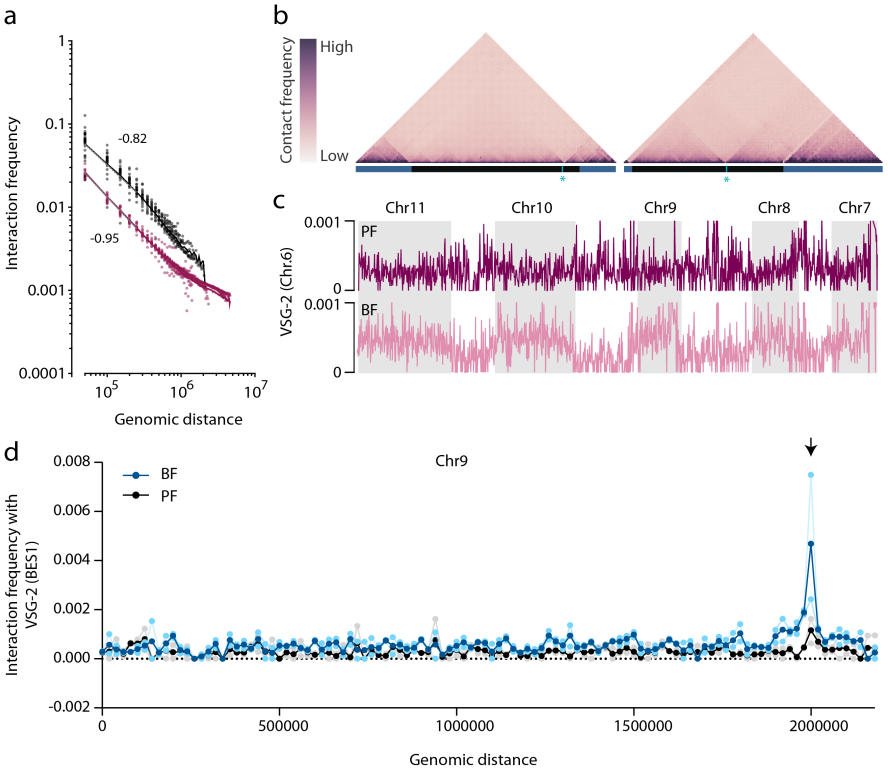


a 4C-like profile of interaction frequencies (20 kb bin size) across chromosome 9 using the active *VSG*-2 (BES1) as bait. Plotted is the mean and standard deviation between three replicates. **b** 4C-like profile of interaction frequencies across chromosome 9 (20 kb bin size) using *VSG*-2 as bait, which is active in PN221 (black) and inactive in PN12 (grey). **c** 4C-like profile of interaction frequencies across chromosome 9 (20 kb bin size) using *VSG-13* as bait, which is active PN12 (grey) and inactive in PN221 (black). In c and d, the mean and standard deviation between two replicates are indicated.

interactions with chromosomal core regions (shaded in grey) and low interaction with the silenced subtelomeres became apparent (Figure 5.5 c, lower panel), in procyclic form parasites this pattern was lost. Here, *VSG*-2 interacted mostly uniformly with both subtelomeres and core regions and partially showed higher interaction frequencies with subtelomeric regions (Figure 5.5 c, upper panel). In line with these observations, the strong interaction between the active expression site and the spliced-leader locus located on chr9 that was found for the bloodstream form cells, was strongly diminished in procyclic nuclei (Figure 5.5 d).

During differentiation to procyclic form parasites, the *VSG* cell surface coat is replaced by a dense coat of invariant GPEET or EP procyclins (also called procyclic acidic repetitive proteins (PARPs)) [229]. Genes belonging to the multigene families of *procyclins* and *procyclin-associated genes* (PAGs) are encoded group-wise in RNA Pol I

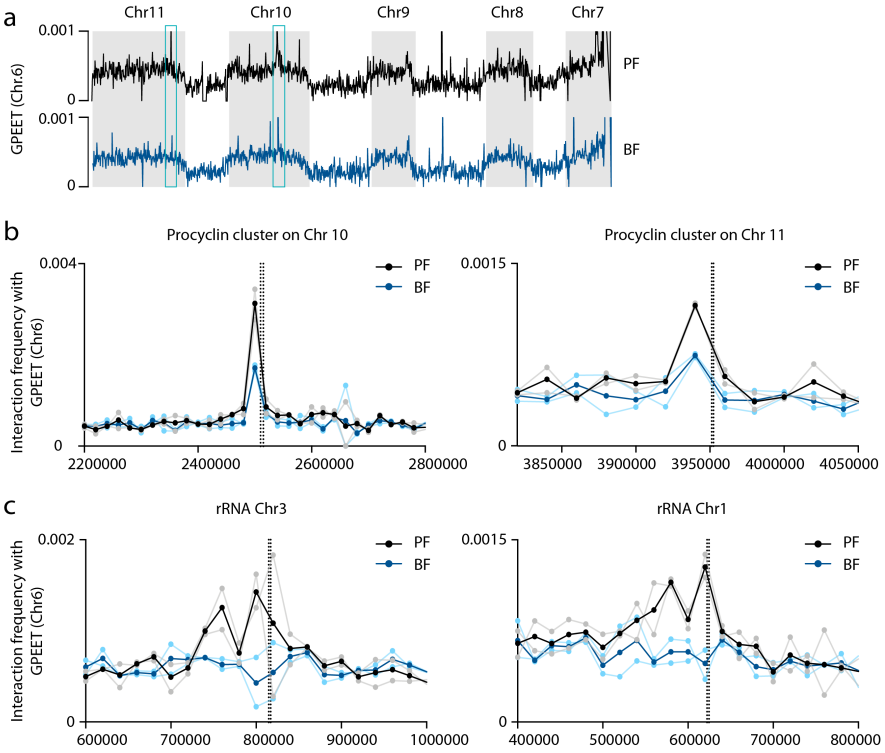
Figure 5.5: Spatial DNA configurations in procyclic nuclei



a Distance-dependent decay of interaction frequencies. For each chromosome, intra-chromosomal interaction frequencies (interaction data from two procyclic wild-type replicates merged) were plotted as a function of genomic distance (50 kb step size) for core regions (red) and subtelomeric regions (black), respectively. Average inter-chromosomal interactions were calculated from a 50 kb interaction matrix. A linear regression of interaction frequencies for distances below 0.25 Mb was calculated from the log10 values of interaction frequencies using.**b** Intra-chromosomal interactions across chromosomes 8 and 4 (20 kb bin size) depicted as heat maps. Core regions are demarked in black, subtelomeric regions in blue, expression sites in red. Centromeres are indicated by blue asterisks. **c** 4C-like profile of interaction frequencies across chromosome 9 (20 kb bin size) using *VSG-2* as bait in bloodstream form (dark blue) and procyclic nuclei (black). Shown are the mean values from two replicates (light blue and grey, respectively).

polycistronic transcription units located on chromosomes 6, 10 [59] and 11 [270]. Previous publications have found *procyclins* to be transcribed by RNA Pol I inside the nucleolus [144]. I generated genome-wide 4C-like interaction profiles of a *procyclin* array under the control of a *GPEET* promoter located on chr6. The 4C-like interaction profile recapitulated the wave-like interaction pattern in both procyclin and bloodstream form nuclei (Figure 5.6 a). These findings indicated that the *procyclin* locus does not relocate entirely towards the nuclear periphery in bloodstream form parasites. Furthermore, the *procyclin* cluster located on chr6 appeared to engage with other *procyclin* genes in *trans* in both life-cycle stages, however, to a lesser extent in bloodstream form parasites (Figure 5.6 b).
The *procyclin* locus also revealed high *trans*-interaction frequencies with *rRNA* genes located on chr3 and chr1 in procyclic form parasites, recapitulating the spatial localization in the nucleolar compartment (Figure 5.6 c, black and grey lines). In bloodstream form parasites, this interaction was lost (Figure 5.6 c, blue lines), indicating a developmental repositioning of the procyclin genes away from the nucleolus.
These findings suggested, that the nuclear architecture of *T. brucei* was intimately and functionally linked with the parasite life-cycle stage. Specifically, the life-cycle dependent expression of *VSGs* and *procyclins* was reflected by developmental repositioning of these loci between a transcriptionally active RNA Pol I compartment (ESB and nucleolus) and a transcriptionally silent compartment. The repositioning of procyclins during developmental silencing was less clear but appeared to involve a dissociation from the nucleolus.

Figure 5.6: *Trans*-interactions of procyclin genes



a 4C-like interaction profile (20 kb bin size) of a procyclin gene array located on chromosome 6 across chromosomes 7-11 (core regions shaded in grey) in procyclic (black) and bloodstream form nuclei (blue). Shown are the interaction values from two merged replicates, respectively. Blue boxes highlight the location of two procyclin loci on chr11 and chr10. **b** 4C-like interaction profiles (20 kb bin size) of the procyclin locus on chromosome 6 across a procyclin gene array on chromosome 10 and 11 and **c** across two exemplary rRNA arrays located on chromosomes 3 and 1. Depicted are the mean values (bloodstream form: dark blue, procyclics: black) from two independent replicates (bloodstream form: light blue, proyclics: grey) at 20 kb matrix resolution. Vertical bars indicate the location of *procyclin* and *rRNA* genes, respectively.

6

Architectural proteins linking 3D genome structure and antigenic variation in *T. brucei*

6.1 ATACseq reveals differential local chromatin accessibility across the *T. brucei* genome

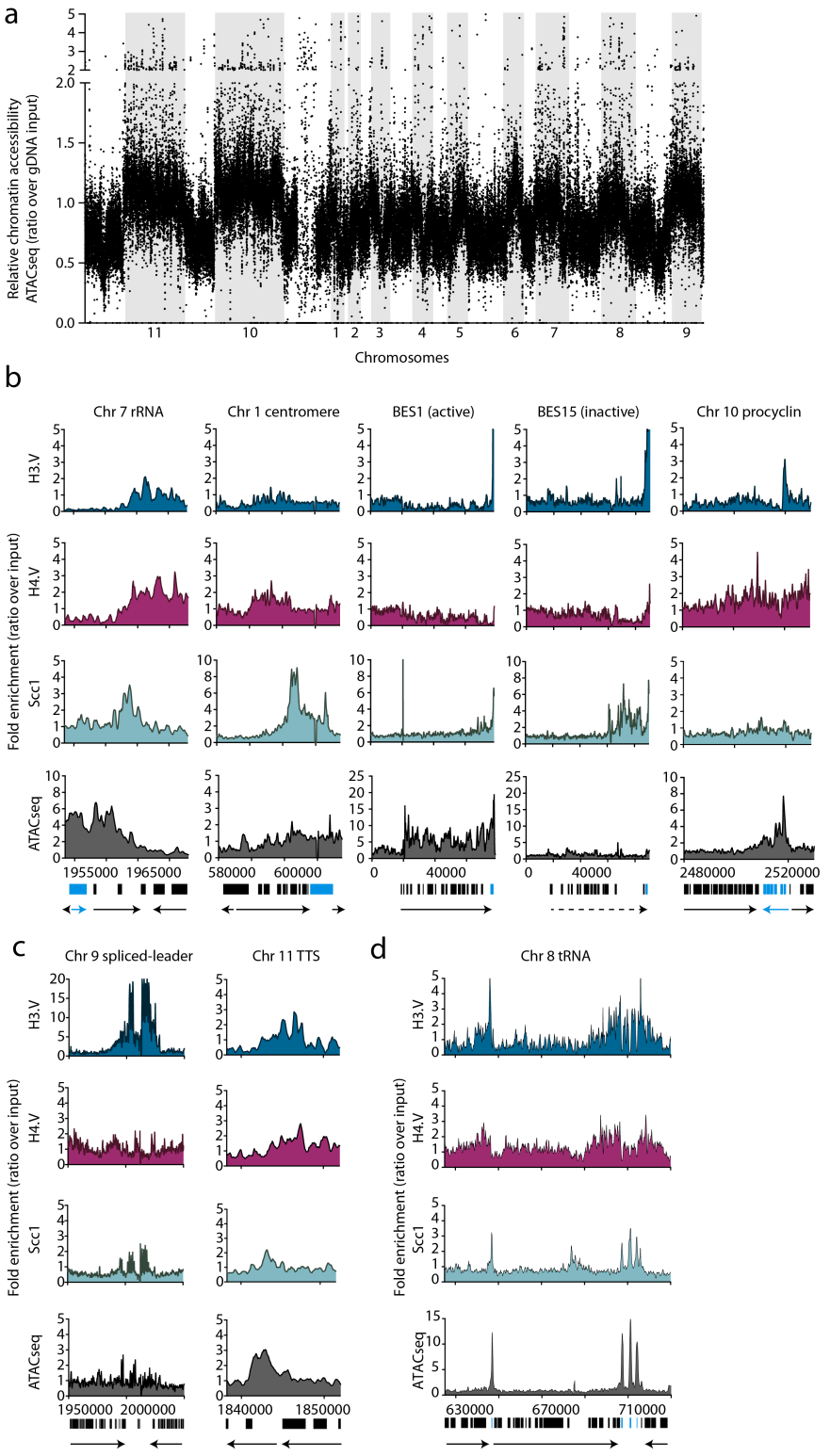
Hi-C and RNAseq generated as part of my PhD work had revealed a prominent spatial segregation of the *T. brucei* genome into transcriptionally active (e.g. chromosomal core regions and active nuclear bodies such as the nucleolus and the ESB) and silent domains (e.g. silent expression sites and the *VSG* repertoire). To test whether spatial folding and transcriptional activity were reflected in chromatin accessibility across the genome, we performed genome-wide ATACseq in *T. brucei* bloodstream-form wild-type cells. ATACseq was performed in duplicates with a gDNA control to normalize for bias intrinsic to the DNA sequence.

ATACseq revealed reduced chromatin accessibility at transcriptionally silent and presumably heterochromatic subtelomeres (Figure 6.1 a) and centromeres (Figure 6.1 b, 2nd panel) as compared to chromosomal core regions. ATACseq reflected also the heterochromatic nature of silenced expression sites with inaccessible chromatin throughout the expression site PTU (Figure 6.1 b, 4th panel). The actively transcribed expression site, in contrast, appeared highly accessible (~15-fold over gDNA control), starting from the ES promoter towards the telomeric repeats (Figure 6.1 b, 3rd panel).

Furthermore, ATACseq revealed start and termination sites of RNA Pol II transcribed PTUs to be demarked with moderately accessible chromatin (Figure 6.1 c, supplementary Figure a.13. Likewise, RNA Pol I transcribed *rRNA* and *procyclin* loci displayed increased chromatin accessibility (6.1 b). RNA Pol III transcribed *tRNA* genes in particular reveal sharp ATACseq peaks across the cohesin-rich *tRNA* gene bodies (Figure 6.1 d).

In summary, the ATACseq results pointed to a spatial partitioning of active and inactive chromatin inside the nucleus. Importantly, the increased compaction of silenced loci, e.g. at subtelomeres as shown by Hi-C was reflected by a high degree of inaccessibility as determined by ATACseq.

Figure 6.1: Cohesin is localized at sites of H3.V and H4.V enrichment.



a Relative chromatin accessibility across subtelomeric regions (white) and core regions (grey) across the megabase chromosomes. Plotted are the individual values of ATACseq in wild-type over gDNA control (window size 501 bp, step size 101 bp).

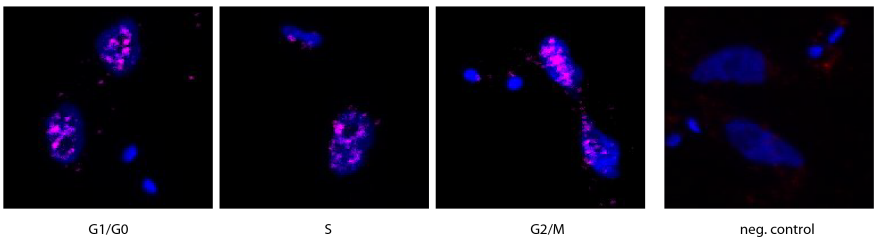
b, **c** and **d**. ChIPseq data showing the enrichment of histone variants H3.V (blue) and H4.V (red) and the cohesin subunit Scc1 (green) at representative genomic sites (window size 501 bp, step size 101 bp). ATACseq indicates relative chromatin accessibility (window size 501 bp, step size 101 bp). Protein coding genes are represented as black boxes underneath. The location of *rRNA* genes, *tRNA* genes, centromeres, *VSG* genes and *procyclins* are indicated as blue boxes. Arrows indicate direction of transcription.

6.2 Characterization of cohesin binding sites in the *T. brucei* genome

Beyond its role in sister chromatid cohesion during mitosis, an additional architectural function of cohesin is well-conserved among eukaryotes.
To study the genomic distribution of the cohesin complex in *T. brucei*, I generated a transgenic *T. brucei* cell line with one allele of the cohesin subunit Scc1 N-terminally tagged with a Ty1-tag. The second Scc1 allele was deleted (Table 7.3, supplementary Figure a.12). ChIPseq using a monoclonal anti-Ty1 antibody [18] showed a marked enrichment of cohesin at centromeres (6.1 b). Further, Scc1-ChIPseq revealed distinct peaks of cohesin enrichment specifically at sites of RNA Pol II transcription termination (but not at divergent transcription start sites) (Figure 6.1 c) as well as at the spliced-leader locus on chromosome 9, at *rRNA* (Figure 6.1 b) and *tRNA* genes (Figure 6.1 d).
Cohesin was also enriched at expression sites with increasing enrichment over the *VSG* gene and towards the telomeric repeats (Figure 6.1 b). Cohesin did not show an overall enrichment along the subtelomeres containing the *VSG* repertoire, but was increasingly bound to DNA directly upstream of the telomeric repeats. Notably, cohesin enrichment at PTU boundaries and at sites of *rRNA* and *tRNA* genes coincided with increased chromatin accessibility as determined by ATACseq (Figure 6.1 b-d).
I performed immunofluorescence imaging of the Ty1-tag in the Ty1-Scc1 cell line, which revealed a punctuate distribution of cohesin across the nucleus in all cell-cycle stages (Figure 6.1). This finding reflected the deposition at distinct genomic sites as revealed by ChIPseq. Further, its presence throughout the cell cycle suggested an additional role of cohesin in establishing nuclear architecture beyond sister chromatid cohesion in S and G2 phase in *T. brucei*.

Since the above-mentioned distribution of cohesin reminded us of the distribution of histone variants H3.V and H4.V [245], I performed MNase-ChIPseq experiments in two transgenic cells lines each expressing a respective histone variant with an N-terminal Ty1-tag (Table 7.3, supplementary Figure a.12). MNase-ChIPseq revealed that at the aforementioned sites – with the exception of megabase centromeres –

Figure 6.1: Cohesin immunofluorescence imaging



Immunofluorescence against the Ty1-tagged cohesin subunit Scc1 (magenta) across the cell cycle. DNA (blue) is stained with DAPI. 27-39 stacks per image 0.3 μm stack height) Shown is the maximum projection. A wild-type cell line was treated in the same way and used as negative control.

cohesin enrichment across the genome was in most cases associated with the presence of histone H3 variant (H3.V) and H4 variant (H4.V) (Figure 6.1 b-d, Figure a.13).
At convergent transcription termination sites, enrichment of cohesin was congruent with that of H3.V. However, at directional termination sites, cohesin appeared juxtaposed to and directly downstream of histone variant deposition (in direction of transcription).
At sites of *rRNA* and particularly at sites of *tRNA* genes, the cohesin signal was located directly at the coding DNA sequence (CDS) and was flanked by a strong enrichment of H3.V and H4.V directly up- and downstream. The *tRNA* and *rRNA* CDSs themselves were depleted of H3.V and H4.V.
At subtelomeres and both at the active and silent *VSG* expressions sites, the cohesin signal increased towards the telomeric repeats, was accompanied by an increase in H3.V and – to a lesser extent – by H4.V enrichment. The actively transcribed expression site showed reduced nucleosome occupancy from the promoter all the way to the telomere, as had been described previously [82].
In summary, these results indicated that distinct genomic loci interacting in three dimensions across the *T. brucei* nucleus were demarked by cohesin and that localization of cohesin at these sites coincided with the deposition of histone variants H3.V and H4.V.

6.3 Simultaneous knockout of histone variants H3.V and H4.V results in perturbed chromosomal architecture and increased *VSG* switching

6.3.1 *Generation of a ΔH3.VΔH4.V knockout cell line*

The CCCTC-binding factor CTCF has been described to function in cohesin recruitment to non-centromeric loci [277]. While loss of CTCF

has been described to confound contact domains predominantly at the submegabase-scale [232] [166] [211], loss of cohesin strengthens compartment boundaries and colocalization of super-enhancers [223]. Both proteins are, however, essential for cell proliferation and development, which hampers their genetic manipulation or depletion [203].
CTCF is absent in non-metazoans such as *T. brucei*. Direct depletion of cohesin in *T. brucei* has been shown to strongly affect cell cycle progression due to its central role in sister chromatid cohesion during mitosis and is thus lethal to the parasite. Since my ChIPseq data suggested that cohesin distribution was reminiscent of histone variant H3.V and H4.V distribution, I sought to study the effect of H3.V and H4.V on chromatin architecture in *T. brucei*.
Cell lines with one of the two histone variant forms deleted (ΔH3.V and ΔH4.V) had been published before and no effect on cell viability had been observed [162] [245]. As both histone variants colocalized at sites of cohesin enrichment, which suggested a functional redundancy, I additionally generated a ΔH3.VΔH4.V 'double'-knockout cell line (Table 7.3) to study the effect of histone variant deletion on chromatin architecture and transcription. As for the single deletion strains, the resulting double deletion cell-line did not show aberrations in cell-cycle progression or growth (supplementary Figure a.11).

6.3.2 *VSG expression is strongly deregulated in ΔH3.VΔH4.V cells*

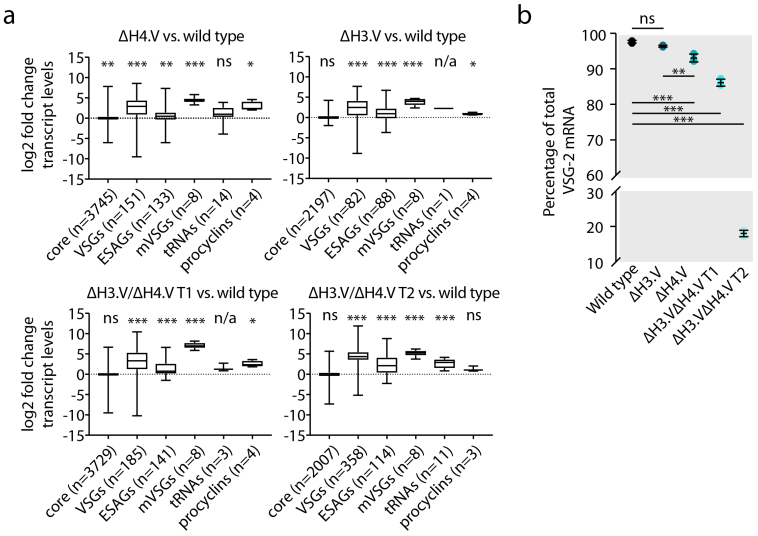
Cohesin depletion in *T. brucei* has been shown to deregulate mutually exclusive expression of *VSGs* while affecting proper division of the ESB [199].
To investigate a possible effect of *H3.V* and *H4.V* deletion on *VSG* expression, I performed RNAseq including non-polyadenylated transcripts in triplicates in wild-type (single-marker), ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells (Table 7.3).
Whole transcriptome analysis with DESeq2 [161] revealed a broad transcriptional deregulation of *VSG* expression in the single (ΔH3.V and ΔH4.V) deletion cell line, which was even more pronounced in the double deletion ΔH3.VΔH4.V cell line as compared to wild-type cells (Figure 6.2, supplementary Figure a.14). Examination of the fold-change transcript levels with an adjusted p-value (p_{adj})<0.1 (Figure 6.2 a) showed that in both single deletion mutants, predominantly *VSG* genes located in bloodstream form ES and subtelomeres (3.65-fold, p_{adj}<0.0001 in ΔH3.V cells and 3.68-fold, p_{adj}<0.0001 in ΔH4.V) as well as metacyclic *VSGs* (8.57-fold, p_{adj}<0.0001 in ΔH3.V cells and 13.22-fold, p_{adj}<0.0001 in ΔH4.V) were strongly upregulated. Expression of *VSG-2* was only mildly altered in ΔH3.V cells (96.41% of all *VSG* mRNAs) and ΔH4.V cells (93.06% of all *VSG* mRNAs) as compared to wild-type cells (97.58% of all *VSG* mRNAs) (Figure 6.2 b). Expression site associated genes (ESAGs) were only mildly, but significantly upregulated in

ΔH3.V (2.19-fold, p_{adj}<0.0001) and ΔH4.V cells (1.41-fold, p_{adj}=0.0007) (Figure 6.2 a). tRNA transcripts were upregulated, however, not significantly (4.7-fold, p_{adj} n/a in ΔH3.V cells and 1.78-fold in ΔH4.V cells). Procyclin (associated) gene transcript levels were upregulated in the single deletion mutants (1.79-fold, p_{adj}=0.01 in ΔH3.V cells and 7.21-fold, p_{adj}=0.017 in ΔH4.V). The housekeeping portion of genes located in the chromosomal core showed tendencies towards both up- and downregulation and was on average not altered in comparison to the wild-type strain (0.99-fold, p_{adj}=0.19 in ΔH3.V cells and 1.02-fold, p_{adj}=0.001 in ΔH4.V) (Figure 6.2 a).
The same phenotype was observed in a more pronounce manner in the ΔH3.VΔH4.V double deletion mutants. Here, expression of *VSG-2* was reduced to 86% of total *VSG* mRNA (Figure 6.2 b).
A repeated RNAseq experiment using the same clone with additional ~three weeks in culture (T3, supplementary Figure a.14 a) showed *VSG-2* expression further reduced to 18% of total *VSG* mRNA (Figure 6.2 b). *ESAGs* and *tRNA* genes were also more clearly upregulated compared to the first timepoint (T2) (Figure 6.2 a). Again, transcription of housekeeping genes was not altered in the double deletion mutants at both time points.
In summary, the RNAseq experiments indicated a role of histone variants H3.V and H4.V in *VSG* transcription regulation. Thereby, H3.V and H4.V possibly acted in a synergistic or additive manner, as *VSG* derepression was enforced in the double deletion mutant. The strong decrease of *VSG-2* expression in the double deletion mutant, that could not be observed in the single deletion mutants, was indicative of a switching event.

6.3.2.1 *Immunofluorescence microscopy and FACS confirmed exchange of VSG-2 at the surface of ΔH3.VΔH4.V cells*

To investigate the possibility that a *VSG* switching event had taken place in the histone variant deletion cells, I performed immunofluorescence microscopy in wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells (~1 week in culture (T2, supplementary Figure a.14 a)). Specifically, I used the monoclonal TAT1 antibody recognizing trypanosomal α-tubulin (magenta) [283] to stain the cytoplasm and subsequently investigated the cells for loss of *VSG-2* expression on the cell surface using a polyclonal α-VSG-2 antibody (green) [83], e.g. cells solely labelled for α-tubulin represented *VSG*-switchers (Figure 6.3 a). Switching frequency was determined independently by two blinded persons, which evaluated cells for loss of the VSG-signal in the microscopy images. Whereas no increase in switching frequency was observed for ΔH3.V (0.3%, n=329) and ΔH4.V (0.4%, n=363) cells in comparison to the wild type (2.6%, n=135), ΔH3.VΔH4.V cells cultured for ~1 week (T2) showed loss of *VSG-2* expression in 40.6% of the cells (n=186).
These findings were recapitulated in a FACS flow cytometry experi-

Figure 6.2: RNAseq in ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells



a RNAseq of ΔH3.V, ΔH4.V and ΔH3.VΔH4.V (T1 and T2). Shown is the log2 fold-change for significantly regulated transcripts (p>0.1, three replicates for each condition) in comparison to wild type. Boxplots indicate median, 25% and 75% percentile, whiskers show minimum and maximum values. Deviation from 0 (no differential gene regulation) was determined using one Sample t test (ns = not significant, *** p<0.0001, ** p<0.002, * p<0.02). **b** Change in *VSG*-2 expression in ΔH3.V, ΔH4.V and ΔH3.V/ΔH4.V (T1 and T2) cells as determined by RNAseq. Shown are the percentages of *VSG*-2 mRNA levels of total *VSG* mRNA in three RNAseq replicates. Significance was determined using one-way anova (ns = not significant, **** p<0.0001, ** p=0.0015).

ment. Here, cells were first incubated with the same α-VSG-2 antibody following incubation with a labeled secondary antibody. Subsequently, 100,000 gated cells were analyzed for presence or loss of VSG-2 on the cell surface using a FACSort flow cytometer. Quantification of VSG-2 negative and positive cells revealed a switch in *VSG* expression in 51.28% of ΔH3.VΔH4.V cells and no increase in *VSG* switching in the single deletion mutants compared to the wild type (Figure 6.3 b).

In conclusion, immunofluorescence and flow cytometry analyses pointed to an increased *VSG* switching frequency upon simultaneous deletion of histone variants *H3.V* and *H4.V*. These experiments also showed that deletion of one of the two histone variants alone was not sufficient to trigger *VSG* switching. As the cells could only be tested for loss of *VSG-2* expression, it remained unclear what *VSGs* were activated and how hetero- or homogenous *VSG* activation was at the single-cell level.

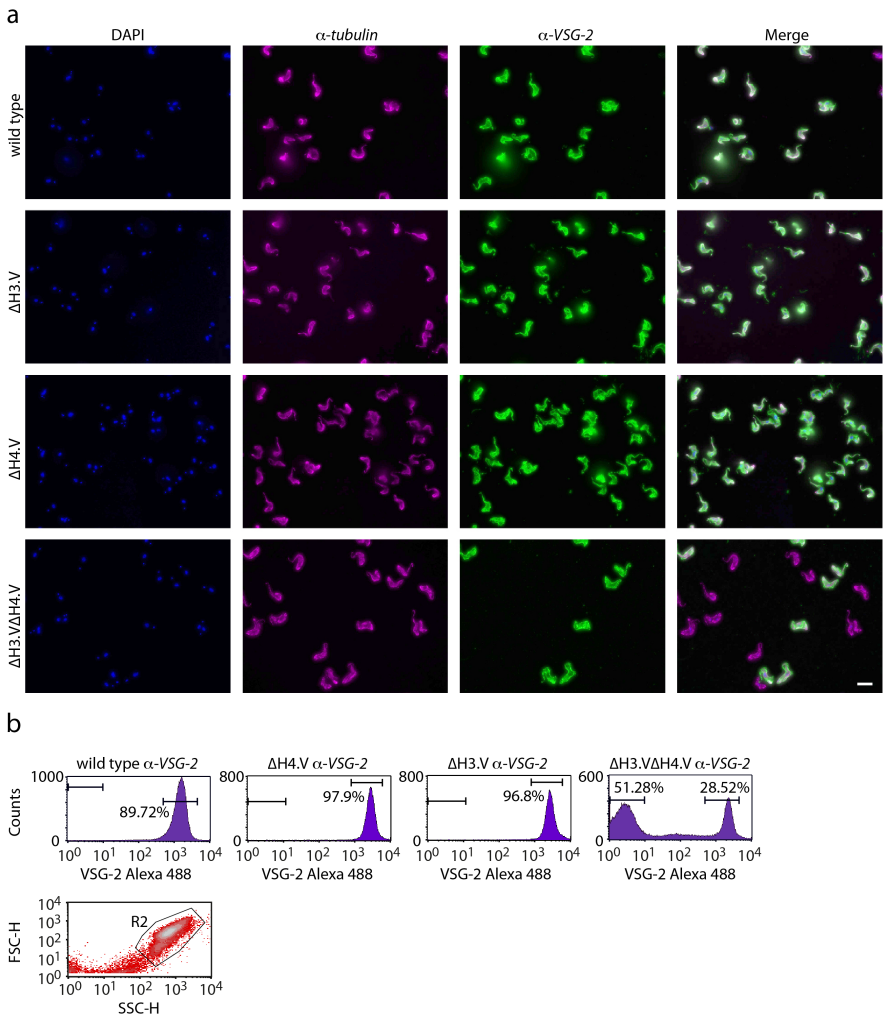
6.3.3 *Single-cell RNAseq revealed increased VSG switching upon deletion of H3.V and H4.V*

Total RNAseq represents the average transcriptome profile over a population of cells. Therefore, increased heterogeneity in *VSG* transcripts can either result from increased switching at the population level or loss of mutually exclusive expression at the single cell level (multiple *VGSs* are expressed in a single cell) (Figure 6.4 a). To discern between these two possibilities, we performed single-cell RNA sequencing (scRNAseq) in wild-type and in ΔH3.VΔH4.V cells and analyzed *VSG* expression in individual *T. brucei* cells by mapping the sequencing reads against the previously published *VSG*nome [56].

While the analysis of 40 wild-type cells showed a homogenous expression of *VSG-2*, data from 23 out of 44 ΔH3.VΔH4.V cells (T1, supplementary Figure a.14 a) revealed activation of new *VSGs* and a concomitant loss or reduction of *VSG-2* mRNA. Furthermore, *VSG* activation did not occur at random but appeared biased towards the activation of *VSG-11*, *VSG-8* and the metacyclic VSG *mVSG-531* (Figure 6.4 b). *VSG* activation was not mutually exclusive and a combination of up to four different *VSGs* expressed per single cell could be observed. However, expression of additional *VSGs* (*VSG-6*, *VSG-15*, *VSG-17*, *VSG-14*, *mVSG-397*, *mVSG-653*, *mVSG-631* and *mVSG-559*) was only observed in combination with *VSG-2*, *VSG-11* or *VSG-531* in this dataset.

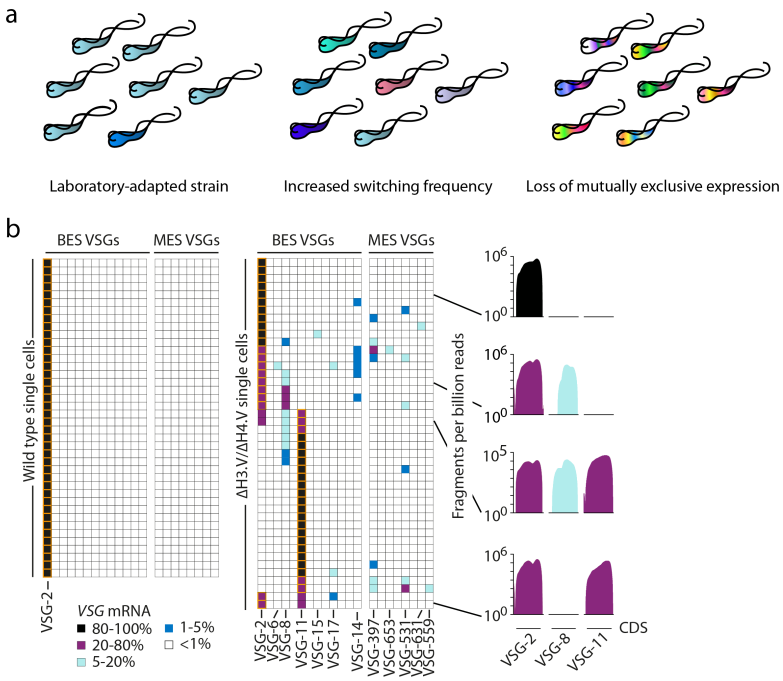
A second scRNAseq experiment with additional 334 single ΔH3.VΔH4.V cells (~50 population doublings apart (T3, supplementary Figure a.14 a)) showed increased deactivation of *VSG-2* (~20% of total *VSG-2* mRNA transcripts) and an increasing manifestation of a single dominant *VSG* being expressed in the switchers, which was *VSG-11* in 44 % of the cells (supplementary Figure a.15).

Figure 6.3: Mutual exclusive expression of *VSGs* is lost in ΔH3.VΔH4.V, but not in ΔH3.V or ΔH4.V cells.



a Representative immunofluorescence images targeting VSG-2 (green) on the cell surface. Cell bodies are stained with Tat1 (magenta) against tubulin. Scale bar: 10 μm, maximum projection of 26-28 stacks (0.1976 μm stack height). **b** Flow cytometry analysis of *VSG-2* expression on the cell surface of wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells. Gating strategy is depicted below.

Figure 6.4: Single-cell RNAseq reveals loss of mutually exclusive expression in individual cells.



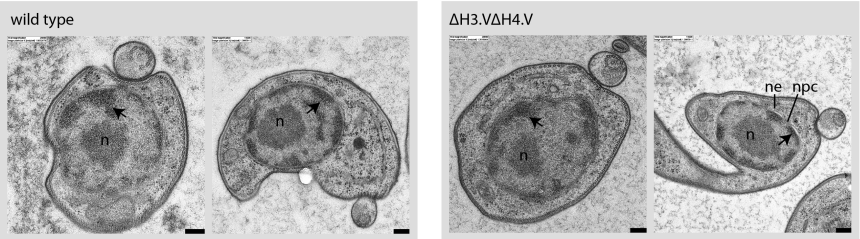
a Few cells (~1$x10^{-5}$ [119]) cells change *VSG* expression in laboratory-adapted strains. Increased heterogeneity of *VSG* expression can result from increased switching at the single-cell level or by loss of mutually exclusive expression on the population level (multiple *VSGs* expressed per single cell). **b** Analysis of *VSG* expression in n=40 wild-type and n=44 ΔH3.VΔH4.V cells using single-cell RNAseq. Each row corresponds to a single cell, columns correspond to selected *VSG* isoforms. Given is the percentage of a distinct *VSG* mRNA of all *VSG* mRNAs per single cell (considering uniquely mapped reads only). Sequencing reads were normalized for sequencing depth, library size, *VSG* gene length and uniqueness. Orange border indicates the dominantly expressed *VSG*.

In conclusion, scRNAseq showed that loss of histone variants H3.V and H4.V led primarily to increased *VSG* switching. However, the mutual presence of different *VSG* isoforms in individual cells also pointed to a partial loss of mutually exclusive expression. The analysis of a ΔH3.VΔH4.V population at two time points showed that *VSG* switching was ongoing in these cells and that certain *VSG* isoforms established as new dominant *VSGs* in the population of cells.

6.3.4 *Simultaneous deletion of H3.V and H4.V affected inter-chromosomal interactions*

To investigate the effect of histone variant H3.V and H4.V deletion on chromosome architecture, I examined wild-type and ΔH3.VΔH4.V

Figure 6.5: Electron microscopy across representative wild-type and ΔH3.VΔH4.V cells.



(n) nucleolus, (ne) nuclear envelope, (npc) nuclear pore complex. Black arrows indicate electron-dense chromatin at the nuclear periphery. 15,000-fold magnification. Scale bar: 200 nm.

nuclei by electron microscopy and re-analyzed the folding behavior of the *T. brucei* chromosomes by performing Hi-C in ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells.

Electron microscopy of n=8 wild-type interphase nuclei revealed the characteristic nuclear structure with a prominent nucleolus in the nuclear center and clustered electron-dense domains in close association with the nuclear envelope [207]. This pattern was not altered in ΔH3.VΔH4.V cells (n=66), suggesting that integrity of the overall nuclear architecture was not perturbed (Figure 6.5).

Modeling the folding behavior of chromosomal core regions and subtelomeres by measuring intra-chromosomal interaction frequencies over increasing genomic distances indicated on average a power-law scaling of -0.90 and -0.83 for core and subtelomeric regions for distances below 0.25 Mb (not shown). This indicated a slightly slower distance-dependent decay of interactions, but the folding behavior was not significantly different from wild-type cells (Wilcoxon matched-pairs signed rank test; p=0.44 between cores and p=0.31 between subtelomeres).

To investigate changes in inter-BES interactions, we measured the inter-chromosomal interaction frequency between BESs in wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells. Together with the interacting bins of interest, a randomized background sample was picked from the respective matrix and the ratio between the mean interaction frequency among BESs and randomized loci was determined. The background interaction frequency was permuted 100 times and the fold-change in interaction frequency of BESs over background was plotted. These calculations showed slightly reduced inter-BES interaction frequencies in ΔH4.V and a significant increase in inter-BES interactions in ΔH3.V and ΔH3.VΔH4.V cells (Figure 6.6 a).

To validate these findings, we performed FISH against the telomeric repeats as a proxy for the spatial distribution of chromosomal ends, which include the expression sites. Data obtained by FISH revealed

an even peripheral distribution of telomeres in wild-type and ΔH4.V cells and a marked clustering of telomeres upon deletion of H3.V. We quantified the occurrence of telomeric clusters using Imaris 8 (Oxford Instruments) and found them in more than 60% of ΔH3.V cells and more than 40% of ΔH3.VΔH4.V cells (Figure 6.6 b-c).
I subsequently investigated other genomic loci that strongly interacted in trans under wild-type conditions. Comparing the interaction behavior of *rRNA* and centromeric loci in wild-type and histone variant deletion cells by background modeling gave inconsistent results across different bin sizes. This is presumably due to the repetitive nature of thesis loci. Genome-wide 4C-like profiles of centromeric regions (Figure 6.6 d-e) and of *rRNA* genes (Figure 6.6 f-g) as virtual baits did, in contrast, not reveal evident changes in inter-chromosomal point-to-point interactions upon *H3.V* and *H4.V* deletion (across different matrix resolutions).
These findings indicated that the overall nuclear architecture in ΔH3.VΔH4.V cells was retained and that distinct inter-chromosomal interactions representing 'hallmarks' of an intact nuclear architecture such as the spatial association *rRNA* and centromeres were not disturbed. However, the increase of inter-BES interactions upon deletion of histone variants *H3.V* and *H4.V* may point to a loss of constraints that anchored the BESs to distinct nuclear sites. Thereby, mainly the deletion of *H3.V*, but not of *H4.V* appeared to be responsible for the architectural changes observed.

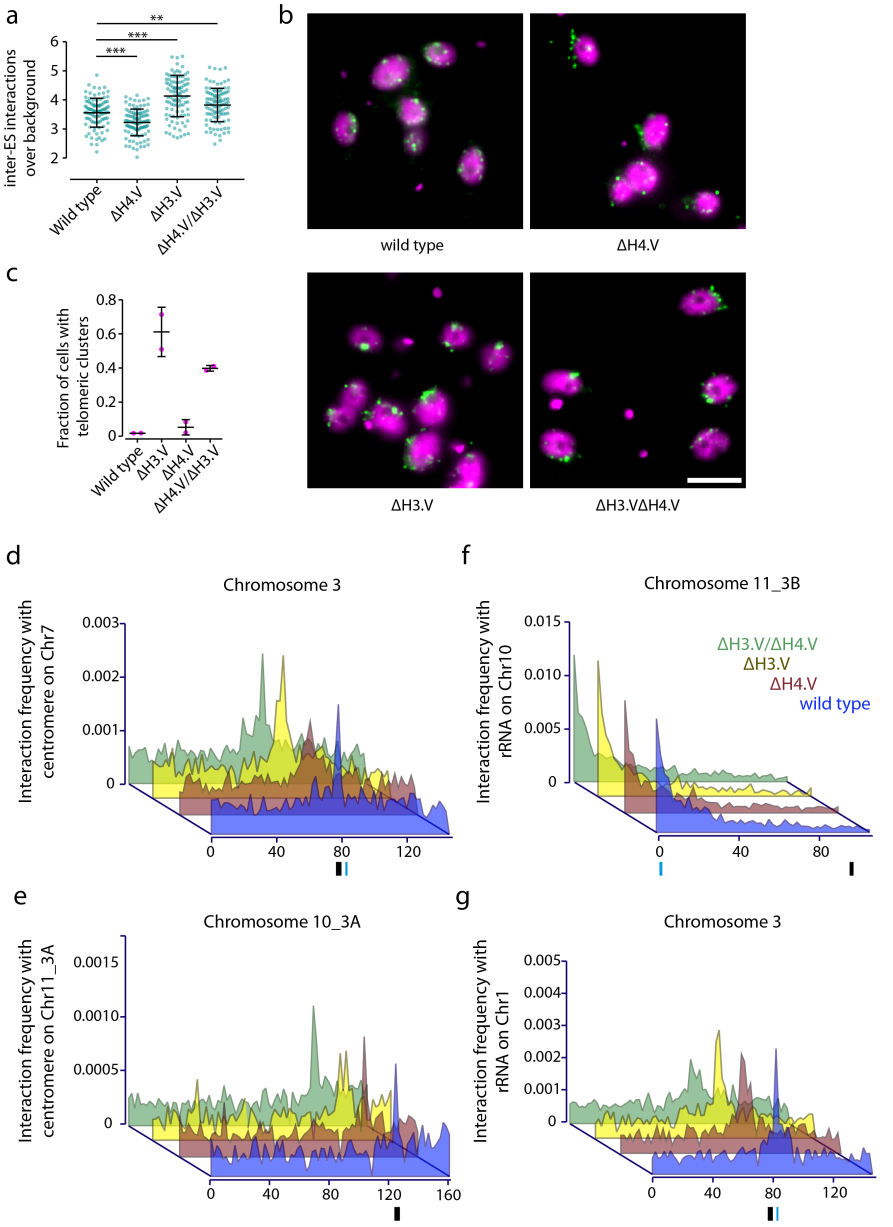
6.3.5 *Cohesin deposition was not dependent on H3.V or H4.V alone*

The close association of histone variants H3.V and H4.V and cohesin raised the possibility of an interdependence in deposition. To test the hypothesis that one of the histone variants is involved in cohesin recruitment, I performed Scc1-ChIPseq in ΔH3.V and ΔH4.V cells in duplicates. Cohesin deposition was, however, not altered in any of the two single knockout cell lines (data not shown). Still, H3.V and H4.V could act synergistically on cohesin deposition or positioning and deletion of one of the two histone variants could be rescued by the presence of the remaining one. Conducting Scc1-ChIPseq in the double deletion mutant was at this point not possible as available resistance markers were already exhausted in the ΔH3.V/ΔH4.V cell line.

6.3.6 *Deletion of H3.V and H4.V altered chromatin accessibility at distinct genomic sites in an additive manner*

Our previous results had indicated differential accessibility of chromatin at sites of cohesin and histone variant H3.V and H4.V enrichment. Also, RNAseq had indicated an activation of these sites in the

Figure 6.6: Silent expression sites relocate upon deletion of *H3.V*.

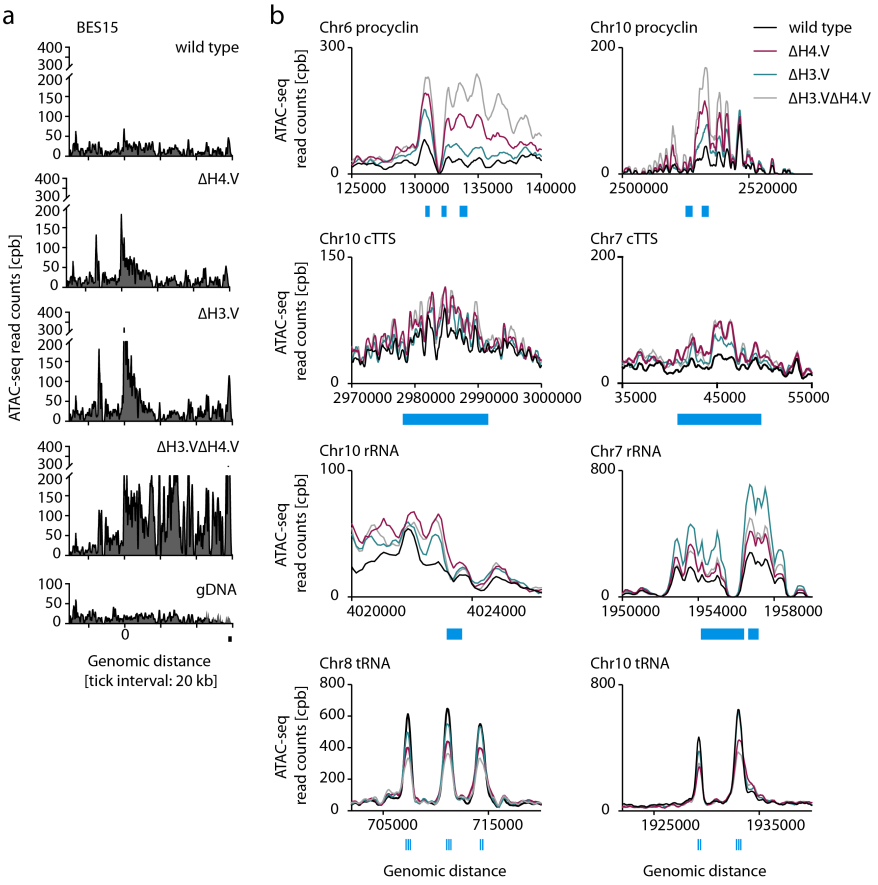


a Fold-change of interaction frequencies among silent *VSG* expression sites in wild-type, ΔH4.V, ΔH3.V and ΔH3.VΔH4.V cells compared to background interaction frequencies. The ratio between the query set and a randomized, fitted background sample was resampled and calculated 100 times. Significance was calculated using Welch's *t* test (two-sided; ** p<0.001, *** p<0.0001). Error bars indicate mean values and standard deviation. **b** Fluorescence *in situ* hybridization (FISH) against telomeric repeats (green, Alexa-488). Scale bar: 5 μm. **c** Quantification of large telomere clusters as determined by FISH in wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells. Mean values and standard deviation of n=2 replicates are given. **d-g** 4C-like profiles of centromeres and *rRNA* genes (20 kb bin size) in wild-type (blue), of ΔH3.V (yellow), ΔH4.V (red) and ΔH3.VΔH4.V (green) cells. Genomic location of centromeric repeats and *rRNA* are marked by black and blue boxes, respectively.

single, more pronounced in the double deletion mutant (e.g. *ESAGs*, *VSGs* and *mVSGs* located in expression sites, as well as *tRNA* and *procyclin* genes). To investigate a putative effect of H3.V and H4.V on local chromatin compaction, we performed ATACseq in ΔH3.V and ΔH4.V single and double deletion mutants.
The most dominant increase in chromatin accessibility upon histone variant deletion could be observed at previously silent expression sites: The deletion of only one of the two histone variants caused an 'opening' of chromatin at the site of expression site promoters but was not sufficient to entirely open chromatin at these strongly repressed sites. However, simultaneous deletion of both histone variants led to increased chromatin accessibility throughout the expression sites until the *VSG* gene (Figure 6.7 a and supplementary Figure a.16).
Little increase in local chromatin accessibility was also observed at other RNA Pol I transcribed loci, such as *procyclins* and *rRNA* genes (Figure 6.7 b). Again, the simultaneous deletion of H3.V and H4.V had the strongest effect on accessibility at these sites. However, at these loci the sole deletion of H4.V caused more increased opening of chromatin than deletion of H3.V. Chromatin accessibility was also slightly increased at RNA Pol II termination sites (Figure 11b).
In wild-type cells, *tRNA* genes, that are flanked by H3.V and H4.V showed very distinct and high peaks of chromatin accessibility across the tRNA CDS in ATACseq. In the histone variant deletion strains, ATAC signal was decreased, again, in an additive manner (Figure 6.7 b).

In summary, local chromatin accessibility was differentially altered at sites of cohesin and histone variant H3.V and H4.V deposition upon histone variant deletion. Irrespectively of whether H3.V and H4.V promoted an increase or decrease of chromatin accessibility at these sites, H3.V and H4.V acted in a synergistic manner at all sites concerned.
The massive opening of silent expression sites upon H3.V and H4.V deletion clearly indicated a role of the two histone variants in expression site silencing and control of antigenic variation.
The increase in chromatin accessibility at RNA Pol II termination sites and RNA Pol I transcribed sites upon H3.V and H4.V deletion showed that the two histone variants were not responsible for chromatin accessibility as observed in the wild-type condition in the first place, but rather had a repressive function at these sites.
At *tRNA* genes accessibility decreased upon knockout of flanking H3.V and H4.V. This may rather point to a role of H3.V and H4.V in insulating *tRNA* genes for RNA Pol III transcription.

Figure 6.7: H3.V and H4.V mediate local chromatin accessibility.



a ATACseq across BES15 (repressed in wild-type cells and activated upon deletion of H3.V and H4.V) in wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells. Only uniquely mapping reads are considered. gDNA read coverage (below) serves the illustration of mappability across BES15 (window size 501, step size 101 bp). The BES promoter region is designated by 'o', location of the *VSG* is marked by a black box. **b** ATACseq in wild-type (black), ΔH3.V (red), ΔH4.V (green) and ΔH3.VΔH4.V (grey) cells at representative sites *procyclin* genes, transcription termination sites, *rRNA* and *tRNA* genes. Respective genes or domains are marked by blue boxes.

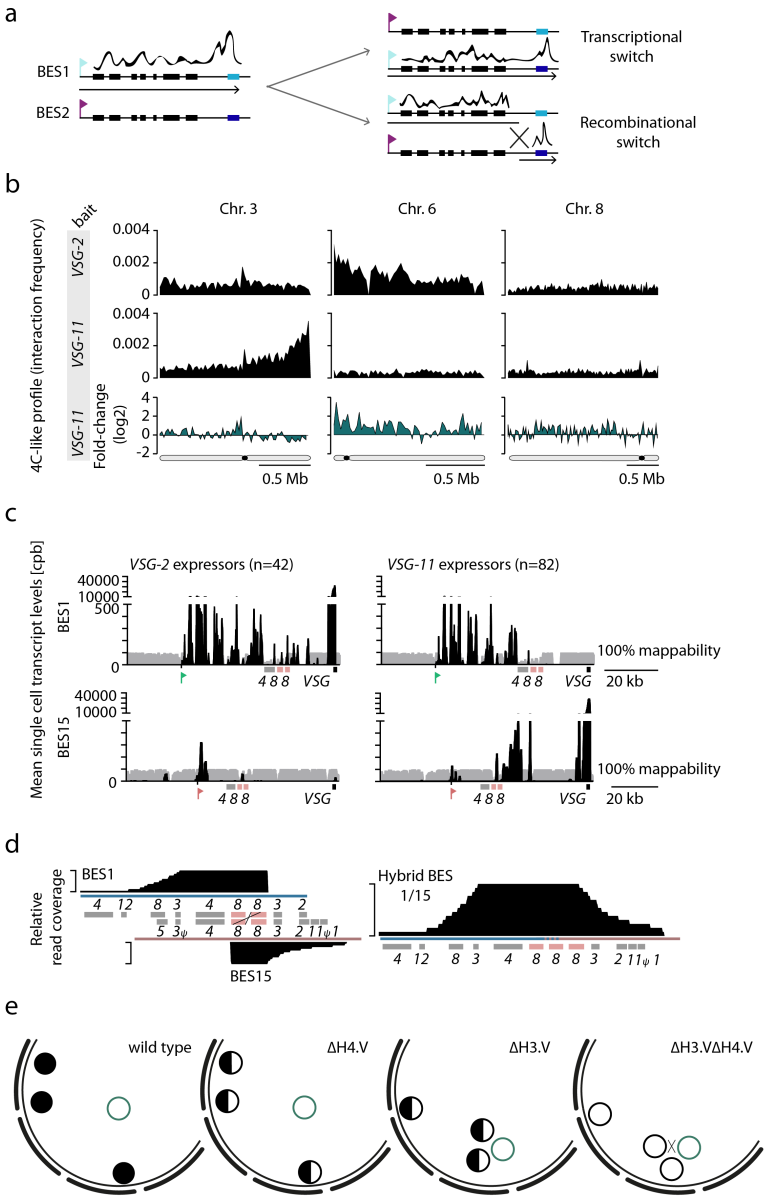
6.3.7 *Changes in global nuclear architecture and local chromatin accessibility mediated by H3.V and H4.V facilitated VSG switching via recombination*

In *T. brucei*, a switch in *VSG* expression can either occur by a transcriptional (*in situ*) or recombinational switch between expression sites (Figure 6.8 a).
To determine the mechanism of *VSG* switching in the ΔH3.VΔH4.V cell line, we investigated expression and interaction patterns of expression sites BES1 and BES15. BES1 contains *VSG-2*, that is active in wild-type cells and BES15 contains *VSG-11*, that was the most often activated *VSG* in the ΔH3.VΔH4.V cell line.
To distinguish between a transcriptional switch and a recombination event that had led to the activation of *VSG-11*, I analyzed 4C-like profiles of *VSG-2* and *VSG-11* across the populations of wild-type and ΔH3.VΔH4.V cells.
In wild-type cells (Figure 6.8 b, black plot), *VSG-2* showed high interaction with the 5′-end of chr6, where BES1 is located. *VSG-11* showed highest interactions with the 3′-end of chr3, which is juxtaposed to BES15. Plotting the log2-fold change between the 4C-like profiles of *VSG-11* in wild-type and ΔH3.VΔH4.V cells across chromosomes 3 and 6 revealed a ~16-fold increase in interaction between *VSG-11* and the 5′-end of chr6. This pointed towards a recombination event between BES15 and BES1 (6.8 b, green plot).
To analyze the transcription profiles of BES1 and BES15, we selected and pooled the reads from n= 82 ΔH3.VΔH4.V single cells that had switched towards an activation of *VSG-11* and compared the expression profile to that of n=42 pooled *VSG-2* expressors.
The average expression profiles revealed high transcription of BES1 from the ES promoter towards *VSG-2* and only sparse transcription around the promoter region of BES15 in cells that had retained *VSG-2* transcription. In the *VSG-11* expressing cells, transcription of BES1 abruptly ended at the first *ESAG-8* gene and re-initiated in the middle of BES15 around the location of *ESAG-8* until the telomere-proximal *VSG-11*. These findings supported a recombination event between BES1 and BES15 at the sites of their respective *ESAG-8* gene (Figure 6.8 c).
To pin down the exact site of recombination between BES1 and BES15, we performed long-read sequencing (SMRTseq) in ΔH3.VΔH4.V cells. We extracted reads larger than 10 kb and split them down to ~2,5 kb fragments, which were mapped independently to the genome. Long reads containing fragments that mapped to both, BES1 and BES15 were selected and examined for a break in the alignment using the Integrative Genomics Viewer (IGV). The site of breakage pointed to a recombination that had occurred between the ESAG8 gene pair present in both expression sites. We virtually constructed a chimeric

expression site from BES1 and BES15 by referring to the breakpoint identified in the long reads. This chimeric ES contained three copies of *ESAG-8*, one retained from BES1 and two originating from BES15. Mapping the SMRT reads in their full length against the chimeric ES bridged the recombination junction entirely (Figure 6.8 d).

Taken together, scRNAseq transcriptome and Hi-C interaction data revealed that recombination between BES1 and BES15 had led to activation of *VSG-11* in ΔH3.VΔH4.V cells. Further, the exact point of recombination was identified using long-read sequencing.
We suggest that in ΔH3.VΔH4.V cells, recombination between expression sites was facilitated by the combination of both, an increase in local chromatin accessibility and an increase in expression site clustering. Although most architectural constraints seemed mediated by H3.V, only the concurrent deletion of *H3.V* together with *H4.V* sufficiently opened chromatin at expression sites and thus poised them for recombination (Figure 6.8 e).

Figure 6.8: *VSG* recombination is promoted in ΔH3.VΔH4.V cells.



a Illustration of possible switching mechanisms between two expression sites. In a transcriptional switch, transcription of one expression site is shut down with concomitant activation of a new expression site. Recombinational switches involve an exchange of the *VSG* (blue box) located in the active expression site by a new *VSG* gene (green box). The black line outlines the expected transcription profiles for the two switching scenarios. **b** Inter-chromosomal 4C-like interaction profiles of BES1 (top) and BES15 (middle) with chr3, 6 and 8 in wild-type cells. The lower panel shows the log2 fold-change of interactions of BES15 across chr3, 6 and 8 upon deletion of *H3.V* and *H4.V*.

c Transcript profiles of individual ΔH3.VΔH4.V cells expressing either *VSG-2* (n=42) or *VSG-11* (n=82) were merged and the average transcript profiles (uniquely mapping reads only) across BES1 (top) and BES15 (bottom) are shown. The degree of 'uniqueness' was determined by mapping of *in silico* generated reads (grey bars). **d** Coverage of PacBio SMRT sequencing reads (gDNA from ΔH3.VΔH4.V cells, filtered for reads larger than 10 kb) across BES1 (blue, top left) and BES15 (red, bottom left). A BES1-BES15 hybrid BES was generated according to the site of recombination (marked by a black cross). *ESAG* genes are represented as boxes. **e** Hypothetical model of the roles of H3.V and H4.V on local chromatin accessibly and global genome architecture. Knockout of the individual *H3.V* or *H4.V* genes causes partial opening of the expression sites. Additionally, deletion of *H3.V* leads to a spatial rearrangement and clustering of expression sites. Deletion of both histone variants provides critical proximity between expression sites combined with increased accessibility which facilitates *VSG* switching by DNA recombination among them.

Part III

Materials and methods

7 Cell culture

7.1 *T. brucei* cell culture

All strains used for this work are derivates of the Lister 427 isolate (see Table 7.3). Bloodstream form parasites were cultured in HMI-11 (HMI-9 without serum plus) [114] at 37°C up to a density of 10^6 cells/ml and procyclic form parasites were cultured in SDM-79 [35] at 27°C. Drug selection was applied if required at standard concentrations.

7.2 Cell lines used in this study

7.2.1 *ΔH3.V and ΔH3.V deletion strains*

The histone variant H3.V and H4.V single deletion strains had been published previously [162] [245] (Table 7.3).

7.2.2 *ΔH4.V/Ty1-H4.V ectopic expression*

The *ΔH4.V/Ty1-H4.V ectopic expression* cell line had been published previously [245]. Here, both endogenous *H4.V* alleles were knocked out and a Ty1-tagged version of *H4.V* is inducibly overexpressed from an ectopic (ect.) rRNA spacer (Table 7.3).

7.2.3 *Generation of a ΔH3.VΔH4.V cell line*

To generate a *H3.V* and *H4.V* double deletion mutant (Table 7.4), both *H4.V* alleles were deleted in the *ΔH3.V* [162] cell line. Therefore, the regions flanking the *H4.V* (Tb927.10.15350) CDS upstream (H4.V_01_F, H4.V_02_R) and downstream (H4.V_03_F, H4.V_04_R) (Table 7.1) were amplified from *T. brucei* wild-type genomic DNA and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). The parental plasmid *pyrFEKO-Neo* [233] (see Table 7.2)was linearized using HindIII and AgeI restriction sites and the upstream region was inserted using InFusion HD Cloning Plus reagents (Clontech). The downstream region was inserted in a ligation reaction using BamHI and SbfI restriction sites. To be able to use the plasmid construct for deletion of both *H4.V* alleles, the G418 resistance cassette was exchanged with a blasticidin and phleomycin resistance marker, respectively. The resistance cassettes were excised from the plasmids *pyrFEKO-BSD* or *pyrFEKO-Phleo* [233] (Table 7.2) by restriction digest (BglII and XbaI) and ligated into the *H4.V* deletion construct using the

same restriction sites.
The resulting plasmids *pyrFEKO-BSD_H4V_KO* and *pyrFEKO-Phleo_H4V_KO* (see Table 7.2) were linearized using NheI and SbfI and transfected into the Δ*H3.V* [162] cell line using a Nucleofector apparatus (Lonza).

7.2.4 *Generation of a Ty1-H3.V/ΔH3.V cell line*

For deletion of the first allele of *H3.V* (Tb927.10.15350), the *H3.V* CDS flanking regions were amplified using the primer pairs H3.V_01_F, H3.V_02_R and H3.V_03_F, H3.V_04_R (see Table 7.1) and cloned into the plasmid *pyrFEKO-Puro* [233] (see Table 7.2) using PvuII/HindIII and BamHI/SbfI restrictions sites, respectively and InFusion HD Cloning Plus reagents (Clontech). Prior to transfection into wild-type cells, the resulting plasmid was linearized using PvuII and SbfI.
For the N-terminal tagging of the second H3.V allele with 2x *Ty1*, a region upstream of the H3.V CDS was amplified (H3.V_05_F, H3.V_06_R) and cloned into *pPOTv3-2xTy1* [61] (Table 7.2) at ApaI and NotI sites with InFusion HD Cloning Plus reagents (Clontech). Likewise, the 5′-end of the *H3.V* gene was amplified (H3.V_07_F, H3.V_08_R, Table 7.1) leaving out the ATG start codon and inserted into the plasmid at SacI and NheI sites. Subsequently, the synthetic *Ty1* tag was exchanged by a codon-optimized version. Therefore, two complementary oligonucleotides (H3.V_09_F and H3.V_10_R, Table 7.1) were annealed, HindII and SacI digested at ligated into the plasmid using the same restriction sites. The plasmid was linearized with ApaI and NheI and stably transfected into Δ*H3.V/H3.V T. brucei*, yielding *Ty1-H3.V/ΔH3.V* cells (see Table 7.4).

7.2.5 *Generation of a Ty1-Scc1/ΔScc1 cell line*

For deletion of the first *Scc1* allele (Tb927.7.6900), regions flanking the *Scc1* CDS upstream (Scc1_01_F, Scc1_02_R, 7.1) and downstream (Scc1_03_F, Scc1_04_R, Table 7.1) were amplified, digested with PvuII/HindIII and BamHI/SbfI restriction enzymes, respectively and ligated into the parental vector *pyrFEKO-Hyg* [233] (Table 7.2) using the same restriction sites. Transfection of wild-type *T. brucei* bloodstream form cells resulted in *Scc1*/Δ*Scc1* cells. To tag the second *Scc1* allele N-terminally, the 3′-end of the Scc1 5′-UTR (Scc1_05_F, Scc1_06_R, Table 7.1) and the 3′-end of the Scc1 CDS (Scc1_07_F, Scc1_08_R, Table 7.1) were amplified, thereby omitting the Scc1 ATG start codon. The PCR products were digested using ApaI/NotI and SacI/NheI restriction enzymes and were in subsequent ligation reactions inserted into the vector *pPOTv3-2xTy1* [61] using the same restriction sites. As described for N-terminal tagging of *H3.V*, the *Ty1* tag was exchanged for a codon optimized version. The plasmid was linearized with ApaI

and NheI and was stably transfected into *Scc1*/Δ*Scc1 T. brucei* cells to generate *Ty1-Scc1*/Δ*Scc1* (see Table 7.4).

Table 7.1: Oligonucleotides used for cloning of plasmid constructs.

Name	**Sequence (5′ -3′ end)**
H3.V_01_F	AGGGCGAATTGGGCCCAGAAATACCTCATTACGTACCA
H3.V_02_R	GCATTATACGCGGCCGCGAATATAATCTAAGATATATATATATATATATCTTTTTA CTTAATCTCT
H3.V_03_F	AAAACCTGCAGGATGTAAATGTGCAGGAGATG
H3.V_04_R	AAAAGGATCCGAAAACAGCGAAAGGGAAAA
H3.V_05_F	AGGGCGAATTGGGCCCAGAAATACCTCATTACGTACCA
H3.V_06_R	GCATTATACGCGGCCGCGAATATAATCTAAGATATATATATATATATATCTTTTTA CTTAATCTCT
H3.V_07_F	GATCGGGTAGTTAAGAGCTCGCGCAAATGAAGAAAATAACACCG
H3.V_08_R	GCTTTTTCATGCTAGCTTAGTTACGCTCGCCTCGGA
H3.V_09_F	AGCTTATGGAGGTGCACACGAACCAGGACCCGCTGGACGCCGAGGTGCACA CGAACCAGGACCCGCTGGACGGATCCGGATCAGGATCTGGATCAGGATCGG GTAGTGAGCT
H3.V_10_R	CACTACCCGATCCTGATCCAGATCCTGATCCGGATCCGTCCAGCGGGTCCTGG TTCGTGTGCACCTCGGCGTCCAGCGGGTCCTGGTTCGTGTGCACCTCCATA
H4.V_01_F	ATCGGGATCCGCAAGGAAGCGAGCGTAATA
H4.V_02_R	GATCCCTGCAGGCGTTCACTACTCCATACAAC
H4.V_03_F	TAAGTTGGATAAGCTGTGTAGCGGCTGGGGTTG
H4.V_04_R	GAGAACGTGTACCGGCTCGGGTTGCTTGTGATTTT
Scc1_01_F	GATCCAGCTGCCTATCGGTGCCTGTTTGTT
Scc1_02_R	GATCAAGCTTCCTTCGGGAATGCTTACTCC
Scc1_03_F	GATCGGATCCGACAAGAACGACCCAACGAA
Scc1_04_R	GATCCCTGCAGGCCCCGGCGCAAATATGTA
Scc1_05_F	GATCGGGCCCCCTATCGGTGCCTGTTTGTT
Scc1_06_R	GATCGCGGCCGCCTTGTTTTATTCAAGTGCAATATTTTACCTTCGGGAATGCTT
Scc1_07_F	GATCGAGCTCTTCTACTCAACCTACGTGCTGACGA
Scc1_08_R	GATCGCTAGCCAGGGCAATGGGAACCAC

Table 7.2: Table of plasmids generated in this study.

Plasmid	Purpose	Resistance in *E. coli*	Resistances in *T. brucei*	Parental plasmid	References
pyrFEKO-BSD_H4V_KO	Deletion of H4.V (Tb927.10.15350)	ampicillin	blasticidin, HSVTK	pyrFEKO-Neo	Scahill et al., 2008; Müller et al. 2018
pyrFEKO-Phleo_H4V_KO	Deletion of H4.V (Tb927.10.15350)	ampicillin	phleomycin, HSVTK	pyrFEKO-Neo	Scahill et al., 2008; Müller et al. 2018
pyrFEKO-Puro_H3V_KO	Deletion of H3.V (Tb927.10.15350)	ampicillin	puromycin, HSVTK	pyrFEKO-Puro	Scahill et al., 2008; Müller et al. 2018
pyrFEKO-Hyg_Scc1_KO	Deletion of Scc1 (Tb927.7.6900)	ampicillin	hygromycin, HSVTK	pyrFEKO-Hyg	Scahill et al., 2008; Müller et al. 2018
pPOTv3_Ty1_H3V	N-terminal tagging of H3.V with 2x Ty1	ampicillin	BSD	pPOTv3-2xTy1	Dean et al., 2015; Müller et al., 2018
pPOTv3_Ty1_Scc1	N-terminal tagging of Scc1 with 2x Ty1	ampicillin	BSD	pPOTv3_Ty1_H3V	Dean et al., 2015; Müller et al., 2018

Abbreviations: HSVTK, herpes simplex virus thymidine kinase; BSD, blasticidin.

Table 7.3: Table of previously published cell lines used in this study.

Cell line	Parental cell line	Constructs	Genotype	Selection	Feature	Reference
Lister 427 BF	Lister 427 BF	-	Lister 427, MITat1.2	-	wild type	Cross, 1975
Lister 427 PF	Lister 427 PF	-	Lister 427, MITat1.2	-	wild type	Cross, 1975
single-marker (SM)	Lister 427 BF	pHD328, pLew114hyg5	Lister 427, MITat1.2, TETR T7RNAP NEO	G418	Inducible ectopic overexpression of T7 polymerase	Wirtz et al., 1999
ΔH3.V (BFjel25)	single-marker	pJEL89, pJEL38	SM, ΔH3V::PAC ΔH3V::HYG	G418, puromycin, hygromycin	H3.V deletion	Lowell et al., 2004
ΔH4.V	single-marker	n/a	SM, ΔH4V::PAC ΔH4V::HYG	G418, puromycin, hygromycin	H4.V deletion	Siegel et al., 2009
PN221	Lister 427 BF	pLF12, pLF13	Lister 427 BF, BES1 PAC, BES17 NEO	puromycin [G418]	VSG-2 expressed from a puromycin-tagged BES1; BES17 is tagged with G418	Figueiredo et al., 2008
PN13	Lister 427 BF	pLF12, pLF13	Lister 427 BF, BES1 PAC, BES17 NEO	G418 [puromycin]	VSG-13 expressed from a G418-tagged BES17; BES17 is tagged with puromycin	Figueiredo et al., 2008
ΔH4.V/Ty1-H4.V ect.	single-marker	n/a	SM, RRNA::Ty1-H4.V BLE, ΔH4V::PAC ΔH4.V::HYG	G418, blasticidin, puromycin, hygromycin, doxycycline	Both endogenous H4.V alleles are deleted. A Ty1-tagged version of H4.V is inducibly ectopically overexpressed	Siegel et al., 2009

BF, bloodstream-form; PF, procyclic-form; SM, single-marker; MITat1.2, Molteno Institute Trypanozoon Antigen Type 1.2; ect., ectopic overexpression; TETR, tetracycline-repressor; T7RNAP, T7 RNA polymerase; Δ, gene deletion; PAC, puromycin-N-acetyltransferase; BLE, phleomycin resistance gene; BSD, blasticidin S deaminase; NEO, aminoglycoside phosphotransferase (confers resistance to G418); HYG, hygromycin phosphotransferase. When required for selection, the cell culture medium was supplemented with antibiotics as follows: 2 μg/ml G418 (Roth), 0.1 μg/ml puromycin (Sigma Aldrich), 5 μg/ml hygromycin (Roth), 25 μg/ml blasticidin (Invivogen), 2.5 μg/ml phleomycin, 1 μg/ml doxycycline (Sigma Aldrich). Antibiotics written in braces indicate that resistance was conferred to the cell line, but the according antibiotic was not added to the culture medium.

Table 7.4: Table of cell lines generated in this study.

Cell line	Parental cell line	Constructs	Genotype	Selection	Feature	Reference
ΔH3.VΔH4.V	ΔH3.V (BFjel25.2)	pyrFEKO-BSD_H4V_KO, pyrFEKO-Phleo_H4V_KO	SM, ΔH3V::PAC ΔH3V::HYG ΔH4V::BLE ΔH4V::BSD	G418 [puromycin, hygromycin, phleomycin, blasticidin]	H3.V and H4.V double deletion mutant	Müller et al., 2018
Ty1-H3.V/ΔH3.V	Lister 427 BF	pyrFEKO-Puro_H3V_KO, pPOTv3_Ty1_H3V	Lister 427 BF, ΔH3V::PAC Ty1-H3.V BSD	puromycin, blasticidin	One H3.V allele is endogenously N-terminally tagged with 2x Ty1; the other allele is deleted	Müller et al., 2018
Ty1-Scc1/ΔScc1	Lister 427 BF	pyrFEKO-Hyg_Scc1_KO, pPOTv3_Ty1_Scc1	Lister 427 BF, ΔScc1::HYG Ty1-Scc1 BSD	hygromycin, blasticidin	One Scc1 allele is endogenously N-terminally tagged with 2x Ty1; the other allele is deleted	Müller et al., 2018

BF, bloodstream-form; SM, single-marker; MITat1.2, Molteno Institute Trypanozoon antigenic Type 1.2; TETR, tetracycline-repressor; T7RNAP, T7 RNA Polymerase; Δ, gene deletion; PAC, puromycin-N-acetyltransferase; BLE, phleomycin resistance gene; BSD, blasticidin S deaminase; NEO, aminoglycoside phosphotransferase (confers resistance to G418); HYG, hygromycin phosphotransferase. When required for selection, the cell culture medium was supplemented with antibiotics as follows: 2 μg/ml G418 (Roth), 0.1 μg/ml puromycin (Sigma Aldrich), 5 μg/ml hygromycin (Roth), 25 μg/ml blasticidin (Invivogen), 2.5 μg/ml phleomycin, 1 μg/ml doxycycline (Sigma Aldrich). Antibiotics written in braces indicate that resistance was conferred to the cell line, but the according antibiotic was not added to the culture medium.

8

NGS-based approaches

8.1 *In situ* Hi-C in *T. brucei*

Hi-C was performed in triplicates in single-marker, Δ*H3.V*, Δ*H4.V* and Δ*H3.V*Δ*H4.V* cells, as well as in duplicates in *PN221*, *PN13* [83] and procyclic *T. brucei* cells (see Table 7.3). All *T. brucei* cell lines used are derivatives of the Lister 427 cell line.
The experimental Hi-C workflow was adapted to *T. brucei* and performed based on previously published protocols [156] [222] [22]. Per Hi-C experiment, $2x10^8$ cells were spun down (10 min, RT, 800xg) and resuspended in 40 ml of 1x trypanosome dilution buffer (TDB; 0.005 M KCl, 0.08 M NaCl, 0.001 M $MgSO_8$ $x7H_2O$, 0.02 M Na_2PO_4, 0.002 M Na_2HPO_4 $x2H_2O$, 0.02 M glucose, Table 9.2). 4 ml of formaldehyde solution (50 mM Hepes-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 11% formaldehyde, Table 9.2) were added to fix the cells in the presence of 1% formaldehyde for 20 minutes at room temperature (RT). The cross-linking reaction was stopped with 3 ml of 2 M glycine and incubation for 5 minutes at RT following 15 minutes on ice. After washing the cells twice with 1x TDB, the cell pellet was snap frozen in liquid nitrogen and stored at −80°C until further use.
The cell pellet was thawed on ice and cells were resuspended in 1 ml of permeabilization buffer (100 mM KCl, 10 mM Tris pH 8.0, 25 mM EDTA, Table 9.2) supplied with protease inhibitors (1.5 mM pepstatin A, 4.25 mM leupeptin, 1.06 mM PMSF, 1.06 mM TLCK) and 200 μM digitonin and permeabilized for 5 minutes at RT. Cells were subsequently washed twice in 1x NEBuffer3.1 (NEB, B7003S) and were finally resuspended in 342 μl of 1x NEBuffer3.1. Then 28 μl of 1% SDS were added, cells were incubated for exactly 10 minutes at 65°C. The SDS was quenched by addition of 43 μl of 10% Triton-X 100 (Sigma) and incubation at RT for 15 minutes. Another 13 μl of 10x NEBuffer3.1 and 100 units of MboI (NEB, R0147M) and 35 μl of water were added and the chromatin was digested *in situ* at 37°C overnight while shaking. The next day, residual MboI was heat-inactivated for 20 minutes at 65°C. By addition of 60 μl of fill-in mix (0.25 mM biotin-14-dATP (Life Technologies, 19524016), 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP (Fermentas), 40 U of DNA polymerase I Large Klenow Fragment (NEB, M0210)) and incubation at 23°C for 4 hours, restriction fragments were blunted and thereby biotinylated. Subsequently, 665 μl of ligation mix (1.8% Triton-X 100, 0.027 % (w/v) BSA, 1.8x T4 DNA Ligase Buffer (Invitrogen, 46300018) and 5 U of T4 DNA ligase

(Invitrogen, 15224025), Table 9.2) were added and proximity-ligation of end-repaired chromatin was performed at 16°C for four hours using a ThermoMixer (Eppendorf) with interval shake (900 rpm for 10 seconds every 5 minutes). Crosslink reversal was performed by adding 50 μl of 10 mg/ml Proteinase K, incubation at 65°C for four hours, another addition of 50 μl of 10 mg/ml Proteinase K, 80 μl of 5M NaCl and 70 μl of 10% SDS and continued incubation at 65°C overnight.
The next day, DNA was purified by ethanol precipitated and resuspended in 257 μl Tris low EDTA (TLE) buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0, Table 9.2). The sample concentration was determined using Qubit (Qubit dsDNA HS Assay Kit, Thermo Fisher) and 50 ng were reserved for a gel check (see below). Prior to sonication (Covaris S220, 175 W peak incident power, 10% duty factor, 200 cycles per burst, 240 seconds treatment) SDS was added to a final concentration of 0.1% and the sample was split among two microtubes as the DNA content exceeded the capacity of a single sonication vial. After sonication, the split sample was recombined in an Eppendorf tube and the volume was adjusted to 300 μl with TLE. Again, 50 ng were kept for a gel check (see below) to control for successful sonication. To size-select for fragments between 100 and 400 bp in size, 165 μl of Agencourt AMPure XP beads (Beckman Coulter, A63881) were added to the sample and vortexed. After incubation for 10 minutes at RT, the sample was separated on a magnetic rack for 5 minutes. DNA fragments larger than 400 bp were kept on the beads and removed from the sample. Another 500 μl of AMPure XP mixture (without sample) were incubated on the magnetic rack for 5 minutes, the supernatant was removed and the beads were resuspended in 225 μl AMPure XP mixture. The sample supernatant was added to the AMPure XP mixture and incubated for 10 minutes at RT. Then, the sample was separated on the magnetic rack and the supernatant containing fragments smaller than 100 bp was discarded. The Hi-C sample was eluted from the beads by addition of 55 μl of TLE, incubation at RT for 10 minutes and separation on the magnetic rack. The eluate was transferred to a fresh Eppendorf tube and the DNA content was measured using Qubit dsDNA High Senisitivity Assay (Life Technologies, Q32854). 50 ng of sample were kept for analysis on a polyacrylamide gel.
To remove biotin from unligated ends and to blunt the sheared fragments, 70 μl of end-repair mix (Table 9.2; 1x Ligation buffer (NEB), 357 μM dNTPs, 25U T4 PNK (NEB, M0201), 7.5U T4 DNA polymerase I (NEB, M0203), 2.5U DNA polymerase I, large (Klenow) fragment (NEB, M0210)) were added and the reaction was incubated for 30 minutes at 20°C, then 10 mM (final concentration) EDTA was added and incubation was continued for 20 minutes at 75°C to heat-inactivate the enzymes.
Next, fragments containing biotin-labeled ligation junctions were enriched in a biotin-pull down step. Therefore, 50 μl of 10 mg/ml MyOne

Streptavidin C1 Dynabeads (Life Technologies, 65001) were washed once with 400 μl of 1x Tween washing buffer (TWB, Table 9.2; 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20), then collected on a magnetic rack and resuspended in 400 μl of 2x Binding Buffer (BB) (Table 9.2; 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl). The Dynabeads and additional 330 μl of TLE were added to 70 μl of sample and incubated for 15 minutes at RT with rotation. The beads were reclaimed on the magnetic rack and the supernatant was discarded. The sample was subsequently washed on the magnetic rack with 400 μl of 1x BB and 100 μl of 1x TLE with transferal to a fresh tube after each washing step. The beads were finally reclaimed again and resuspended in 41 μl of 1x TLE.
The following steps served the library preparation. Therefore, the Hi-C fragments were polyadenylated by adding 5 μl of 10x NEBuffer 2.1, 1 μl of 10 mM dATP and 3 μl of 5 U/μl of Klenow fragment (3′ - 5′ exo-) (NEB, M0212) and incubation for 30 minutes at 37°C followed by a deactivation step for 20 minutes at 65°C. The sample bound to the beads was cooled on ice and was subsequently collected on a magnet, washed once with 400 μl 1x Quick ligation buffer (NEB, M2200) and resuspended in 46.5 μl of 1x Quick ligation buffer. For adapter ligation, 0.5 μl of 50 μM annealed Illumina TruSeq adapters and 2.5 μl of DNA Quick ligase (NEB, M2200) were added and ligation was performed for 1 hour at RT. Then, beads were collected on a magnet and resuspended in 400 μl of 1x Tween wash buffer (TWB) (Table 9.2; 5 mM Tris-HCl, 0.5 M EDTA, 1 M NaCl, 0.05% Tween-20) and incubated for 5 minutes at RT with rotation. Beads were washed with 200 μl of 1x BB and twice with 200 μl of 1x NEBuffer 2.1, each time reclaiming the beads on a magnet and transferal to a fresh tube. After the last wash, the beads were resuspended in 20 μl of 1x NEBuffer 2.1 and transferred to a fresh tube.
To titrate the number of library amplification cycles, 2 μl of the sample were mixed with 5 μl of TruSeq PCR primer cocktail (25 μM; TruSeq PCR primer cocktail_F, TruSeq PCR primer cocktail_R; Table 8.1), 75 μl of 2x KAPA HiFi HotStart Ready Mix (Kapa Biosystems) and 60 μl of water. This master mix was split in 5 x 25 μl reactions which were amplified with 6, 8, 10, 12 or 14 cycles as follows:
Denaturation for 3 minutes at 95°C, 6-14 cycles of 20 seconds at 98°C, 30 seconds at 63°C and 30 seconds at 72°C, then final elongation for 1 minute at 72°C. The whole reaction volume was loaded on a 2% agarose gel, run at 90 V for ~1 hour and post-stained with ethidium bromide (For an example see Table a.2 a). As proposed by Dekker and colleagues [22], the suggested number of cycles is 1-2 cycles below the number of cycles of the fewest PCR product visible on the gel. Note that with library over-amplification, PCR product sizes shift towards higher molecular weights. In this dissertation, all final library amplifications were done as follows: the remaining 18 μl of library

were readjusted to 20 μl volume with 1x TLE and were split among 8 separate reactions of 50 μl total volume. Per reaction, 1.5 μl of 25 μM TruSeq PCR primer cocktail (Table 8.1) 25 μl of 2x Kapa HiFi HotStart Ready Mix (Kapa Biosystems, KR0370) and 21.5 μl of water were added to 2 μl of library bound to the beads. The PCR was performed as before, but with 5 cycles of amplification. Afterwards, the reactions were pooled and the beads were reclaimed on the magnet. The PCR products reside in the supernatant, as they are not biotinylated. The supernatant was kept and purified using Agencourt AMPure XP beads (Beckman Coulter). Therefore, 1.5x volumes of AMPure XP mixture were added to the supernatant and incubated for 10 minutes at RT. Then the sample was placed on the magnet for 5 minutes and the supernatant was discarded. The beads were washed twice with 70% ethanol on the magnet and were briefly air-dried. Then the beads were resuspended in 25 μl of TLE buffer, transferred to a fresh Eppendorf tube and DNA concentration was determined using Qubit (Qubit dsDNA HS Assay Kit, ThermoFisher). 50 ng were reserved for a gel check.
To determine quality of the library and to control for successful shearing, size selection and adapter-ligation, the 50 ng control samples reserved throughout the protocol were loaded onto a 5% polyacrylamide gel, which was post-stained with SYBR Gold (ThermoFisher) (Table a.2 b).

8.2 Quantification and sequencing of NGS-libraries

Library molarities were quantified by qPCR (KAPA SYBR FAST qPCR Master Mix, Kapa Biosystems) and a sequencing pool was generated accordingly. Roughly, for Hi-C experiments, $1x10^8$ reads were aimed at. For libraries generated in ChIP experiments, between $2x10^6$ and $1x10^7$ reads were required. RNAseq experiments required ~$15x10^6$ reads. For all libraries, paired-end sequencing (76 bp) was performed using the Illumina NextSeq 500 platform and following the manufacturer's instructions.
Programs used for the different steps of NGS data analysis are listed in Table 8.3.

Table 8.1: Oligonucleotides used for library indexing and amplification.

TruSeq	**Sequence (5′ -3′ end)**
TruSeq_F	AATGATACGGCGACCACCGA*G
TruSeq_R	CAAGCAGAAGACGGCATACGA*G
Nextera	**Sequence**
Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCATCAGATGTG
Ad2.1	CAAGCAGAAGACGGCATAGCAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2	CAAGCAGAAGACGGCATAGCAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT

Ad2.3	CAAGCAGAAGACGGCATAGCAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4	CAAGCAGAAGACGGCATAGCAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5	CAAGCAGAAGACGGCATAGCAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6	CAAGCAGAAGACGGCATAGCAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7	CAAGCAGAAGACGGCATAGCAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8	CAAGCAGAAGACGGCATAGCAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9	CAAGCAGAAGACGGCATAGCAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10	CAAGCAGAAGACGGCATAGCAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT

Illumina TruSeq oligonucleotides used for amplification of libraries with TruSeq adapters and Nextera barcoded primers used for indexing of ATAC-seq samples. *, phosphothioate bond

8.3 Computational analysis of Hi-C libraries

8.3.1 *Mapping of Hi-C reads and matrix generation*

Hi-C reads were trimmed at the site of adapter ligation using cutadapt (version 1.10) [172] and truncated at the site of their ligation junction using the HiCUP (version 0.5.9 devel). The paired-end reads were mapped independently using BWA-mem (https://arxiv.org/abs/1303.3997) of the bwa kit version 0.7.12-r1039 to the Lister 427 genome v9 [194] and only 'uniquely' mapping reads were retained (MAPQ>0). The genome version used did not contain allelic variants of a respective assembled allele (Tb427v9_without_allelic_variants) to ensure that every locus is represented once in the interaction matrices and not discarded due to being non-unique. In contrast, during read mapping all unassembled contigs were present to make sure each read maps to the best possible locus thereby reducing false positive hits. Reads were paired and unpaired reads were discarded. For removal of invalid read pairs, read binning, matrix generation and balancing (iterative correction and eigenvector decomposition [125], default settings) HiC-Pro (version 2.10.0) was used [240]. Prior to matrix balancing, the *T. brucei* matrices were corrected for differences in ploidy between the core and subtelomeric regions. Therefore, all 'intra-core' interactions were multiplied by 0.5. As Hi-C Pro provides interaction frequencies in a sparse two-column format, we converted the matrices for some applications to a classical redundant matrix format as employed by Homer [110] using a custom python script. Subsequently, each value in a balanced matrix was divided by its respective column sum to allow a direct comparison between different matrices.

8.3.2 *Analysis of Hi-C matrices*

8.3.2.1 *Distance-dependent decay of interaction frequencies and calculation of the power-law decay*

To analyze the distance-dependent decay of intra-chromosomal interaction frequencies [156], all interaction values were binned according to their genomic distance and the average interaction value was calculated for each respective bin. The distance-dependent interaction value was plotted for each individual chromosome and averaged across all chromosomes. For these analyses, Hi-C matrices binned to 50 kb distances were used. To infer the power-law scaling, a linear function was fitted to the log-log values of the decay function for distances below 0.25 Mb using Prism GraphPad. This analysis was done on each chromosome arm separately and for core and subtelomeric regions, respectively. Average inter-chromosomal interaction frequencies were determined by calculating the mean inter-chromosomal interaction frequency across all core or subtelomeres, respectively, from a Hi-C interaction matrix at 50 kb resolution.

8.3.2.2 *Generation of 4C-like interaction profiles*

To generate 4C-like interaction profiles, genome-wide interactions of a respective genomic region (bait) from a Hi-C matrix of choice (here 20-kb resolution were extracted. All bins overlapping with the bait region were selected from the matrix. In case the bait region spanned multiple bins, an average interaction value across all bins was calculated.

8.3.2.3 *Colocalization analysis using background modeling*

To investigate *trans*-interactions between different genomic loci (this can be single genes or broader regions such as centromeres or gene arrays), we developed a custom python script to select regions of interest from a normalized (ploidy-normalized and ICEd, see Section 8.3.1) Hi-C interaction matrix (50-kb bin size). Thereby, the genome annotation file was searched for selective entries and the corresponding genomic regions were translated into the overlapping genomic bins. Alternatively to screening the genome annotation file for keywords, a list with regions of interest could be fed into the tool in gff-format. Inter-chromosomal interaction frequencies between these genomic bins were plotted. As a control, a sample that mirrored the selection of genomic bins in size and distribution was randomly selected from the same matrix, considering likewise only inter-chromosomal interactions. For both sets, zero values were excluded and the mean and median interaction frequencies were determined. The genomic interaction between the bins of interest over background was tested using a two-sided Welch's *t*-test.

Differential colocalization was used to compare distinct inter-chromosomal interactions in two different matrices. Here, the interaction frequencies of genomic loci of interest compared to a randomized background sample (as described above) in two matrices of choice were tested for significant differences in signal-over-background among them. We tested the query sample against a randomized background sample in 100 rounds of subsampling. In each round of subsampling, the ratio between average signal and average background was compiled, thus resulting in two respective populations of signal-over-background ratios (one for each matrix). The significance of difference between these two populations was assessed using Welch's *t*-test (two-sided).

8.3.2.4 *Visualization of Hi-C interaction matrices and sub-matrices*

For matrix visualization, we plotted the values of a homer-formatted matrix according to the color palettes provided by seaborn (https://seaborn.pydata.org). To visualize sub-matrices (e.g. individual chromosomes only), the respective region was beforehand excised from a genome-wide Hi-C matrix using a custom python script.

8.4 Recombination analysis using SMRTseq

To analyze the site of recombination in the ΔH3.VΔH4.V cells, gDNA was isolated from $3x20^8$ cells using the Blood & Cell Culture DNA Midi Kit (Qiagen) following the manufacturer's instructions. The genomic DNA was subsequently purified in an additional phenol-chlorophorm extraction step using Manual Phase Lock Gel tubes (5Prime, 2 ml, heavy/light) following an ethanol precipitation step. The sample was resuspended in 100 μl of TE buffer (Table 9.2). To prevent shearing of DNA fragments for long-read sequencing, vortexing of the sample was avoided during the extraction procedure. Preparation of the library and PacBio SMRT sequencing was performed at the Icahn School of Medicine at Mount Sinai (New York City, New York) using the PacBio Sequel system by diffusion at 5 pM (v2.1 chemistry, 10h movie).
To map the site of recombination between BES1 and BES15, we extracted reads >10 kb from the sequencing data and split them into 2.5 kb fragments. These fragments were mapped independently to the genome. We filtered for reads with fragments mapping to both BES1 and BES15 and re-mapped them in their full length against BES1 and BES15. By visual inspection of the mapping pattern, the breakpoints in BES1 and BES15 were identified and a virtual BES1-BES15 hybrid expression site was constructed.

8.5 Chromatin Immunoprecipitation with sequencing (ChIPseq)

In this study, all proteins subjected to ChIPseq experiments were Ty1-tagged. The ChIP experiment was conducted as described previously [275] with the following alterations: 50 μl of Dynabeads Protein G (ThermoFisher) were coupled with 10 μg of an anti-Ty1 antibody (BB2) [18] in 200 μl of 0.02% Tween in PBS with slow rotation at 4°C overnight. The Protein G beads coupled to the antibody were washed three times with 0.02% Tween in PBS on a magnetic rack and 500 μl of ChIP sample was subsequently added. The immunoprecipitation was performed at 4°C overnight with slow rotation. During the ChIP sample preparation, chromatin was fragmented using micrococcal nuclease (MNase) if histone variants were immunoprecipitated. Otherwise chromatin was sheared using a Covaris instrument (Covaris S220, 150 peak incident power (PIP), 5 watts, 10% duty factor, 200 cycles per burst, 600 seconds treatment)).
Library construction and paired-end sequencing was performed as described previously [276]. Paired-end reads were mapped using BWA-mem. Read counts were normalized and translated into coverage plots using COVERnant (v0.3.2) (https://github.com/konrad/COVERnant) [276].

8.6 Assay for transposase-accessible chromatin using sequencing (ATACseq)

To perform ATACseq, $3x10^7$ *T. brucei* cells were harvested, washed with 30 ml cold 1x TDB (Table 9.2)and the pellet was resuspended in 300 μl of permeabilization buffer (Table 9.2) supplied with protease inhibitors. Cells were permeabilized by addition of 3 μl of 4 mM Digitonin (Roth, 4005.1) and incubation for 5 minutes at RT. Cells were pelleted and resuspended in 600 μl isotonic buffer (Table 9.2 [163]) supplemented with protease inhibitors. Each sample was split for two independent assays of unequal cell number, containing $1x10^7$ and $2x10^7$ cells, respectively. Transposition was performed by adding 50 μl of Nextera TDE1 (Tn5) transposase in TD reaction buffer (Illumina) and incubation for 30 minutes at 37°C. gDNA controls to normalize for 'mappability' were included for wild-type and ΔH3.VΔH4.V cells, respectively, and were treated in the same manner.
For library generation, the samples were purified using the MinElute PCR Purification Kit (Qiagen) and were eluted in 10 μl of EB (Table 9.2). The DNA fragments were amplified using the NEBNext High-Fidelity 2X PCR Master Mix and 2.5 μl of 25 μM of barcoded primers (NEBNext multiplex oligos for Illumina, Ad1, Ad2.1-Ad2.10, Table 8.1) following the manufacturer's instructions. Amplification was conducted for 13 cycles. After purification of the library using AMPure

XP beads (Beckman Coulter), a size-selection step was performed by excising fragments of 150-1000 bp in size from a 6% polyacrylamide gel. The libraries were paired-end sequenced (2x76 bp) on the Illumina NextSeq 500 platform. Sequencing reads were mapped with BWA-mem and further processed using COVERnant (v0.3.2) (https://github.com/konrad/COVERnant) [276].

8.7 Single-cell RNAseq

For RNAseq on the single-cell level, *T. brucei* wild-type and ΔH3.VΔH4.V cells were sorted (0 cell, 1 cell, 50 cells) using forward-scatter (FCS-A) and side-scatter area (SCC-A) of a FACSAria III (BD Biosciences; precision: single-cell; nozzle: 100 μm). Cells were directly sorted in 2.6 μl of 1x lysis buffer (Takara) supplied with 0.01 μl of RNAse inhibitor (40U/μl, Takara) and frozen at $-80°$C.
For cDNA synthesis, the cell lysates were supplemented with 0.2 μl of $1{:}2x10^6$ or $1{:}2x10^7$ dilution of ERCC RNA Spike-In Control Mix 1 (ThermoFisher, 4456740). cDNA libraries were prepared using the SMART-Seq v4 Ultra Low Input RNA Kit (Takara) following the manufacturer's instructions but using only a quarter of the recommended reagent volumes. Libraries were PCR amplified using 26 or 22 cycles, respectively and following the manufacturer's recommendations. After purification of cDNA with Agencourt AMPure XP beads (Beckman Coulter), each library was taken up in 15 μl of elution buffer (Takara) and was quantified using Qubit High Sensitivity DNA Assay (Life Technologies). Library qualities were assessed with a Bioanalyzer (Agilent). 0.5 ng of cDNA were subjected to tagmentation-based adapter ligation (Nextera XT, Illumina). Tagmentation was performed as recommended by the manufacturer but using only one quarter of the reagent volumes. Tagmentation was conducted for 10 minutes at 55°C and PCR amplification was performed with 1 minute of extension time. For paired-end sequencing (2 x 75 bp), 96 libraries were poolend and subjected to a NextSeq 500 apparatus (Illumina).

8.7.1 *Single-cell RNAseq data analysis*

For analysis of single-cell RNAseq data, reads were mapped with BWA-mem (version 0.7.16) to the Tb427v9 genome including ERCC spike-in sequences. To identify PCR duplication artifacts, the mapped data was processed with SAMtools (version 1.8) [153] and MarkDuplicates (Picard, version 2.18.3-SNAPSHOT, http://broadinstitute.github.io/picard/). Read counts for distinct features were assessed using BEDtools (version 2.25.0) [220]. To assess library quality, reads mapping to rRNA genes and protein-coding genes (CDS including 89 bp of the 5′-UTR and 400 bp of the 3′-UTR; as defined in [246]), respectively, were determined. Reads that did not map or mapped to both of the above-

mentioned features were categorized into 'mapping to other regions of the genome'. Reads were designated as 'unmapped', if reads could not be mapped at all. Per single cell, we determined the number of genes with more than 10 read counts and for subsequent analysis of *VSG* expression, we only included libraries with sufficient complexity, e.g. containing more than 500 genes with each more than 10 read counts. To account for the high degree of homology shared between different *VSG* genes, we determined uniqueness of each locus by mapping *in silico* generated reads to the genome and normalized scRNAseq transcript levels with this 'uniqueness' score. Transcript levels of each individual *VSG* gene per cell is indicated as percentage share of all *VSG* transcripts.

8.8 Total RNAseq including non-polyadenylated transcripts

Total RNAseq was conducted in triplicates in *T. brucei* wild-type, ΔH4.V, ΔH3.V and ΔH3.VΔH4.V (T2 and T3, supplementary Figure a.14 a) cells. Therefore, cells were grown to densities below $1x10^6$ cells/ml and prior to harvesting of $4.5x10^7$ cells, cell concentration was determined nine times with a Coulter Cell Counter (Beckman Coulter) per replicate. Cells were washed in 1x TDB and resuspended in 350 μl RA1 buffer (Macherey Nagel, NucleoSpin RNA extraction kit) and 38 μl of 0.1 M DTT and 1 μl of 1:10 diluted ERCC RNA Spike-In control Mix 1 (ThermoFisher, 4456740) were added. After RNA purification following the manufacturer's instructions, the RNA was taken up in 30 μl of nuclease-free water. rRNA was depleted by hybridization of 131 50-bp anti-rRNA oligonucleotides (covering 18S rRNA, M1, M2, M3, M4, M5, M6 28S alpha and 28 beta, kindly gifted by Christine Clayton, see Table 8.2) to the rRNA sequences. Therefore, 2.6 μl of 5x hybridization buffer (Table 9.2; 500 mM Tris-HCl pH 7.4, 1 M NaCl), 0.459 pmol of anti-rRNA oligonucleotides and 2.2 μl of nuclease-free water were mixed, denatured at 95°C for 2 minutes and subsequently cooled to 37°C (0.1°C/second). The RNA-DNA hybrids were removed from the sample by digestion with 2 U of RNAseH (ThermoFisher, AM2292) in 50 μl of 1x Turbo DNAse reaction buffer (AM2238, ThermoFisher) for 20 minutes at 37°C. The reaction was inactivated by addition of EDTA (15 mM final concentration) and incubation at 75°C for 10 minutes. The rRNA-depleted RNA sample was purified using RNAeasy Minelute columns (Qiagen) according to the manufacturer's instructions. The sample was eluted in 14 μl of nuclease-free water. cDNA synthesis was conducted by subjecting 100 ng of rRNA-depleted total RNA to the NEBNext Ultra RNA Library Prep Kit for Illumina sequencing (NEB). Adapter ligation to the cDNA fragments and sequencing on a NextSeq 500 Illumina platform was conducted as described previously [275].

Table 8.2: rRNA depletion oligonucleotides

Target rRNA	Sequence (5′ -3′ end)
anti-18S	
1-50	GACTTTTGCTTCCTCTATTGAAGCAATATCGGTGAACTTTCGGTCGCCCG
51-100	AGACGAGCGACGGTCCGATCACCTGTATTGCACCATCATCGGAAACAACG
101-150	ACGGGCGGTGTGTACAAATGGCAGGGACGTAATCGGCACAGTTTGATGAG
151-200	CTGCGCCTACGAGACATTCCTCGTTGCGCGACCAATAATTGCAATACTCG
201-250	GTCCCAAGTGGGTCTACGTGATTCCGTGGTTTCCCCGCTCTTTTGATCAA
251-300	GCGGGAGCGACATTGTGAAGTGCCGCTCGGACTCTTGTTCTCACTGACAT
301-350	TGTAGTGCGCGTGTCGCCCAGAACATTGAGGAGCATCACAGACCTGCTGT
351-400	TGCCCAAAATCTCACCTTGCGCGAAAAGCAAGGAATCCCGCAGAGAAGGG
401-450	GCCTGTAAAAAGAAGGATGACGACCACCGTTGGGGAAAAAATCCCCGCCG
451-500	GAGGGACTGCTGTCCTGTGTGGACAATCCCGGCACCTACTGGGCAGCTTG
501-550	GATCTCGTCCGTTGACGGAATCAACCAAACAAATCACTCCACCGACCAAA
551-600	AGCGGCCATGCACCACCATTCAGGGGATCGAGAAAGAACACTCCATCTGT
601-650	CAATCCTCACCCTGTCCGGATCTGGTAAAGTTCCCCGTGTTGAGTCAAAT
651-700	TAAACCGCACGCTCCACGTCTTGTGGTGCCATTCCGTCAATTTCTTTAAG
701-750	TTTCACTCTTGCGAACGTACTCCCCCCTGAGACTGTAACCTCCAAGCTTT
751-800	CGCGTGAAGAAGTAAGAAACGTGCTGAGGATATTCCCGTAAAAAACCCCG
801-850	GAAGAACCTCAAAAGAGGGAACCGGGGGTGATAGAATAGGCACGTAAATT
851-900	GATAAATAAATTGGGGCCGAGGGTCGGCACAAAAGGCCGCCACACGGGCA
901-950	CCCAATGATGTTCCCCAATTCATGGGTGTCATGGTTTGCAGTGTGGACTA
951-1000	CAATGGTCTCTAATCATCTTTGATCCCCACACTTTGGTTCTTGATTGAGG
1001-1050	AAGGTATCCTTGAAGAATGCCTTCGCTGTAGTTCGTCTTGGTGCGGTCTA
1051-1100	AGAATTTCACCTCTGACGCACCAGTACGTTCTCCCCCGAACTACCCTCCT
1101-1150	TCATTCCTGGAAGCCGCGCGTCGCAATGAATGTAAAAAAACATCACGAAG
1151-1200	AGCCGAATGTGCGCGCCACAAAAATGAGACACGAAACACCGACCCAAGGC
1201-1250	ACGCGGCCCCATAATCTCCCAAGGACTTCTAAAACCAACAAAAGCCGAAA
1251-1300	CGGTGGCCCCAGGGCTGATGCTTCGTTATCCCATGCTTTGTAATTCAAGT
1301-1350	CGGCAGACTGCTTTGGTCGCACTTTTTTGGTCACAGTAAAATTCACGGAC
1351-1400	GCCCCCTCGCATGCATGACATGCGTGAAAGTGAGGAGGAACCGCTCCCGT
1401-1450	GTTTCTTGTATCTGAGTGTGTTCACGGGCCACGAGGGCGTGGGTCAAAAC
1451-1500	ACGTCCGAGGTGGACGGGACTGGCACGGCACAAAACTACGTGGCCCACAG
1501-1550	TTCAACTACGAACCCTTTAACAGCAACAGCATTAATATACGCTTTTGGAG
1551-1600	CTGGAATTACCGCGGGTGCTGGCACCAGACTTGTCCTCCAATTGTTACTC
1601-1650	GATATTGGATGGGTTTGAGTATCCCCCATTGAAAACAACCATGCACTAGG
1651-1700	GCACGGTCGGCTCTATTTCTTTTCGCTGCCTCATCGTATTTTTTCGACAT
1701-1750	TGGGCAATTTGCGCGCCTGCTGCCCTCCGTAGAAGTGGTAGCTATTTCTC
1751-1800	AGGCTCCCTCTCCGGAATCGAACCCTAATCCCCCGCTCCCGTCAACGCCA
1801-1850	TGGCAGTCCACTACACGGCCATCACTGGCTGATAGGGCAGTTGTTCGTCA
1851-1900	GTAGACGCCGAGCAACAAGGCTCGGCTTTCGCACAGAATGAATTGAGTCA
1901-1950	ACACGTCTGAGTGTTACTGCCCCCGGGTGGAGGCGTTGGTTCTAATTTCA
1951-2000	TTCATTCGCTGGACGTCCCATCGCCACGACAAGTGCGTAAACACAAGCCG
2001-2050	CGACATAGAAAAAGAGAGCGTCCGATTGGTTCATGTATTAGCTTGGCGTT
2051-2100	TCGCCAAGTTATCCAGTAATGCGGAGACCGCAAAATTTTGGCGGCAGATT
2101-2150	ACGTCTGATGTAATGAGCCATGCGCAGATTCCTGCAATGCAGTGATTCTG
2151-2188	AGGCATGCATGGCTAAGTCCTTGAAACAAGCATATGAC
anti-M1	
1-50	ACACCCCAGGTTTTTGCTTTAATGCGCCGAACTCACAACAAAGGACAGTA

51-100	CGTAGGGTGAGCTGCATTTTTCAATCTTGAGCCCCCGCTCCGAGTAACGC
101-150	GAGAAGTCGAGAGGCCCAAGCCCCAGCAAGGCCGGAGATGTGGATTATGG
151-200	TGAAAAGAAAGCTCACCGTGGGAGGATGGATGGGTCACCTGGCAAGTGTT
anti-M2	
1-50	CGGGCACCCGGTCTTGGGACTGGTTTCTGGCTTAGAGGCGTTCAGCCGAG
51-100	AACCACCGTGCGATAGCGTCGGAGCACTTGCCCTTCGACAAACTCCAGAA
101-150	ACCAGAGGCACGGCATGGGAGTTCCTCTCGTACTACCCCATACTTCCCTC
151-182	ACGATGCCTGCGAGATCCCTTCACAATCTCAC
anti-M3	
1-50	ATTCGACACTGAGAAATGTGGCATGCACGGGGATGGCTCGAGAGAGCTTC
51-100	TCCCATGCGCCGTTTGCGTTCAAAGATTGGGCAATGAAATGATTCTGCAA
101-150	TTGATACCACTTATCGCACTTTGCTGCGTTCTTCAACGAAATAGGAAGCC
151-172	AAGTCATCCATCGCGACACGTT
anti-M4	
1-50	AAAAGTCCCTGAGCATGAAATTCCAAGCATCAAGGTTCAGCGCCTCGGGA
51-100	CGTAAAGTGCCCCTCGAGCCCCACGGTCGTGGAGCGTGCCGCTCATGCCA
101-136	GCCCATGCATGTACTCTCGTTTGGAGAGGGACAAAA
anti-M5	
1-50	AGAGTACAACACCCCGGGTTCCAGCGCCATCACTGATCGCCGTACTAACG
51-100	AGGCCTGTGGCCGCTTAACTTCACAAATCGGACGGGATATGGTGCATTCG
anti-M6	
1-50	GATCTTGTGAAAATGAAGGTACTGGGCGCCGACAAGGGCGGGCAGAGCCC
51-78	ACCAGATAAGAGAAGTGGCGATTCGAGA
anti-28S α	
1-50	TCCCACATGCAATTTCTTTTTGGAATATCATCCCAGGTTCTCTGTTGCCC
51-100	CTTTTCAGATAAGCCGGTAAAACACACAAACCCCAAAAAAGGGAGAAGTG
101-150	CGCCGACCACCTTCACTTTCCTAACAAAACGGTGTAACGTTGTGCTCAGT
151-200	GACCGAACAACTCGTGTACGAGTGCCGTTGGCACGAAATCCTTCCCCGCT
201-250	CCAGTATCCAACGTTGCATTTAAGAAAAAATATTCCGTTAGATCTTTGCT
251-300	ACGCCAACCAAGATCTGCACTAGAGGCGGGTCCATCCGAGGCTCGCGCCA
301-350	TCGGCTTCGTCCTCACCCCCACGCCACGTCTACATCGGGAGGGCTACACT
351-400	TTCTCACAAAGAAATATAGCCGCCCTCACCAAAGGAATAATGGGCAATCA
401-450	CGCACATGGGGCTTAAGCGTCATCCATTTTCGGGGCCGGTGGATTCGGTT
451-500	GGTGAGTTGTTACACACTCCTTAGCGGATGCCGACTTCCATGGCCACCGT
501-550	CCAACTGTCCCTATCCACCAACACCCTTTTCACGTATCGAGAATGAGCCC
551-600	ACAAGTTTTGACGCCACATATCCCTAAGCGGTTCATTCCGCATCGCCTGC
601-650	TCCGCTTACCGAAGAAGGCCCACTTGGTGCCTCAATTCTTATGTACCCAA
651-700	AAGCCTTGTTAAAATAACTAAGGCTGGGTTGTTTGCACATTTAAAGTTTG
701-750	AGAATAGGTCGATATTACAAGAACACCAATGCCTCTAATCATTTGCTTTA
751-800	CCGCACAATACTTCTTAAAGACAAGCACCAGCTATCCTGAGGGAAACTTC
801-850	GACGTGAACCAGCTACTAGGTGGTTCGATGAGTCTTTCGCTCCGATACCC
851-900	AAATCGTACGAGCGATTTGCACGTCAGCATCGCTACAGGCCTCCACCGGG
901-950	GTTTCCCCGGGCTTCACCTTGTTCAGGCATAGTTCACCACCTTTCGGGTC
951-1000	CAAACAGGCACACTCGCGCGCAGCCCACGGGGGGCCGGCCGGCGTTGGGC
1001-1050	AAAAACACAAAAAAGAGAAATTGCGCACGCCACGGGTGTAAAATCACACC
1051-1100	TCCCTTTCAGTACGAGAAAAGAGTTCTCACATACATACTCCTCCCTTGCG
1101-1150	CGTCTGTTTGACTCCTTGGTCCGTGTTTCAAGACGGGTGGGGGCACCCCA
1151-1200	TTTCCATCGCCAGCGTCGAGACGCACGGACAGGGAAGGAGGAAGAGAGTG
1201-1250	AGAGAAAAAACCCACACACAACACACCGACCCTTTATTCCGCCCACTCAC
1251-1300	GCACCCCATGACCTCAAAAGAGAAACCCGCACAGTGAAAACCCACCGGAT
1301-1350	GTTGAGCCCCGGAAGGAAATTGCGCGGCGAGGTTGGGTCTGGCCGTAAAA

1351-1400	AAAAAGAAATGGAGGTACTTCCATTGTCGTTGTCACGATTTCAGGTTCTA
1401-1450	TGTCACTCTCTTTCCAAAGTACTTTTCAAACTTTCGCTCACGCTACTTGT
1451-1500	TCGCTATCGGTCTTCCTACTCTATATTTAACCTTGGCAGACAAATACCTG
1501-1550	CCGGCTTGACACTGCATTCCCAAACAGTGCTACTCTCTTGGACCGAACAT
1551-1600	GAGGAGGCGCGTAGGCTCCGTGGCTAAGGAATGTAATGGGCTGACACCTT
1601-1650	CTTTGGCTCAATACTCACATGAAGGGTTGTAAATGTGTTCAGCAACGGGT
1651-1700	TTGCTTTGCAATTGAATTCCGCCACGCCTGCACCAGCCCTATAAGAACGG
1701-1750	CCACAAAAATGGTGCCACAGTTCGAGGTCATGTAAGACCCGATTCGGCAA
1751-1800	CGAGTTGAATCCCTCTTCGCTCGCCGCTGACTGAGAGAATCACGGTTGTT
1801-1863	TTCTTTTCCTCCGCTGAGTAATATGCTTAAGTTTAGCGGGTGGTCCTGCC
anti-28S β	
1-50	AGCGCCAAGCACATGCCACCGAACGGCAGACCCAGAAAAAAACGGACTGC
51-100	CTCCGCTTTGCGCGGATAATCTTTTACTCGCAATCTCCAGCCAAGTAGGG
101-150	TAAAACCAACCTGTCTCACGACGGTCTAAACCCAGCTCACGTTCCCTGTC
151-200	AGTGGGTGAACAATCCAACCCTTAGCGTCTTCTGCGGCGCTAGGTTAGGA
201-250	AGAGCCGACATCGAAGGATCAAAAAGCGACGTCGCTATGAACGCTTGGCC
251-300	GCCACAAGCCAGTTATCCCTGTGGTAATTATTCTGACACCTCTGGTATGA
301-350	TTTGTTCATATTTCCTGTTCTTCGCAAAGGATCGCTGGCCTTGCTTTCGC
351-400	AGTCTCGATTCGTACTGGAAATCGTTAATCAAGCAAGCATTTACGCTTGT
401-450	GTTCCACGTGAGATTTCTGTTCCCACTGAGCTCGCCTTAGGACACCTGCG
451-500	TTGTCGTGTAACAGATATGCCGCCCCAGCCAAACTCCCCATCTGGCAGAT
501-550	GTCTCGAAACAAGACTCCGGAGAAGAAACGAAAAGTTCCCACCCCGTTGG
551-600	CGCAGTAGTTCCGCACGGCTGGGCGACCACGGGTCACCAGCCCTCAAACG
601-650	CCCAAAAAAGGGAGAAAGAGAGAAAAGGAGAAACGGAGAGAAGGATAAAA
651-700	AACCAACACCCCTCCCGTGGCGGTGAAAGAAACACCACAATTATGCCCAG
701-750	CCTAGACGGCGCACAGCCCGAAGACTGCGCAAAACAACCATTGGTCTCTT
751-800	CATTCGATAAGTGAAGCAACGTTCCGAGTGGTGGTATTTCATTTGAGCTT
801-850	GCGCTTCCACCTATACTACACCTCCAAAGTCGCCGCACAGAGCCAGACTG
851-900	GAGTCAAACTCAACAGGGTCTTCTTTCCCCGCTTAGCCCTTCAAGTCCGT
901-950	TCCCTTGACTGTCGTGTCGCTAGATAGTAACTCGGGACAGAGGGAATCTC
951-1000	ATTAATCCATTCATGCGCGTCTCTAATTGGAAGACGAGGCATTTGGCTAC
1001-1050	CTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCCGCGCGTCGGTGAATC
1051-1100	TCGTCACGTTGACATTCAGAGCACTGGGCAGAAATCAGTTTGCGACATAC
1101-1150	ACCACTTGGATGCATCACAAACCTATGTTTTAATTATACAGTTGGATTCC
1151-1200	CCGCCGTAAGCAGTTCTGAGTTGGCAATTTGTAGGTGCGCAACCAAAAAA
1201-1250	GGAAGCGAAAACACCTGCGCGGGAGTGAGACGTAAAAACAACACACCAGA
1251-1300	AGGAGGTTAGTAGATACGGCTCGGGAGGGTAGCCTTCACGTTCCGAGCAA
1301-1350	AACGGTCGCCGCGTATGTGCTAAGGCGCGAACGCCAAAACAACAAAACGC
1351-1400	AGAAAACGGCCGCGCACGAAAGGACACTGGCCCTCCTGTGGACCCCCACT
1401-1450	TTGCCCTCAGAGCCAATGCTTCTTGCGAACTTACGCATCTATTTTGCCGA
1451-1500	CTTCCCTGAGTGTACCTTTCTCTATGTACCAGAGCATTCGATGCTTGGGG
1501-1536	ACCTGATGCGGTGATGAGTACGGTCTGCAGTTGGGA

DNA oligonucleotides targeting 18S rRNA, M1, M2, M3, M4, M5, M6 28S alpha and 28 beta rRNAs (kind gift of Christine Clayton).

8.9 RNAseq data analysis

RNAseq reads were clipped to remove the adapter sequence using cutadapt (version 1.10) [172]. Clipped reads were mapped against the

T. brucei genome using BWA-mem (version 0.7.16) retaining uniquely mapped reads only (MAPQ>0). Quantification of feature mapping was performed using the BEDtools multicov subcommand following a differential gene expression analysis using DESeq2 (v.1.20.0) [161]. In subsequent analyses we considered features with an adjusted *p*-values p_{adj}<0.1 [24] as differentially expressed. Transcription profiles were generated using COVERnant (v0.3.2) (https://github.com/konrad/COVERnant) [276].

Table 8.3: Publicly available programs applied for NGS-data processing.

Program	**Version**	**Reference**
bcl2fastq	v.20.0.422	Illumina
cutadapt	v.1.15	Martin et al., 2011
HICUP	v0.5.9 devel	Wingett et al., 2015
BWA-mem	v0.7.16	Li et al., 2010
HiC-Pro	v2.10.0	Servant et al., 2015
HiCsd	v0.1.0	Müller et al., 2018
DeSeq2	v1.20.0	Love et al., 2015
R	V3.5.0	http://www.R-project.org
python	3.6	http://www.python.org
SAMtools	v1.8	Li et al., 2009
BEDtools	v2.26.0	Quinlan et al., 2010
picard	v2.9.0	https://broadinstitute.github.io/picard/
COVERnant	v0.3.2	https://github.com/konrad/COVERnant
Wiggle Tools	v1.2.2	https://github.com/Ensembl/WiggleTools
Minimap	2.1	Li et al., 2016
Fastaq	3.17.0	https://github.com/sanger-pathogens/Fastaq

9

Biochemical approaches

9.1 Fluorescence *in situ* hybridization (FISH)

Cells were grown to densities below $1x10^6$ cells/ml. Per assay, $1x10^7$ cells were harvested, washed with 1x TDB and fixed for 5 minutes in the presence of 4% formaldehyde. The cells were washed again with 1x TDB, resuspended in 50 μl of 1x TDB and let settle on a framed coverslip. Subsequently the coverslip was washed twice for 2 minutes with 90 μl of TDB and the sample was then incubated with quenching solution (Table 9.2; 1 mg/ml $NaBH_4$ in TDB) for 10 minutes at RT. The sample was then washed twice with TDB and cells were permeabilized by addition of 70 μl of 0.1% NP-40 in TDB for 5 minutes. The permeabilization solution was removed in two TDB washing steps. Then, the sample was treated with 1 mg/ml RNAseA in PBS for 30 minutes. The sample was washed with 1x TDB and was equilibrated by incubation at RT for 30 minutes with 50 μl of hybridization buffer (Table 9.2; 50 (v/v) formamide, 10% (w/v) dextran sulfate, 2x SSPE, 250 μg/mL herring sperm DNA) diluted 1:1 with 2x SSC (Invitrogen). Digoxigenin-labeled probes (Telomere-Dig or an equimolar mixture of nine different probes targeting centromeric repeats, see Table 9.1) were diluted to 400 nM in 25 μl of hybridization buffer and added to the framed sample. The frame was sealed with a plastic lid and incubated for 5 minutes on a thermal block heated to 90°C. Incubation was continued at 37°C overnight in a wet chamber.
The following day, samples were washed in 30 ml of 50% deionized formamide in 2X SCC1 at 37°C for 30 minutes, then for 10 minutes in 30 ml of 1x SSC at 50°C, following incubation for 10 minutes in 30 ml of 2x SSC at 50°C and finally in 30 ml of 4x SSC for 10 minutes at RT (all incubation steps were done in a falcon tube). Then, cells were blocked with P1 buffer (100 mM maleic acid, 150 mM sodium chloride, pH 7.5, 4% BSA, 1% milk) [80] for 1 hour and then incubated with the primary anti-digoxigenin antibody (sheep anti-digoxigenin Fab fragment, Roche, diluted 1:2000 in P1) for 60 minutes. The sample was washed with 0.5% Tween-20 in PBS four times for four minutes each. Then, the sample was incubated with the fluorophore-labeled secondary antibody (Alexa Fluor 488 conjugated donkey anti-Sheep IgG (H+L), Life Technologies, A11015, diluted 1:2000 in P1) for 30-60 minutes at RT in the dark. For amplification of the signal, the samples were washed with 0.5% Tween-20 in PBS four times and incubated with rabbit anti-Donkey IgG (H+L) DyLight 488 (Invitrogen) (diluted 1:2000 in P1) for another 30-60 minutes. Finally, samples were washed

twice in 30 ml of 0.5% Tween-20 in PBS for 10 minutes each and for 10 minutes in 30 ml of PBS while shaking. The samples were then mounted using Vectashield Mounting Medium with DAPI (Biozol) on a microscopy slide and sealed using nail polish.

Table 9.1: Probes for fluorescence *in situ* hybridization.

FISH probe	**Sequence (5′ -3′ end)**
cen_11_3A-Dig	Dig/N/TGCACATTATTGCCTATTATTGCACATTAT/N/Dig
cen_10_3A-Dig	Dig/N/AATAATGTGCATTTTACGCAATAATGT/N/Dig
cen_9_3A-Dig	Dig/N/CAATAATGCACATTTTACGCCATAATGTGCAATAAT/N/Dig
cen_9_3B-Dig	Dig/N/AGTCAACAATATGCACTTTACGCCATAATATGCAATAAT/N/Dig
cen_8-Dig	Dig/N/AAAAACAGTGTTGCAATGCAACATGTAA/N/Dig
cen_5-Dig	Dig/N/GCAACATTGTGCAGGTTGTGCAAGAATGCGTGTTTTACAC/N/Dig
cen_2-Dig	Dig/N/TTTATGTACAAAAGCATGTTATTGCTTGT/N/Dig
cen_1-Dig	Dig/N/CACATTAATGCACACATATGCGTGCTTA/N/Dig
cen_4-Dig	Dig/N/CGTATTTGACGTGAAAATACGCAATAGTGCA/N/Dig
Telomere-Dig	Dig/N/TTAGGGTTAGGGTTAGGGTTAGGGTTAGGG

Digoxigenin-labelled probes used for fluorescence *in situ* hybridization. *, phosphothioate bond; Dig, digoxigenin; N, C6- (5′-end) or C7- (3′-end) linker.

9.2 Immunofluorescence

For Immunofluorescence, I followed a previously published protocol [244] with minor alterations. Per assay, $1x10^7$ cells were harvested and crosslinked in the presence of 2% formaldehyde in HMI-11 for 5 minutes at RT. Cells were then washed with TDB and let settle on a coverslip. To determine *VSG* expression status, cell bodies were first stained with the previously published monoclonal antibody Tat1 [283] in a 1:200 dilution against α-tubulin. As secondary antibody, a 1:350 dilution of an Alexa Fluor 594 conjugated chicken anti-mouse IgG (Invitrogen) was used. A previously published CRD-depleted rabbit anti-VSG-2 antibody [83] together with a secondary Alexa Fluor 488 conjugated donkey anti-rabbit IgG (1:350, Invitrogen) was used for staining of VSG-2 on the cell surface.
To stain the Ty1-tagged H3.V, H4.V and Scc1 proteins, the BB2 antibody [18] was used in a 1:400 dilution in PBG and detected with a secondary Alexa Fluor 488 conjugated donkey anti-mouse antibody (Invitrogen, diluted 1:500 in PBG).

9.3 Fluorescence microscopy of FISH and *VSG* IF

Tagged constructs were imaged as stacks with 30-40 slices (0.3 μm slice thickness) using a Leica confocal microscope (DMI8-CS, HC PL APO CS2 63x, 1.40 OIL objective).
A DMI6000 wide-field fluorescence Leica microscope with a mercury metal halide lamp was used for imaging of VSG IF and FISH images

(HCX PL APO CS 100x/1.47 OIL objective, DFC 360 FX camera). Images were taken as stacks with 32 slices of 6.3232 μm in height (0.1976 μm slice thickness).
To quantify the occurrence of large telomere clusters, we used Imaris 8 software (Oxford Instruments) as follows: Nuclei were segmented in the DAPI channel and surfaces were rendered for the FISH signal (quality filter >1000). Telomere clusters were accounted as 'large' when measured as greater than 0.3 μm^3.

9.4 Western blot on Ty1-tagged proteins

To detect Ty1-fusion proteins by western blot, 5 million cells were loaded on a 15% polyacrylamide gel with SDS and run at 25 mA. The proteins were transferred to a nitrocellulose membrane using a wet blot transfer (Mini Trans-blot, BioRad, at 100V for 1,5 hours). The membrane was blocked in 3% BSA in 0.1% PBS-Tween. The membrane was incubated with the BB2 antibody [18] (1:1,000 dilution in 1% milk in 0.1% PBS-Tween) at 4°C overnight, washed and incubated with a secondary horseradish anti-mouse antibody for 1.5 hours at RT (1:10,000 dilution in 1% milk in 0.1% PBS-Tween). Detection was performed using Pierce ECL Western Blotting Substrate (Thermo Scientific) and exposure for 5 minutes (Fusion FX Vilber Lourmat, Peqlab).

9.5 Flow cytometry to determine *VSG* expression status

For determination of the cell surface VSGs we followed a previously published protocol [83]. Therefore, $1x10^6$ cells were harvested (1,500 xg, 4 minutes at 4°C) and cells were resuspended in 100 μl of ice-cold HMI-11 supplemented with an anti-VSG-2 or anti-VSG-13 antibody [83] was added. Incubation was conducted for 60 minutes at 4°C while shaking. Cells were then washed three times with ice-cold HMI-11 following an incubation step for 20 minutes in 100 μl of cold HMI-11 supplemented with a secondary Alexa-Fluor 488-conjugated secondary antibody. Finally, cells were washed with 1x TDB and resuspended in 400 μl of 1x TDB prior to subjecting the cells to a FACSort flow cytometer (Becton Dickinson Biosciences).

9.6 Flow cytometry to determine cell-cycle profiles

Cell cycle profiles were determined by harvesting $5x10^6$ cells (10 minutes, 1300 xg, 4°C), washing the pellet once in ice-cold TDB and fixation of the cells by resuspension in 1 ml of ice-cold 2 mM EDTA in PBS and addition of 2.5 ml ice-cold methanol. Cells were incubated for 1 hour, then washed with 2 mM EDTA in PBS at RT and resuspendend

again in 2 mM EDTA in PBS. Then 1 μl of RNAseA (10 μg/μl) and 10 μl propidium iodide (1 μg/μl) were added and incubated for 30 minutes at 37°C. Cell cycle profiles were analyzed using a FACSCalibur (Becton Dickinson Biosciences).

Table 9.2: List of non-commercial buffers used in this study.

Buffer	Conc.	Reagents
BB (Hi-C)	2x	10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl
Elution buffer (EB)	1x	10 mM Tris-HCl pH 8.0
End-repair mix (Hi-C)	1x	1x Ligation buffer (NEB), 357 μM dNTPs, 25U T4 PNK (NEB, M0201), 7.5U T4 DNA polymerase I (NEB, M0203), 2.5U DNA polymerase I, large (Klenow) fragment (NEB, M0210)
Formaldehyde solution	4x	50 mM Hepes-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 11% formaldehyde
Hybridization buffer (FISH)	1x	50 (v/v) formamide, 10% (w/v) dextran sulfate, 2x SSPE, 250 μg/mL herring sperm DNA
Hybridization buffer (RNAseq)	5x	500 mM Tris-HCl pH 7.4, 1 M NaCl
Fill-in mix (Hi-C)	1x	0.25 mM biotin-14-dATP (Life Technologies, 19524016), 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP, 40 U T4 DNA Polymerase I large Klenow Fragment (NEB, M0210)
Isotonic buffer (ATACseq)	1x	100 M KCl, 10 mM Tris pH 8.0, 10 mM $CaCl_2$, 5% glycerol (store at -20)
Laemmli sample buffer	4x	40 % Glycerol, 0.02% bromophenolblue, 8% SDS, 250 mM Tris, pH 6.5
Ligation mix (Hi-C)	1x	1.8% Triton-X 100, 0.027 % (w/v) BSA, 1.8x T4 DNA Ligase Buffer (Invitrogen, 46300018) and 5 U of T4 DNA ligase (invitrogen, 15224025)
Lysis buffer (Western blot)	1x	1x Laemmli sample buffer, RIPA buffer, 2 mM DTT, 1% (v/v) β-Mercaptoethanol
Permeabilization buffer	1x	100 mM KCl, 10 mM Tris pH 8.0, 25 mM EDTA
P1 (FISH)	1x	100 mM maleic acid, 150 mM NaCl, pH 7.5; add freshly 4% BSA, 1% milk
RIPA (Western blot)	1x	50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.25% Na-Deocycholate, 0.1% SDS
Separation buffer (Western blot)	4x	1.5 M Tris-HCl, pH 8.8
Stacking buffer (Western blot)	4x	0.5 M Tris-HCl, pH 6.8
Running buffer (Western blot)	10x	15.15 g/l Tris, 72 g/l Glycin, 5 g/l SDS
Quenching solution (FISH)	1x	1 mg/ml $NaBH_4$ in 1x TDB
TDB	1x	0.005 M KCl, 0.08 M NaCl, 0.001 M $MgSO_8$ x7H_2O, 0.02 M Na_2PO_4, 0.002 M Na_2HPO_4 x2H_2O, 0.02 M glucose
TE	1x	10 mM Tris-HCl, pH 8.0, 1 mM EDTA
TLE (Hi-C)	1x	10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA
TWB (Hi-C)	1x	5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20
Transfer buffer (Western blot)	1x	1.5 g/l Tris, 7.2 g/l glycin, 10% (v/v) methanol

10

Data and source code availability

All sequencing and the final processed data obtained from RNAseq, scRNAseq, ChIPseq, ATACseq and Hi-C are deposited with the according published manuscript [194] in NCBI's Gene Expression Omnibus [75] and can be accessed via the GEO Series number GSE100896 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100896). Applied workflows, Unix Shell, Python and R scripts are deposited at Zenodo (https://doi.org/10.5281/zenodo.823672) together with instructions on how to reproduce the data analysis. See Table 8.3 for a list of all publicly available software used for data analysis in this study.

Part IV

Discussion

11

Establishing Hi-C in Trypanosomes

11.1 Are the Hi-C libraries obtained for *T. brucei* nuclei of sufficient quality and complexity?

High quality of Hi-C data is critical for deducing correct and informative DNA interaction behavior from it. Quality of Hi-C data correlates, as a first metric, with a high mapping rate of Hi-C reads [240]. After removal of unmapped reads and reads mapping to the unassembled portion of 'unitigs', approximately ~80% of sequenced reads remained. Due to the high degree of repetitiveness of the *T. brucei* genome, approximately 70% of these reads in a respective Hi-C experiment could be mapped as uniquely mapping pairs to the genome, the remainder were discarded as multi-mappers (for all mapping statistics see supplementary Figures a.3 - a.9). Of those, between ~75% and 96% remained as uniquely mapped valid read pairs after removal of invalid interactions. Taken together, after removal of reads mapping to unitigs, mapping non-uniquely and representing invalid interactions, only ~50% of all paired-end reads remain as valid interactions for downstream analysis. Still, this reflects a high proportion of captured valid interactions and is in line with recently published Hi-C data conducted in apicomplexan parasites [39].
A second metric for a high quality Hi-C library is the over-representation of intra-chromosomal interactions that become apparent as the strong diagonal in an interaction heat map. A typical high-resolution Hi-C experiment in human cells yields ~40% of intra-chromosomal interactions [143] [240]. Our data sets are in this range since after duplicate removal, approximately 47% of all valid interactions were *cis*-interactions (averaged over three Hi-C experiments in wild-type *T. brucei*). This value still represents and underestimation since only one of the two heterozygous subtelomeres can be attached to the core, causing that *cis*-interactions occurring between the remainder subtelomere and its respective core will be measured as *trans*-interactions.
Library complexity is foremost determined by the fraction of informative molecules in the Hi-C sample. The efficiencies of digestion, fill-in, re-ligation and biotin pull-down are key to capture meaningful contacts. A Hi-C sample contains invalid interactions (e.g. dangling ends and self-circularized digestion fragments) and valid interactions representing both true interactions and products of random ligation. Valid interactions that origin from random ligation cannot be discerned from true interactions computationally. The noise produced by these interactions was reduced by applying *in situ* Hi-C to *T. brucei*

nuclei based on well-established, previously published protocols for permeabilization of the nuclear membrane [198] [275].
We used Hi-C Pro [240] to detect invalid interactions and filter them computationally from the sequencing data. According to Belton and colleagues [23], a Hi-C library typically comprises 0.5-5% of self-circles and 10-45% of dangling ends. Whereas a high proportion of self-circles reflects insufficient cross-linking, a high proportion of dangling ends accounts for insufficient biotin removal from unligated ends [23]. Our Hi-C libraries typically contained 0.14% of self-circles and 6% of dangling ends, which is far below the tolerated maximum threshold values.
If multiple read pairs map end-to-end to the exact same genomic position, only one of them is considered in following analyses as the remainder represents over-amplification products. However, a high amount of redundant PCR duplicates (> 5%) accounts for low molecular complexity and a the presence of a possible PCR bias [240] [23].
Our first attempts to generate high-resolution Hi-C matrices revealed an extremely high proportion of PCR duplicates (71% of all uniquely mapped reads on average), pointing towards a very poor library complexity. Adjustments in the handling of the Hi-C sample and especially the sample splitting for the final library amplification were key to reduce PCR duplicates. A titration of the Hi-C sample [22] revealed a shift towards higher molecular weights of the PCR products after ~10 cycles. At this point, the amplification reaction is oversaturated and oligonucleotides are exhausted. I amplified all libraries with as few as 5 cycles in eight separate reactions (supplementary Figure a.2 a), which reduced PCR duplicates down to ~1% and yielded between ~0.5 and 3 μg for sequencing. As the amount of PCR duplicates is very low in the final library at a sequencing depth of ~$1x10^8$ reads, I assume that cycle number could still be further reduced and samples could be sequenced even deeper.
In summary, the Hi-C libraries generated here are of high quality and great complexity. The amounts of invalid ligation products were below the generally advised threshold, although this of course varies with organism and application. Despite the high degree of repetitiveness of *T. brucei*, even low complexity regions such as subtelomeres were well represented in the interaction matrices. To improve these regions of high repetitiveness even further, I would suggest to increase read length during the paired-end sequencing, which is currently limited to 75 bp. In this way, more reads corresponding to repetitive regions should be captured in the mapping process and fewer genomic bins should be discarded due to insufficient signal.

11.2 What is the resolution of the *T. brucei* contact matrices?

The effective resolution of an interaction matrix is difficult to determine if the expected interaction patterns are not known *a priori*. Therefore, Rao and colleagues introduced the term 'map resolution' which describes the finest scale at which a distinct local interaction behavior can reliably be determined [222]. The 'matrix resolution', in turn, refers to the size of genomic bins used to construct a contact matrix. To obtain meaningful interaction matrices, it is advised to choose the smallest bin size such that 80% of bins obtain a minimum of 1,000 contacts [222]. To meet this criterion, a 10 kb resolution matrix for a human genome ($3.2x10^9$ bp) would demand ~$3.2x10^8$ reads. The *T. brucei* genome is two magnitudes smaller ($4x10^7$ bp), already due to the two magnitudes smaller $4x10^6$ reads would suffice to obtain a similar resolution. We sequenced our libraries to a depth of ~$1x10^8$ reads. Considering that only half of the reads map uniquely, even $5x10^7$ reads would theoretically yield a resolution of ~1 kb. However, such an assumed linear relationship between resolution and sequencing depth is not given for two-dimensional data [12].
We performed our analyses more 'conservatively' at 10, 20, 50 and 100 kb resolution to also check for the robustness of interactions across a range of binning sizes. Whereas TADs (0.2-2 Mb) have only been observed in metazoans, different types of self-interacting domains have been observed in non-metazoans, including bacteria (CIDs up to 300 kb in size, see **??**) and yeast (50-100 kb globules [188]). Assuming a comparable size of self-interacting domains existed in *T. brucei*, we should have been able to pick them up at 10 kb resolution. Still, given that our libraries are not yet sequenced until saturation, deeper sequencing combined with smaller bin sizes could be informative about folding principles at very small scales at *T. brucei*. Attempts to increase resolution further by exchanging the MboI restriction enzyme e.g. by using DNAseI or micrococcal nuclease could lead to a better representation of (repeat) regions, where the MboI recognition site is underrepresented in the DNA sequence. However, the generation of such small DNA fragments would not allow to improve read mapping by increasing sequencing read length as suggested in section 11.1.
Furthermore, as can be seen in the supplementary Figures a.3-a.9, most of the invalid pairs in the Hi-C matrices generated in this study are (unligated) dangling ends, ranging from 2.7-25.4% of all read pairs. Reducing the fraction of these uninformative reads from the library would effectively increase the amount of valid pairs. Additionally, if digestion is incomplete, a small subset of these reads may falsely appear as interactions between neighboring restriction fragments [22]. Dangling ends can effectively be reduced by improving the biotin removal step in the Hi-C protocol. As dangling ends do not contain a

valid biotinylated ligation junction, insufficient biotin-removal from fragment ends will include dangling ends in the biotin pull-down step. Therefore, Belagzhal and colleagues propose to improve biotin removal from fragment ends by promoting the exonuclease activity of T4 DNA polymerase over its polymerase activity. This can be achieved by providing an incomplete set of dNTPs at low concentrations in the reaction mixture [22]. In my Hi-C protocol, biotin removal is conducted during the end-repair step. It may therefore be advisable to tear these two processes apart and to more 'rigorously' perform the biotin removal step.

12

The Lister 427 genome in three dimensions

12.1 *T. brucei* chromosomes fold as a fractal globule but domain-like structures remain elusive

Analysis of interaction frequencies in dependence of genomic distance revealed a power-law decay of *P(s)*=-0.88 for subtelomeric and *P(s)*=-0.93 for core regions at distances below 0.25 Mb, respectively. This is in agreement with the observation that active regions of the genome are less compacted and therefore exhibit a more rapid decay of interaction frequencies as compared to repressed regions [64].
This pattern is further consistent with a fractal globule (*P(s)*=-1) rather than an equilibrium globule (*P(s)*=-1.5) as it is e.g. observed in yeast [74] and *Babesia* [39]. A fractal-like folding behavior has previously been reported chromosomes of larger mammalian nuclei [156], but also for the protozoan parasites *T. gondii* [39] and *P. falciparum* [149] [13] [38]. The latter is controversial, since Bunnik and colleagues found Plasmodium chromosomes rather stretched out as in budding yeast [38], whereas Lemieux and colleagues described a power-law decay of -0.96 for distances below 0.5 Mb [149]. Our data indicated a territorial organization of *T. brucei* megabase chromosomes, however, no finer sub-compartmentalization of the housekeeping portion of the genome, e.g. into TADs was apparent (see 11.1). This is expected since classical TADs are restricted to metazoans [106] and gene regulation in *T. brucei* does not rely on distal regulatory elements such as enhancers that may interact with their target genes by looping interactions in *cis* or *trans*. This is in agreement with a secondary loss of TADs in *C. elegans* autosomes where genes are transcribed as polycistrons and enhancers seem not involved in developmental gene regulation [64].
A fractal-like genome organization and the presence of CTs in the absence of TADs has recently been described for *T. gondii* [39]. However, as the authors pointed out, it remains unanswered how a fractal-like organization of chromosomes can occur in the absence of comparable coiling structures at a lower level. Still, as outlined above, the establishment of additional domains that represent dynamic, regulatory units of transcription activity may simply be not required across the protozoan kingdom and had therefore not evolved.
Despite the absence of TADs or globules, we observed a strong segregation of each chromosome into the actively transcribed core and the constitutively repressed subtelomeres harboring the *VSG* repertoire with differential folding behavior and degree of compaction of these domains. This organization is reminiscent of an organization of the

genome into A and B compartments as observed in higher eukaryotes [156]. Typically, these compartments segregate e.g. active from inactive genes, epigenomic marks, replication timing and the occurrence of retrotransposable elements [64]. This holds true for *T. brucei* where genomic compartments correlate with transcriptional activity, chromatin accessibility and differential localization in three dimensions. The 'wave-like' pattern of accessibility and of interaction frequencies as measured by ATACseq and by Hi-C across chromosomal core and subtelomeric regions clearly suggested a differential local chromatin structure and spatial positioning of these two compartments. Importantly, Lieberman-Aiden and colleagues found interaction frequencies within one compartment constantly higher than within the other [156]. This applies to *T. brucei* where an analysis of the distance-dependent decay of interaction frequencies of cores and subtelomeres showed higher interaction frequencies at any given genomic distance for the subtelomeric domains.

The segregation of transcribed housekeeping and silenced *VSG* genes in two and three dimensions may not only serve the faithful repression of all but one antigen but may also restrict high levels of recombination to antigen-coding regions as a mechanism to expand the *VSG* repertoire. This spatial restriction of high mutation and recombination rates to the subtelomeres would increase evolvability of the surface molecule repertoire and allow for facilitated adaptation whilst ensuring genome stability in the housekeeping portion of the genome, as it is e.g. the case for other bacterial and protozoal pathogens [87] [62]. Previous studies have pointed out that the presence of repressive heterochromatin is crucial for and in fact favors recombination events in repetitive subtelomeric regions [20]. From this perspective, the particular condensed and self-interacting state of the subtelomeric domains would have a central role in driving antigenic diversity and innovation. Analogously, FISH microscopy [87] and Hi-C in *P. falciparum* [13] [39] revealed a distinct clustering of chromosome-internal and subtelomeric *var* genes in the nuclear periphery. Ay and colleagues termed these 10-50 kb large clusters 'domain-like structures' (DLSs) as analogues to TADs in higher eukaryotes [13]. Whereas *var* gene clusters in *Plasmodium* are enriched in repressive histone marks H3K9me3, H3K36me3 and the heterochromatin-associated protein PfHP1 [159] [58] [262], the factors determining the heterochromatic nature of subtelomeres in *T. brucei* remain unclear. We did not, for example, find cohesin, H3.V or H4.V particularly enriched across subtelomeric sites, except for in the vicinity of telomeric repeats and across expression sites.

The large compacted domains around *VSG* gene arrays described in this study suggest that genome architecture is determinative for *VSG* silencing and antigenic exclusion. Further, it is apparent that comparable structures have evolved analogously e.g. in apicomplexans that

are exposed to comparable environments and selective pressures as *T. brucei*.

12.2 Inter-chromosomal interactions in *T. brucei*

12.2.1 *Centromeres and expression sites are hallmarks of inter-chromosomal interactions in T. brucei*

Similar to what has been described for yeast [188] and apicomplexans [13] [39], centromeres showed strong clustering in *T. brucei*. This raises the question, how this effects and constraints the positioning of megabase chromosomes in 3D space.
Due to the linear proximity of rDNA and centromeric repeats in *T. brucei*, centromeres are expected to be in spatial association with the nucleolus [192]. Such an organization is remarkably reminiscent of the presence of centromeres in heterochromatic nucleolus-associated domains (NADs) in humans (reviewed in [205]).
In *T. brucei*, the centromere spatially separates the two chromosome arms, thus insulating them from each other. These 'arm domains' [67] have been found in other eukaryote species by chromosome painting [67] and Hi-C [188] [154].
Despite concomitant interactions of telomere-proximal expression sites in *T. brucei*, several lines of evidence speak against a Rabl-conformation of chromosomes *T. brucei* as has been described in yeast [74], *Toxoplasma* and *Plasmodium* [39]. First, the chromosomal interaction heat maps do not exhibit a cross-like pattern as has been observed in yeast [74] and *Plasmodium* [13], reflecting subtelomeres of the same chromosome interacting. This is in agreement with the observation that the *VSG* arrays in *T. brucei* are always restricted to only one end of a respective chromosome, making it unlikely that the chromosome folds back to itself. Second, the power-law scaling observed in *T. brucei* does not support a stretched-out conformation of chromosomes as it had been observed in yeast [74] and *Plasmodium* [39].

12.2.2 *Silent expression sites engage in three dimensions in T. brucei*

The elevated interaction frequencies between silent expression sites is novel and contradictory to previous studies investigating the localization of inactive expression sites using FISH against the 50-bp repeat sequences upstream of the *VSG*-ES promoter. Here, a tendency towards a peripheral location, but no clustering of 50-bp repeats had been observed. The discrepancy between the results obtained by FISH microscopy and Hi-C can be explained by considering that FISH probes the 50-bp repeats, whereas Hi-C probes the telomere-proximal *VSG*. From the expression site promoter to the telomere, a fully decondensed ~50 kb long expression site would span 16.5 μm (the *T.*

brucei nucleus has a diameter of ~2-2.5 μm [79]. This implies that the location of the 50 bp repeats is a very rough estimation for the localization of the *VSG*. A limitation (and strength) of Hi-C, in turn, is that contact frequencies represent the 'average nuclear organization' in a cell population. Therefore, interaction frequencies are rather to be interpreted as interaction probabilities. In this particular case, silent expression sites seem to engage into contacts more frequently than observed by chance, indicating that they reside in the same nuclear compartment. This is supported by the similarities of genome-wide 4C-like interaction profiles of different silent expression sites (see Figure 5.1).

12.3 Transcription is spatially segregated in *T. brucei*

12.3.1 *rRNA genes form intra-nucleolar clusters*

Analysis of the spatial arrangement and local chromatin accessibility of RNA Pol I, II and III transcribed genes in *T. brucei* revealed distinct folding properties for these three groups.
First, Hi-C recapitulated the strong clustering of *rRNA* genes in the bipartite nucleolus, organized in a fibrillar and granular component [207]. The ribosomal genes – except for the *5S rRNA* – and RNA Pol I have previously been found concentrated and clustered in the nucleolar periphery [199] [144]. The nucleolar integrity can be uncoupled from RNA Pol I transcription and is ultimately dependent on the interaction of the ribosomal components [224]. Whether the clustering of *rRNA* genes is mediated by cohesin that we found enriched at *rRNA* sites, remains speculative. In trypanosomes, the nucleolus is considered a highly stable nuclear compartment as the nucleolus does not disintegrate during mitosis [207]. The constitutive interaction of *rRNA* genes explains their high levels of inter-chromosomal interaction frequencies observed by Hi-C. We further found chromatin at *rRNA* loci to display increased accessibility as measured by ATACseq, which reflected their active transcription state.
Interestingly, Hi-C identified the *rRNA* genes to be segregated in at least two distinct clusters in the nucleolus and an additional extranucleolar positioning of the *5S rRNA* that is transcribed by RNA Pol III [263]. A similar clustering of *rRNA* had previously been described for *Plasmodium*, where *rRNA* genes preferentially interact dependent on their transcription in either sexual (S-type *rRNA*) or asexual (A-type *rRNA*) stages of the parasite [149]. Such a striking separation by rRNA-type inside the nucleolus was not apparent from our data, except that the cluster of *rRNA* located on chromosomes 10 and 11 contained only *5.8S rRNA*, whereas the other cluster of *rRNA* genes located on chromosomes 1, 3 and 7 encompassed both *5.8S* and *18S* (small subunit (SSU)) *rRNA*.

12.3.2 *Do strand-switch regions represent boundary elements of interaction and transcription?*

In *T. brucei*, positioning and transcription of RNA Pol II and III transcribed genes is interlinked in so far as that *tRNA* tandem arrays are mostly located in between two RNA Pol II polycistronic transcription units.
tRNA loci are supposedly regions of high histone turnover and are therefore nucleosome depleted (reviewed in [138]), which is in agreement with the high accessibility that we found at *tRNA* genes by ATACseq.
This particular accessibility was accompanied with a distinct enrichment of cohesin across *tRNA* genes, framed by H3.V and H4.V histone variants. H3.V has previously been described to function in RNA Pol II termination and H3.V deletion caused a transcriptional read-through at RNA Pol II transcription termination sites [238] [226]. This, in turn, not only suggests a role of histone variants H3.V and H4.V in siRNA production inhibition [226], but also in physically insulating tRNA sites from transcription by RNA Pol II.
tRNA genes and their flanking sequences themselves have been described to function from yeast to human as boundary elements and as barriers for heterochromatin spreading [138], presumably *via* the recruitment of the RNA Pol III associated transcription factor TFIIIC [247]. Both in vertebrates and in lower eukaryotes, RNA Pol III transcribed sites are also characterized by cohesin binding (reviewed in [138]). Whereas in higher eukaryotes, cohesin is preferentially positioned at CTCF binding sites [211] [277], in yeast cohesin is typically actively 'pushed' by active transcription towards sites of convergence [150].
A function of *tRNA* genes as transcriptional barriers had previously been suggested for *T. brucei* [170] [245]. Taking into account the high and distinct enrichment of cohesin at transcription termination sites, but especially at *tRNA* genes observed in this study, cohesin may be stalled at sites, where RNA Pol II elongation is synergistically blocked by H3.V, H4.V and base-J. At tRNA sites cohesin positioning may inhibit spreading of heterochromatic marks or histone variants to provide the necessary accessibility for active *tRNA* transcription by RNA Pol III. The positioning of cohesin make it attractive to speculate on a clustering of these loci in three dimensions. Notably, a co-localization of *tRNA* and an association of these clusters with the nucleolus, the *5S rRNA* and centromeres has been described for *S. cerevisiae* [259] [74] and *S. pombe* [201]. Also in human cells, colocalization of tRNA genes into RNA Pol III transcription factories in association with nucleoli has been described [217] [205]. We found slightly elevated interaction frequencies among *T. brucei tRNA* genes in *trans*, however from our data there were no indications for a co-localization of *tRNA* genes

with *rRNA* loci or centromeric repeats. Given a co-localization of *tRNA* genes across chromosomes, it would be informative to also investigate a probable clustering of *tRNA* genes in *cis*, which was not part of this study. A scenario of intra-chromosomal clustering of transcription start and termination sites would also not only bring together *tRNA* genes, but would also be favorable in terms of efficient RNA Pol II 'reloading' to a transcribed PTU, similarly to the concept of RNA Pol II transcription factories in higher eukaryotes [225]. However, despite a small degree of TSS co-localization in *trans*, looping of PTUs was not apparent from virtual inspection of the interaction maps generated in this study. A thorough investigation of intra-chromosomal folding would require taking into account the distance-dependent decay of interaction frequencies, which was beyond the scope of this study. Still, the spatial segregation observed for RNA Pol I, II and III transcribed genes and the organization of co-transcribed genes in PTUs flanked by boundary-like elements could represent a primordial form of a more regulated co-localization and -expression of genes as observed in higher eukaryotes [225].

12.4 Is *VSG* exclusion mediated by a *trans*-acting element?

In the past, nuclear architecture has been given importance in the process of mutually exclusive expression in trypanosomes as the sole active expression site had been found to reside in a dedicated Pol I transcribed nuclear body, the ESB, whereas the silent expression sites reside in proximity to the nuclear envelope [199]. Upon differentiation to procyclic form parasites, the actively transcribed expression site is transcriptionally shut down and relocated towards the nuclear periphery [145], implying a positional silencing effect.
Whereas the mutual exclusive expression of olfactory receptors in mammalian nuclei relies on the association of multiple enhancer elements to the stochastically chosen olfactory receptor gene [36] [189] [190], enhancer elements have remained elusive in protozoa. In *Plasmodium*, where antigenic variation has been extensively studied, no regulatory DNA elements involved in mutually exclusive expression of a single *var* gene has so far been identified. Here, all but one *var* genes are located within peripheral facultative heterochromatin [86] from which a single *var* gene escapes. Also Hi-C in *Plasmodium* did not reveal any *trans*-acting enhancer element but local reorganization of chromatin at the activated locus [149]. The high degree of accessibility observed at the actively transcribed site is in agreement with the high rate of *VSG* transcription and the nucleosome depletion described previously [82]. Further, the active expression site interacted markedly with the SL repeat in *trans* and, strikingly, a transcriptional switch was accompanied with a respective change in BES-SL interaction. These

findings indicate that the promoter-driven, RNA Pol II transcribed SL locus is part of or in close association with the ESB. However, the nature of this interaction is unclear. The spliced-leader could indeed function as an enhancer in a sense that it recruits transcription factors shared by both polymerases thus forming a transcriptionally highly active hub or compartment. As proposed for the olfactory receptor choice [190], the spliced leader repeats and their transcripts may promote a euchromatic phase [115] confining the actively transcribed site, thus shielding it from heterochromatin spreading and stabilizing *VSG* transcription.

Alternatively (or additionally), the spatial proximity of the spliced-leader locus to the ESB may immediately supply mini-exons at the site of *VSG* RNA production, thus promoting RNA splicing and maturation. Indeed, previous publications have suggested a congregation of actively transcribed genes at splicing-factor enriched speckles [186] [34]. Although *trans*-splicing occurs co-transcriptionally, a previous study has stressed, that spliced-leader RNA maturation in kinetoplastids involves a cytoplasmic cap 4 methylation step [289]. Cytoplasmic trafficking and nuclear re-import of the spliced-leader transcripts would, of course, contradict an immediate processing of the *VSG* transcripts. Still, the spatial co-transcription of spliced-leader and *VSG* RNAs in the same transcriptionally active compartment could ensure, that comparable levels are produced.

Irrespective of whether the SL favors transcription or translation at the active expression site, this association may be required to ensure sufficient *VSG* mRNA production and may provide a possible explanation why the transcribed expression site is not associated to the nucleolus despite the use of RNA Pol I [212].

In procyclic form parasites, the ES-SL association was lost, which is in agreement with a disintegration of the ESB [144]. Instead, the Hi-C data recapitulated the previously described association of *procyclin* loci as a cluster with the nucleolus [144]. For the first time, our data showed a developmental repositioning of the *procyclin* cluster away from the nucleolus in bloodstream form parasites. A previous publication by DuBois and colleagues suggested an association of *procyclins* with the nuclear envelope in bloodstream form cells since the depletion of the lamin-like NUP-1 led to a derepression of *procyclins* [73]. The genome-wide 4C interaction profile of the *GPEET* promoter on Chr6 was, however, widely unchanged in our Hi-C data in procyclic cells and continued to exhibit a wave-like interaction pattern with preferential association with the chromosomal core regions (in contrast to silenced expression sites, for example). This positioning neither in association with the nucleolus nor with the repressed subtelomeres may indicate a 'poised' state of *procyclin* loci in bloodstream form parasites. This is in agreement with previous findings of Vanhamme and colleagues that find transcription of *procyclin* loci maintained at

relatively high levels in the bloodstream form stage [268].
In summary, these findings suggest, that the nuclear architecture of *T. brucei* is intimately and functionally linked with the parasite life-cycle stages. Specifically, the life-cycle dependent expression of *VSGs* and *procyclins* is reflected by developmental repositioning of these loci between a transcriptionally active RNA Pol I compartment and a transcriptionally silent, heterochromatic compartment.

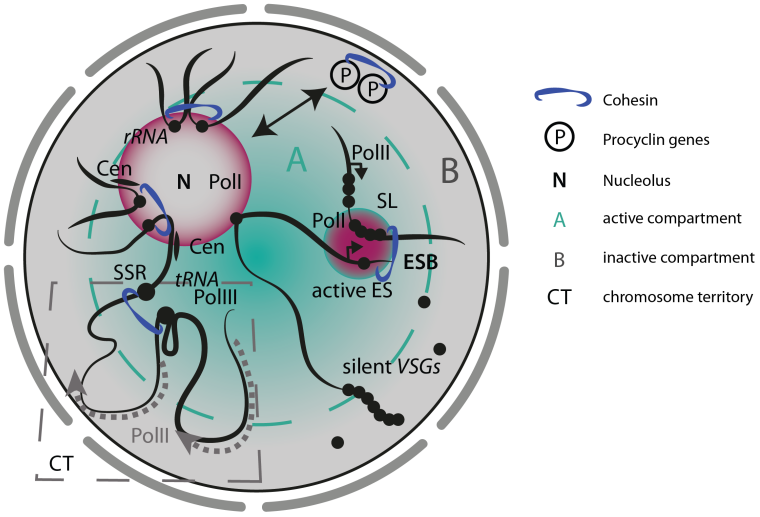
12.5 A regulatory function of *T. brucei* cohesin in nuclear architecture and antigenic variation?

In metazoans, cohesin is a major player in chromosome organization by mediating loop extrusion at CTCF binding sites [232]. The network of chromatin loops established by cohesin guides promoter-enhancer contacts by facilitation and insulation of interactions [72]. Even in the absence of CTCF, cohesin organizes the chromatin fiber, e.g. by mediating formation of globules and chromosome territories in fission yeast [188]. To exert its regulatory function in chromatin organization beyond sister-chromatin cohesion in S- and G2-phase, cohesin has been found present in G1 both in yeast [188] and higher eukaryotes [196]. For *T. brucei*, a previously published transcriptome data set indicated the presence of Scc1 transcripts across all cell cycle stages [10]. This is in agreement with our Scc1 immunofluorescence data indicating that cohesin is present throughout the cell cycle of *T. brucei*. In contrast, previous publications by the Gull and Carrington laboratories suggested onset of Scc1 expression only at late G1-phase in bloodstream form [95] and procyclic *T. brucei*. Whereas these studies used an anti-Scc1 serum for IF, in this study Scc1 was tagged and visualized *via* an anti-Ty1-antibody. This technical difference may account for differences in epitope accessibility and assay sensitivity.
A role of cohesin in antigenic variation is implied by its presence at the spliced-leader repeats, *procyclin* loci and across bloodstream form expression sites. A previous publication found a delayed mitotic division of the replicated active expression site, which was abolished when cohesin was depleted [145]. Our data indicated the presence of cohesin at both active and silent expression sites. Thus, cohesin enrichment is not an exclusive characteristic of the active expression site and cannot be responsible for the late division phenotype alone. Landeira and colleagues also report increased *VSG* switching frequencies upon cohesin depletion [145]. In the light of our data this might indicate that cohesin loss affected not only the active expression site, but also disrupted spatial insulation of silent expression sites in a repressive compartment, thus facilitating switching among them.

12.6 A hypothetical 3D model of genome organization in *T. brucei*

The integration of Hi-C data gained from this study with previously published principles of genome folding in *T. brucei* draws a conclusive picture of territorially organized megabase chromosomes whose core regions are part of an active (A) and condensed subtelomeres part of an inactive (B) nuclear compartment. Analogously to the apicomplexan parasite *P. falciparum* where domain-like structure (DLS) around virulence genes were described, we found *VSG* genes at subtelomeres organized into densely compacted domains with reduced interactions with the transcribed housekeeping portion of the genome, giving raise to the assumption of two main compartments in the *T. brucei* nucleus. Little regulation at the level of RNA Pol II transcription was reflected by the absence of strong chromosome-internal constraints (e.g. TADs). However, the analogies in chromatin structure found between higher eukaryote contact domains and PTU borders in *T. brucei* give reason to investigate their intra-chromosomal relative positioning further.
In contrast, megabase chromosomes interacted in *trans* at 'hotspots' of active transcription in membrane-less nuclear bodies. This includes the spliced-leader locus with the active transcribed ES at the ESB in bloodstream form parasites, the colocalization of rRNA loci (and centromeres) in the perinucleolar space and the association of procyclin loci with the nucleolus in procyclic cells. Notably, all genes colocalizing at these sites are transcriptionally regulated, promoter-driven and developmentally repositioned. Their active transcriptional state was represented by high accessibility in ATACseq.
A putative model of the *T. brucei* nucleus in three dimensions is depicted in Figure 12.1.
Although Hi-C data represents only an averaged interaction profile, genome organization in *T. brucei* demonstrated a strong connection between pathogenicity and genome architecture, as had previously been shown for apicomplexan parasites [39]. Thereby, transcriptional repression of the *VSG* family repertoire whilst selecting a single gene for surface protein expression appears primarily determinative for genome organization in *T. brucei*.

Figure 12.1: A putative model of the *T. brucei* nucleus



P, procyclin genes that are developmentally repositioned (arrows), N, nucleolus; A/green, active compartment; B/grey, inactive compartment; red, highly transcribed nuclear bodies; SL; spliced-leader; ESB, expression site body; Cen, centromere. *tRNA*, *rRNA*, *SL* and *VSG* genes are marked as black dots. Chromosomes are depicted as black lines. Cohesin rings (blue) stabilize interactions between distinct genomic sites.

13

Histone variants link genome organization and antigenic variation in *T. brucei*

13.1 How do histone variants H3.V and H4.V modulate chromatin at PTU boundaries and expression sites?

We found that the pattern of cohesin deposition across the *T. brucei* genome coincided with an enrichment of histone variants H3.V and H4.V. Scc1-ChIP in *H3.V* and *H4.V* deletion backgrounds however indicated, that cohesin loading to the DNA was not dependent on the presence of either H3.V or H4.V alone. It seems more likely, that cohesin rings are loaded to the DNA in an independent mechanism and are subsequently pushed towards PTU borders by active RNA Pol II transcription. At transcription termination sites, histone variants seem to function as speed-bumps for RNA Pol II transcription [245] [226] [238], which, in conclusion, positions cohesin at PTU borders just adjacent to H3.V and H4.V. To what extent cohesin stalling was affected in the *H3.V/H4.V* double deletion mutant could not be assessed in this study as resistances required for further genetic manipulation of the double deletion mutant were exhausted.
Previous publications have reported a transcriptional read-through effect at transcription termination sites and increased production of antisense transcripts in the absence of H3.V [238] [226]. This is in agreement with our findings that while transcription of the 'bulk' protein-coding genes, that reside PTU-internally, was not affected in the deletion mutant cell lines, transcription of genes located at PTU borders were, possibly indicating promiscuous transcription of these sites by RNA Pol II. A recent study investigating transcriptional changes in *H3.V* deletion mutants described a transcriptional repression of RNA Pol I transcribed procyclins and procyclin-associated genes (PAGs) upon *H3.V* deletion [226]. The authors proposed a downregulation of these transcripts as a result of increased dsRNA formation due to increased RNA Pol II transcription of an upstream opposing PTU. In contrast, investigation of significantly regulated genes (P_{adj}>0.1) in the *H3.V* and *H4.V* single and double deletion mutants indicated an upregulation of procyclins and PAGs. Fittingly, ATACseq indicated increased accessibility at sites of RNA Pol I transcription (*rRNA* genes, expression sites and *procyclin* genes) and at transcription termination sites. However, whether opening of chromatin at these sites is a direct effect of histone variant deletion and therefore a prerequisite for transcription, or rather a consequence of increased transcription after

histone variant deletion itself, remained unclear.
Notably, *tRNA* genes behaved differently than RNA Pol I transcribed loci in the deletion mutant cell lines. *tRNA* transcript levels were only significantly increased altered in the late double deletion mutant, but not in the single deletion strains or the early double deletion strain. Here, ATACseq indicated a *reduced* accessibility at these sites upon *H3.V* and/or *H4.V* deletion. This may indicate that the sharp insulation of *tRNA* genes for RNA Pol III transcription loci was mediated by H3.V and H4.V and disturbed upon histone variant deletion.
Despite reduced levels of H3.V and H4.V downstream of the active expression site promoter, all expression sites were characterized by increased levels of both histone variants and cohesin towards the telomeric repeats. As most expression sites are transcriptionally repressed, cohesin can – in contrast to chromosome internal sites – not be positioned by active transcription at these sites. Probably, additional pathways function in the deposition of subtelomeric cohesin. Previous studies have reported that histone variant H3.3 is deposited to heterochromatic loci by the ATRX-Daxx histone chaperone complex [151] [97]. Also, ATRX has been found to colocalize with subtelomeric cohesin and inactivation of ATRX led to decreasing amount of cohesin located at subtelomeric regions [77]. While our data did not reveal and interdependence between H3.V and cohesin deposition, it would be interesting to investigate the role of ATRX in subtelomere integrity and antigenic variation.
A function of subtelomeric H3.V in *VSG* and *mVSG* silencing has been described previously [238] [226]. Our data confirmed these findings as deletion of *H3.V* led to a significant upregulation of *VSG*, *mVSG* and *ESAG* transcripts from both, expression sites and subtelomeres. However, we and others found that the sole deletion of *H3.V* alone was not sufficient to induce a full switch of *VSG* expression [162] [238] [226], but modulated local chromatin accessibility at the promoter regions at silenced expression sites as determined by ATACseq. This is consistent with the finding by Reynolds and colleagues that from all *ESAG* genes, H3.V deletion mainly derepressed the ES promoter-proximal *ESAG-8* gene [226].
H3.V has therefore a major role in the maintenance of a silencing chromatin environment at transcriptionally repressed expression sites and subtelomeres.
H4.V deletion revealed a comparable phenotype as observed for H3.V deletion mutants, indicating that H3.V and H4.V, possibly together with base-J [226] [238] have synergistic barrier functions between sites of RNA Pol I, II and III transcription as well as in the attenuation of basal transcription at silent expression sites. Notably, it is only after the deletion of both histone variants, that chromatin at ESs displayed increased accessibility across the whole length of the expression site. The fact that accessibility was distinctly increased at the promoter

region in single deletion mutants is in agreement with the assumption that expression site silencing is achieved through attenuation of transcription elongation rather that inhibition of transcription initiation [269]. Still, I assume that the local increase in accessibility mediated by *H3.V* and *H4.V* deletion is alone not sufficient to cause transcriptional switching *VSG* expression, since this would require a ~10,000-fold excess in transcription levels [117].

13.2 Histone variant H3.V maintains genome integrity and controls recombination

Beyond a function in local chromatin structure, we found deletion of *H3.V*, but not *H4.V*, to affect nuclear architecture.
Whereas inter-chromosomal interactions e.g. between *rRNA* genes and centromeres persisted in the ΔH3.V and ΔH3.VΔH4.V deletion mutants, positioning of telomeres and expression sites was disturbed in three-dimensions.
The importance of histone H3 variant deposition for telomere integrity has previously been demonstrated in mouse embryonic stem cells [282]. Here, the depletion of H3.3 led to increased DNA-damage at telomeres and telomere fusions. Additionally, a study by Batté and colleagues in yeast revealed that the combined formation of telomeric hyperclusters (induced by sir3A2Q overexpression) and the loss of heterochromatic silencing (induced by sir3 deletion) are required to promote telomere clustering and recombination [20].
Although we were not able to unveil the mechanistic mode of action by which H3.V shapes nuclear architecture, it became clear that H3.V has a similar bivalent function, encompassing the maintenance spatial constraints and the regulation of local chromatin accessibility. Release of these constraints, in turn, favored *VSG* switching by recombination, which is the most frequent event driving antigenic variation in trypanosomes [269].
In addition, previous studies had found that recombinational events in T. brucei are predominantly facilitated between actively transcribed regions [5] and had speculated about a role of H3.V in controlling recombination of *VSGs* indirectly by regulating transcription at subtelomeric regions [226].
A phenotype comparable to the *H3.V* deletion in *T. brucei* has previously been observed in *P. falciparum* after knockout of the 'sirtuin' histone deacetylase *PfSir2a*. Loss of *PfSir2a* leads to a derepression of *var* genes and an increase in chromosomal rearrangements [183]. Although the SIR2rp1 homologue in *T. brucei* is not associated with *VSG* repression [7], the studies in *P. falciparum* highlight the importance of subtelomeric heterochromatin for chromosome integrity and maintenance of monoallelic exclusion.
Notably, deletion of *H3.V* was reminiscent of NUP-1 depletion in

T. brucei, which led to a derepression of silenced *VSGs*, increased switching and clustering of telomeres [73]. This may indicate that subtelomeric H3.V and NUP-1 work in the same pathway of expression site and subtelomere sequestration and requires further attention.
The stronger phenotype of *H3.V* deletion as compared to *H4.V* deletion in regulating chromatin accessibility and BES positioning is reflected by a more pronounced enrichment of H3.V at expression sites in our ChIPseq data. Still, it required the additional deletion of *H4.V* to sufficiently release local constraints, thereby promoting recombination and transcriptional activation.
The required combination of BES repositioning and concomitant BES chromatin opening points towards a control of mutually exclusive expression at multiple levels. H3.V seems to act at the interface of these two control mechanisms. In line with that, targeting of a single chromatin factor (e.g. removal or over-expression) in the past have only yielded *VSG* derepression or mildly increased switching rates in *T. brucei*, but never as extensively as observed in this study.

13.3 Is recombination the sole mode of *VSG* activation in ΔH3.VΔH4.V cells?

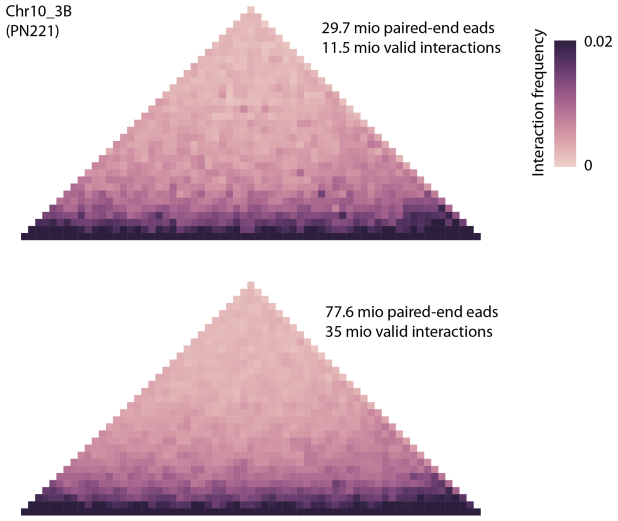
The frequency of chromosomal translocations correlates with the spatial proximity between the double-strand break (DSB) and the repair template [230] and the outcome of a DSB repair is therefore biased by chromosomal conformation. Non-random activation of *VSGs* in the course of *T. brucei* infections has been described previously [193]. Likewise, in our study, we observed a non-random activation of *VSGs* by single-cell RNAseq in ΔH3.VΔH4.V deletion mutants with a preferential activation of *VSG-11* and *VSG-8*. It would be interesting to investigate a putative relation between spatial positioning of expression sites and the likelihood of their respective activation. However, in the switching cell line presented here, the extensive repositioning of expression sites does not allow to infer any putative 'spatial predisposition' from the frequency of *VSG* activation.
Additionally, we observed transcripts of multiple *VSGs* in a single cell. Since the proportion of cells with transcripts of multiple different *VSG* had declined in the second scRNAseq experiment, these switchers most likely represented intermediate forms in the process of switching with the former *VSG* transcripts still around. However, if these cells indeed stably co-expressed two different *VSG* isoforms, it would be informative to investigate, whether co-expression of multiple *VSGs* was restricted to distinct combinations of VSGs due to sterical 'compatibilities' between different glycoproteins on a mosaic cell surface [216]. From a genomic point of view, co-activated *VSG* genes may need to reside in spatial proximity to share the same subcompartment and associated activating factors, e.g. such as VEX-1 [94].

We only investigated the recombination between BES1 and BES15 as the cause for activation of *VSG-11*. We cannot say whether recombination was the only source of *VSG* switching and can therefore not exclude that transcriptional *in situ* switches occurred in other cells.
In contrast to previous publications stating that the 70-bp repeats represent the main site of recombination among expression sites [93] [118], we found the point of recombination located inside the *ESAG-8* genes. That gene conversion among expression sites can take place in the absence of the 70-bp repeats had been shown before [178]. It would be informative to extract from the SMRTseq data if recombination also led to the activation of the other predominant *VSG* isoforms expressed and to investigate their respective point of recombination.
We found H3.V and H4.V to maintain genome integrity and chromosomal stability in *T. brucei*, thereby controlling mutually exclusive expression of *VSGs* at two levels: local chromatin compaction and global spatial sequestration of expression sites. These findings link the utilization of unusual histone variants, genome architecture and gene expression in *T. brucei*.

Part V

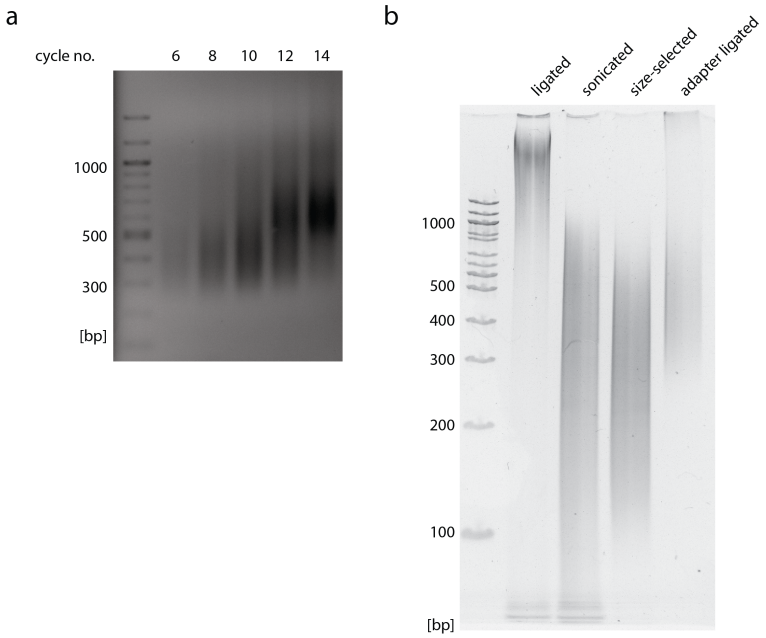
Appendix

Figure a.1: Hi-C sequencing depth and matrix resolution.



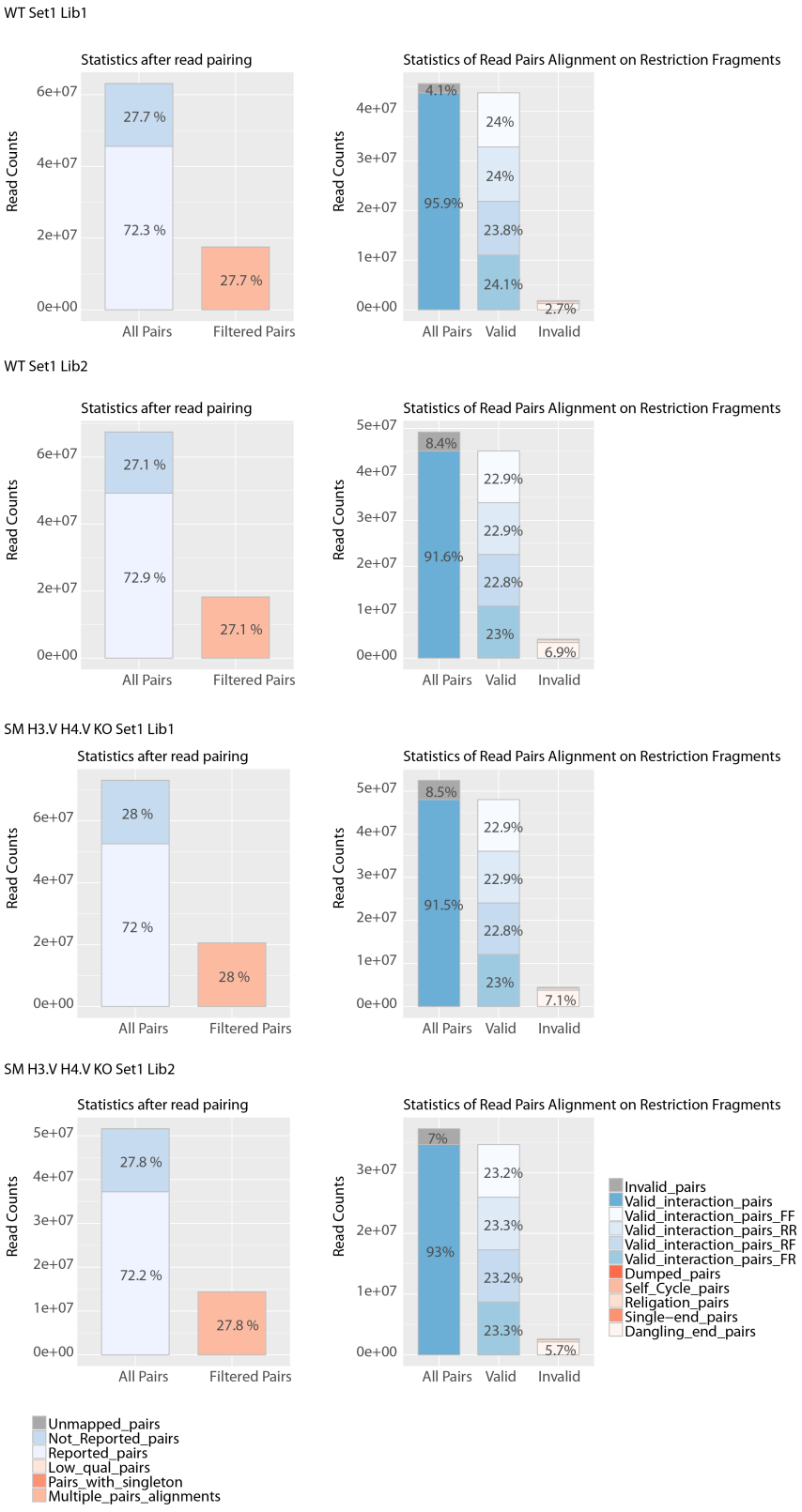
Shown is the subtelomeric arm Chr10_3B (PN221 cell line, 20 kb bin size) with 11.5 million (top) and 35 million reads (bottom) unique reads mapped to the whole genome (after removal of unitigs).

Figure a.2: Quality control of Hi-C libraries.



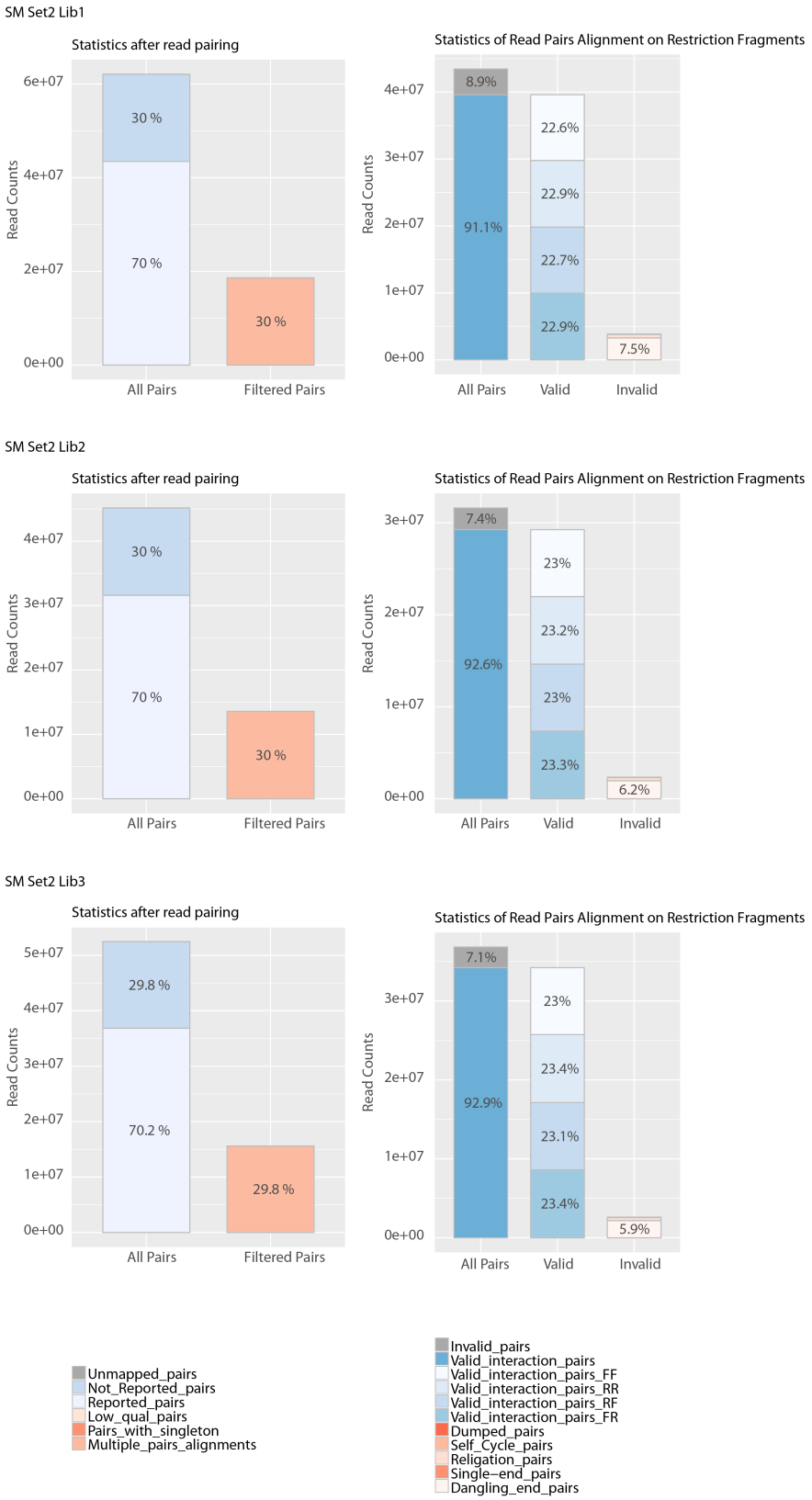
a Titration of amplification cycles of a Hi-C library. **b** Polyacrylamide gel showing 50 ng of digested and ligated, sonicated, size-selected and adapter-ligated Hi-C library.

Figure a.3: Alignment statistics of Hi-C in wild-type and ΔH3.VΔH4.V cells (Set1).



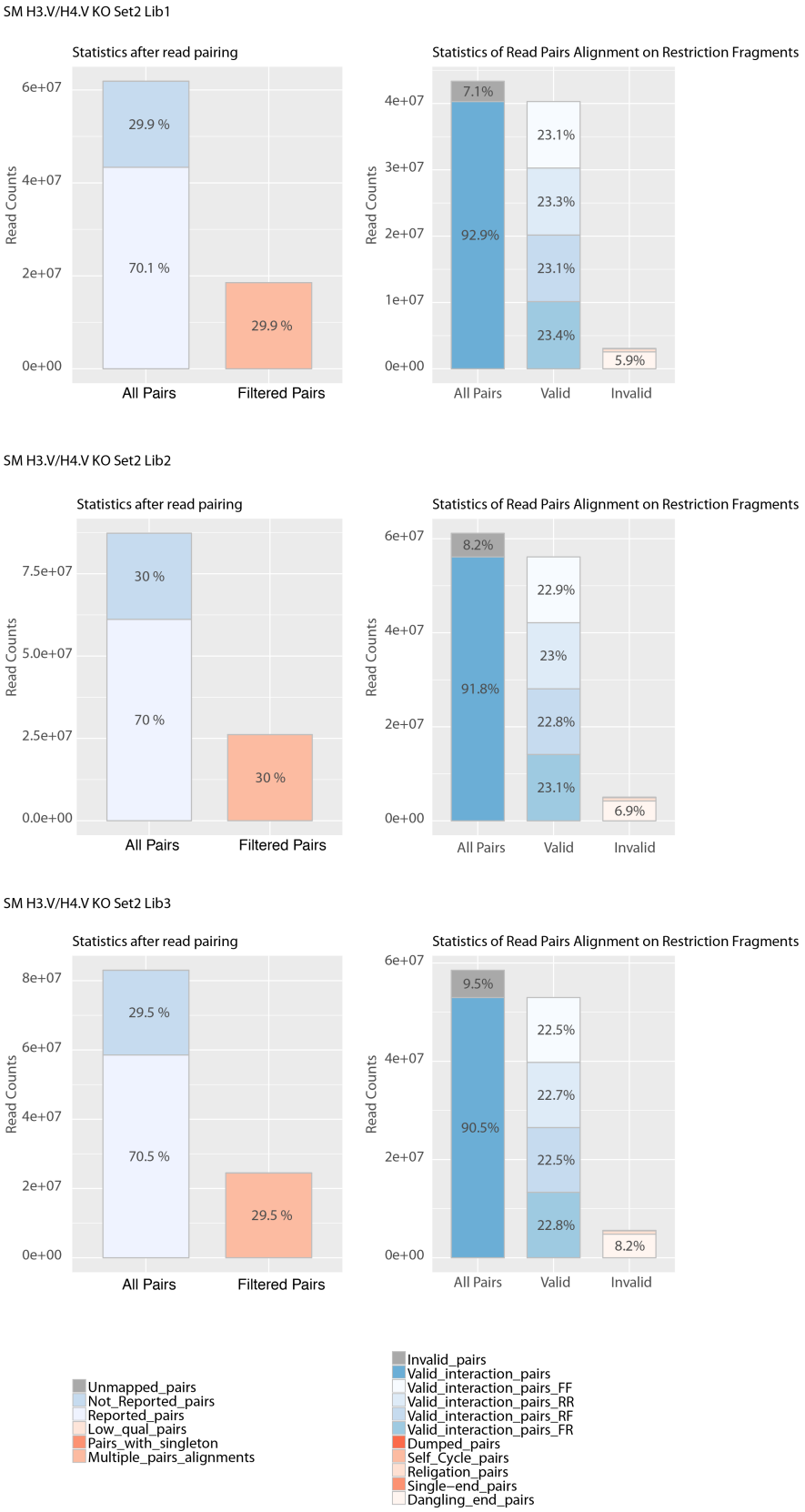
Hi-C Pro mapping statistics after read mapping and pairing of two replicates (Lib1, Lib2).

Figure a.4: Alignment statistics of Hi-C in single-marker cells (Set2).



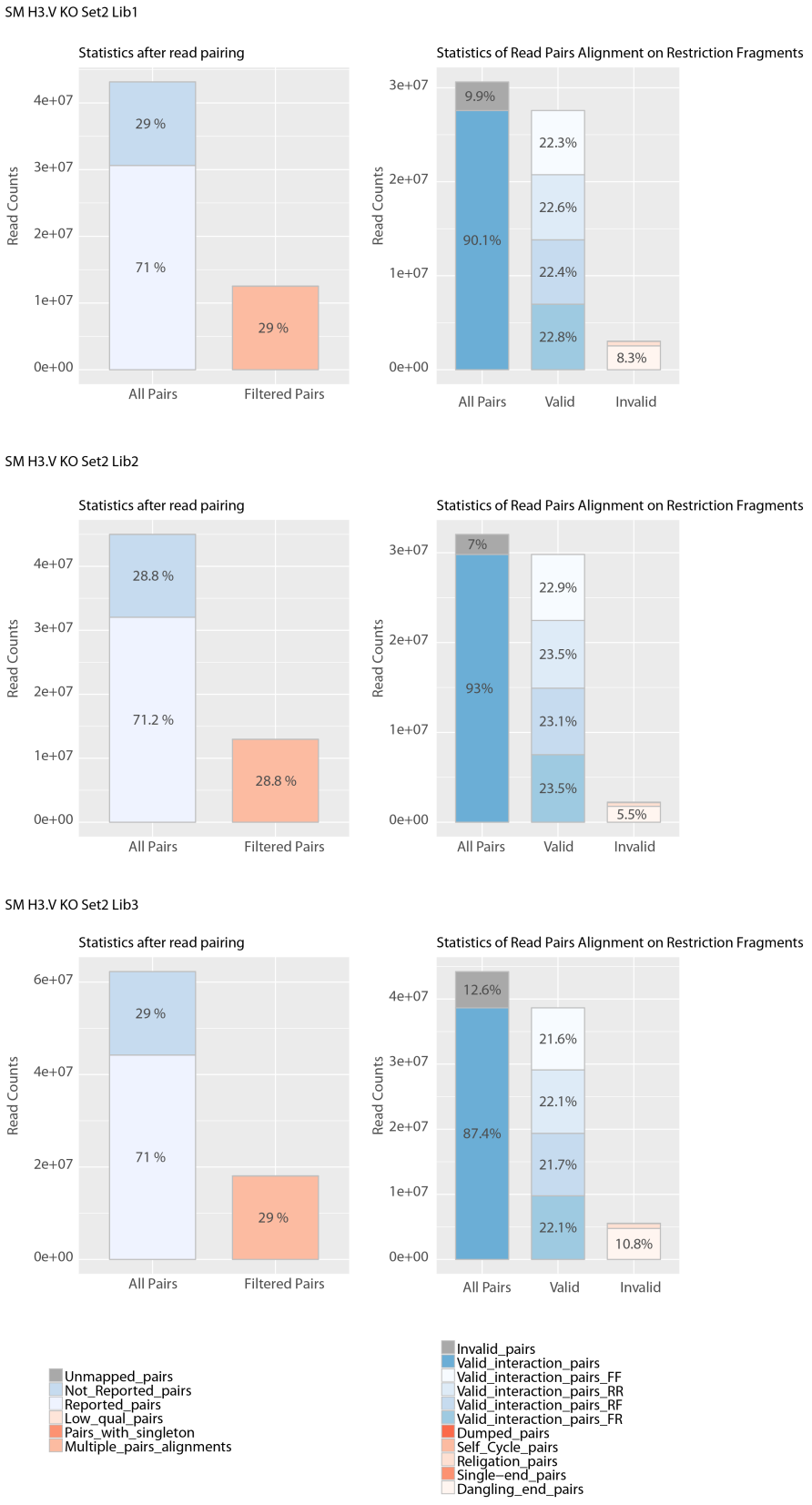
Hi-C Pro mapping statistics after read mapping and pairing of three replicates (Lib1, Lib2, Lib3).

Figure a.5: Alignment statistics of Hi-C in ΔH3.VΔH4.V cells (Set2).



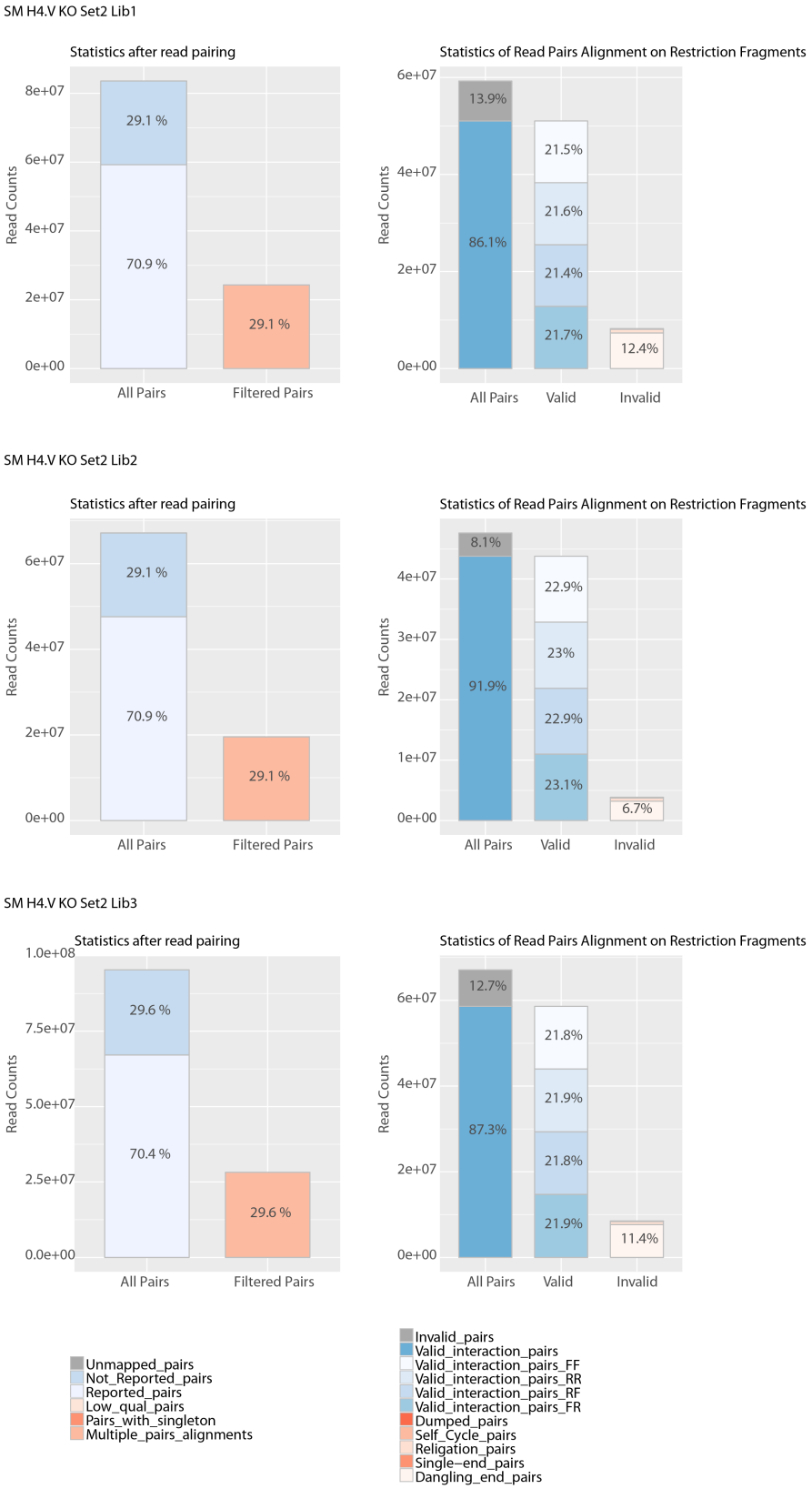
Hi-C Pro mapping statistics after read mapping and pairing of three replicates (Lib1, Lib2, Lib3).

Figure a.6: Alignment statistics of Hi-C in ΔH3.V cells (Set2).



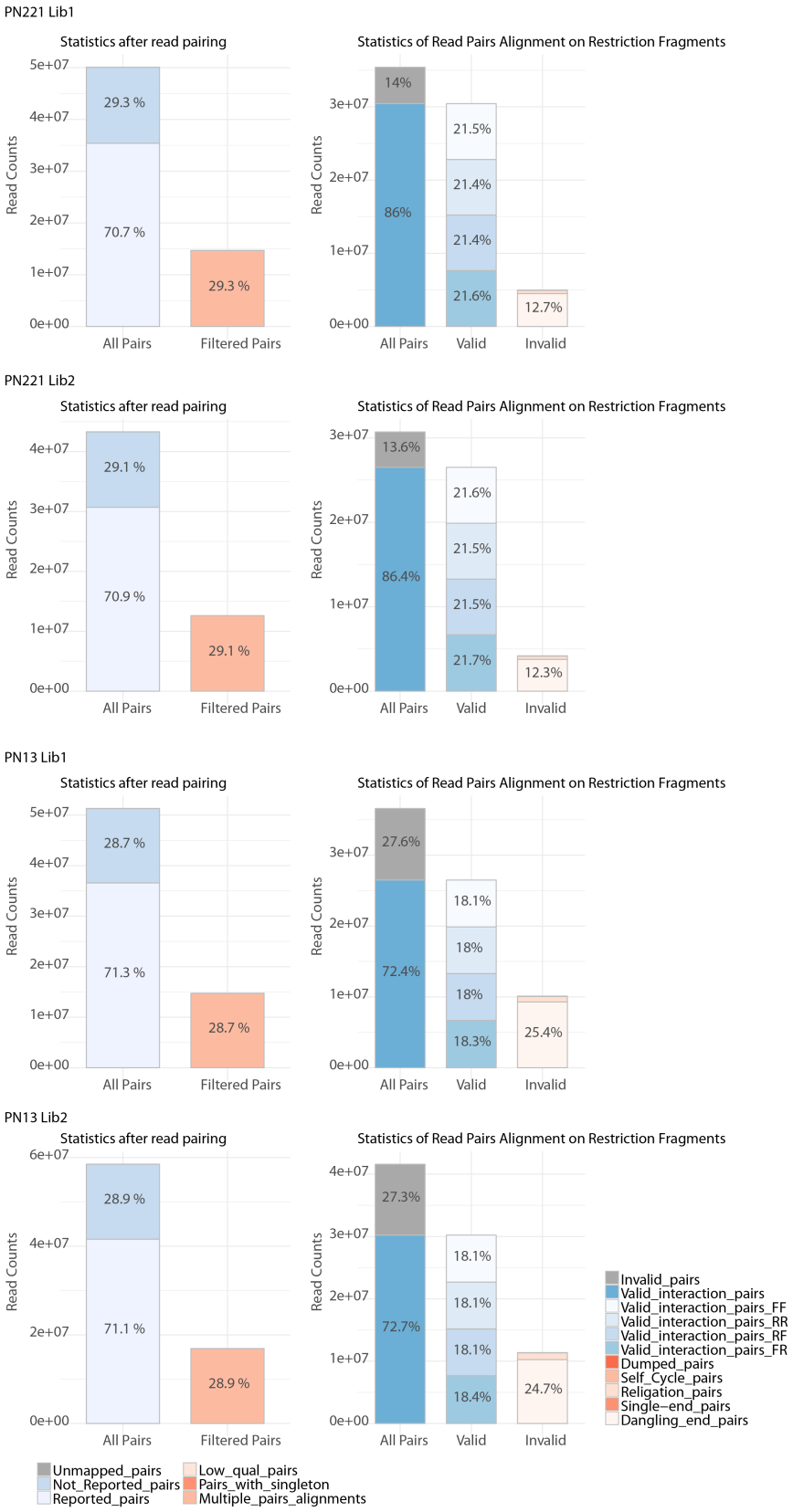
Hi-C Pro mapping statistics after read mapping and pairing of three replicates (Lib1, Lib2, Lib3).

Figure a.7: Alignment statistics of Hi-C in ΔH4.V cells (Set2).



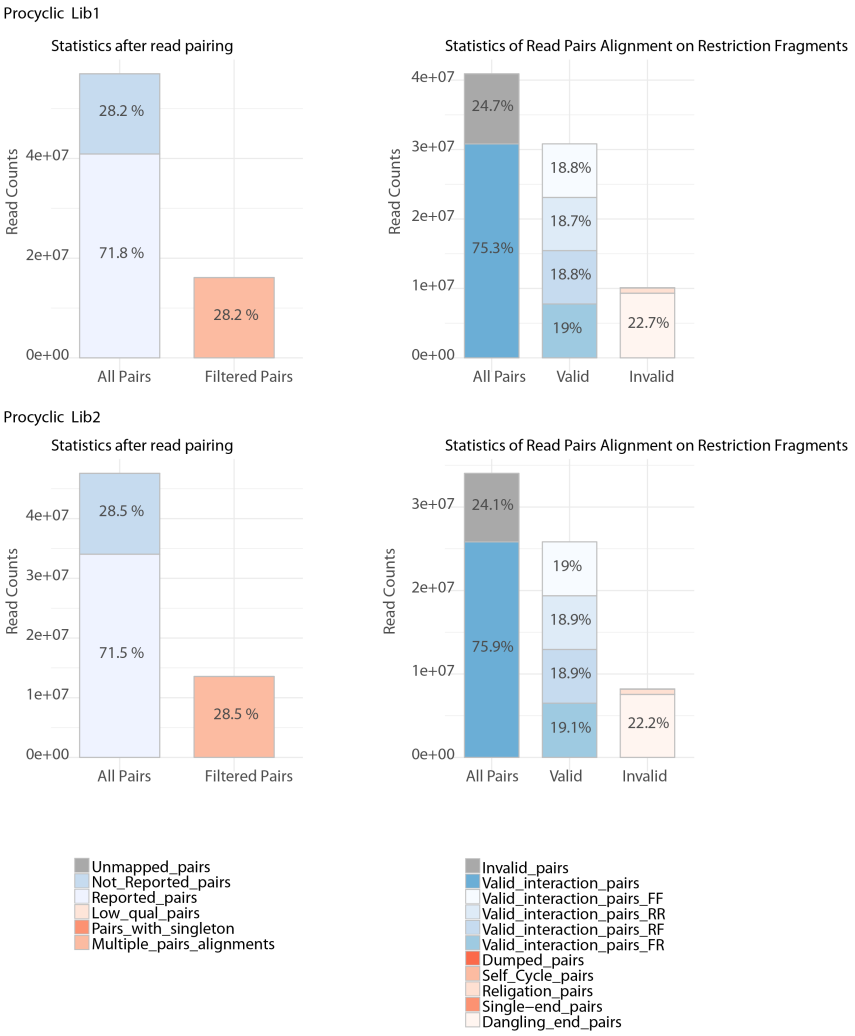
Hi-C Pro mapping statistics after read mapping and pairing of three replicates (Lib1, Lib2, Lib3).

Figure a.8: Alignment statistics of Hi-C in PN221 and PN13 cells.



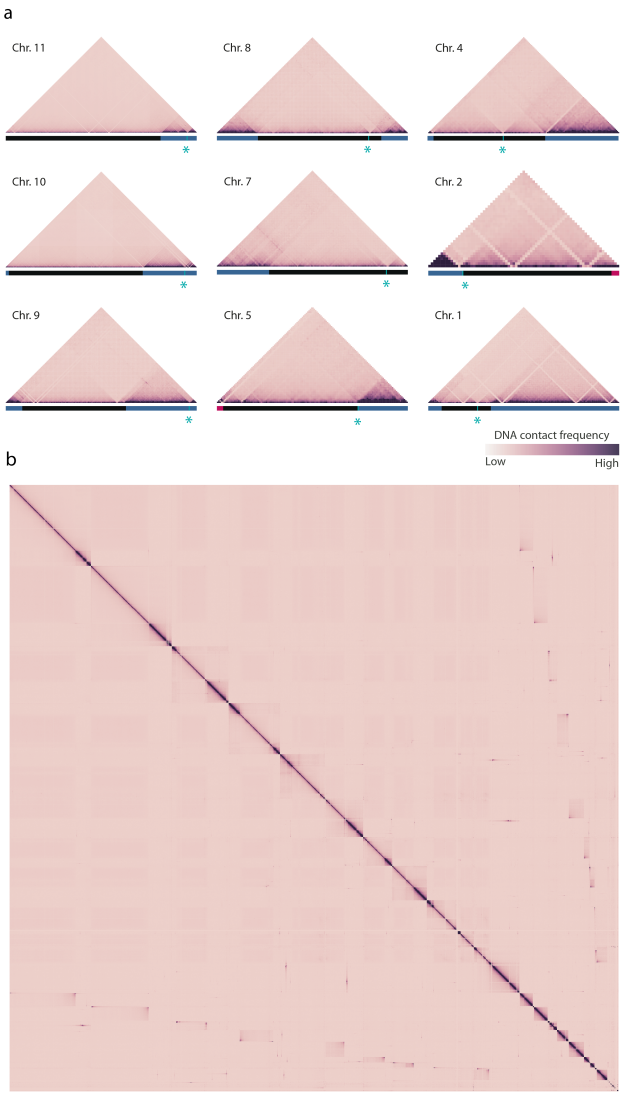
Hi-C Pro mapping statistics after read mapping and pairing of two replicates each (Lib1, Lib2).

Figure a.9: Alignment statistics of Hi-C in procyclic cells.



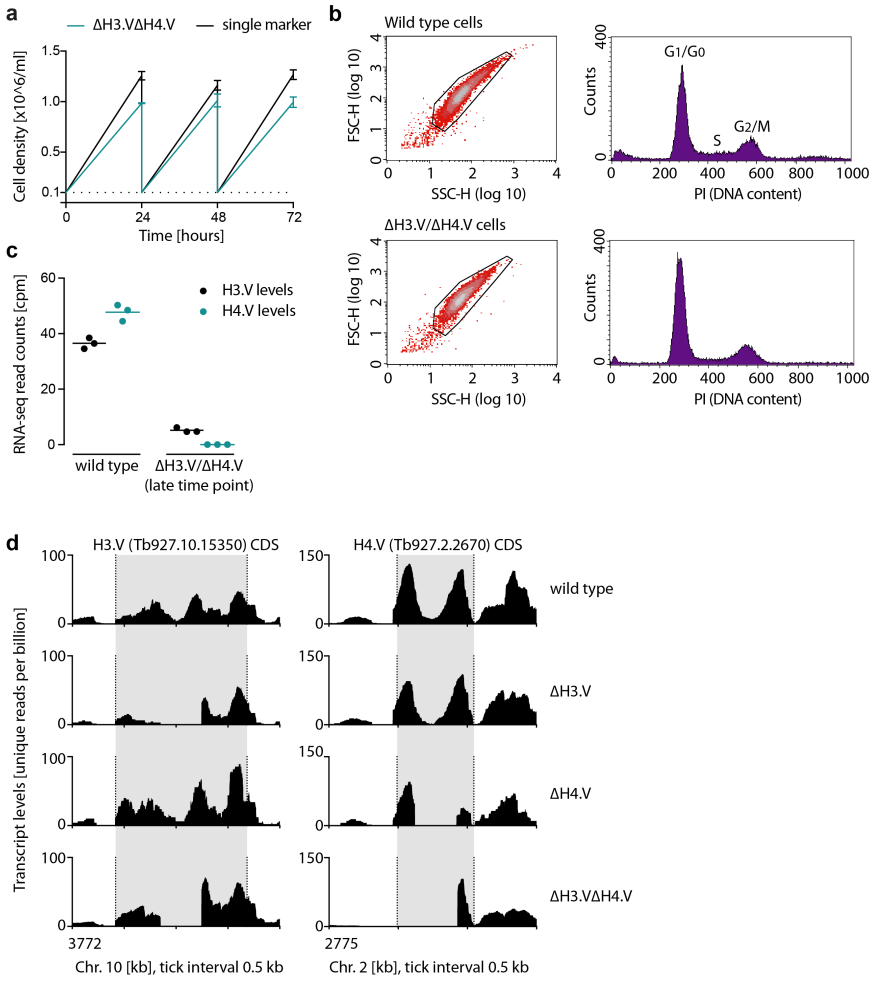
Hi-C Pro mapping statistics after read mapping and pairing of two replicates each (Lib1, Lib2).

Figure a.10: Interaction heat maps of the bloodstream-form genome.



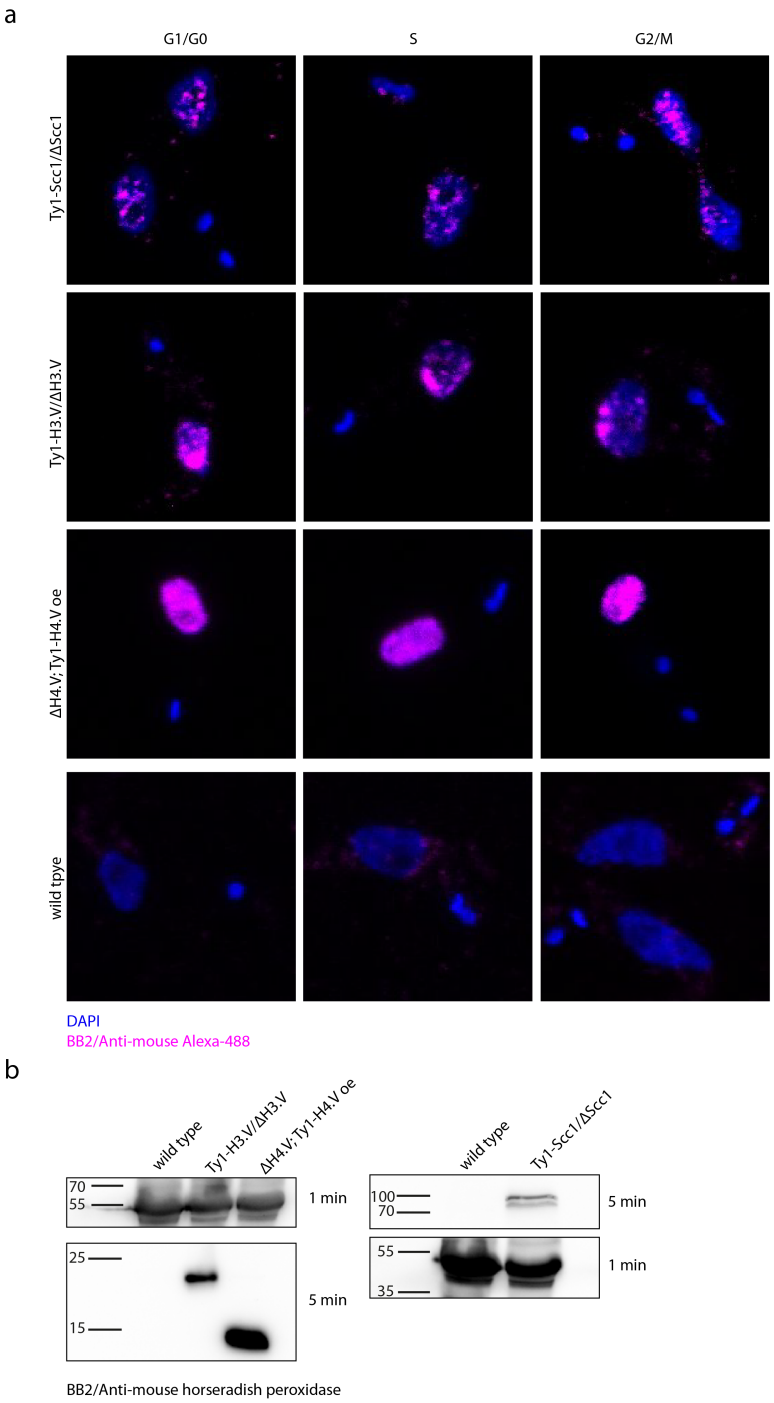
a Heat maps of intra-chromosomal interactions of individual chromosomes (20 kb resolution, data merged from three replicates). Core (black), subtelomeric (blue) and expression sites (red) are marked by horizontal lines. Centromeric regions are marked by a blue asterisk. **b** heat map of the whole haploid *T. brucei* genome at 20 kb resolution. One of the two alternative subtelomeres is joined to the respective core region.

Figure a.11: Characterization of *H3.V* and *H4.V* deletion cell lines.



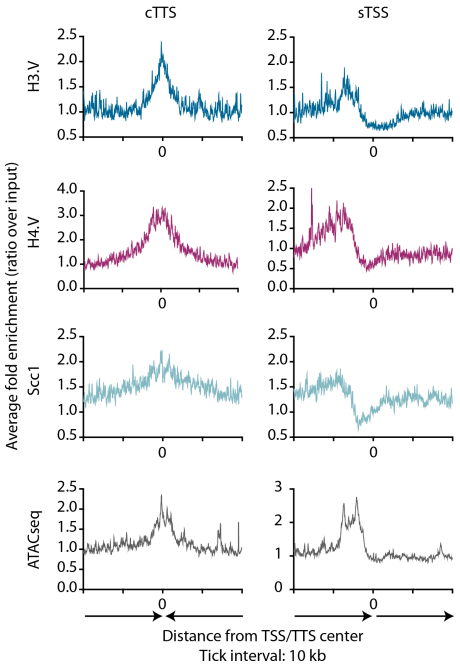
a Growth curve of the parental single marker cell line (black) and ΔH3.VΔH4.V cells (green). Shown is the mean with standard deviation from three replicates per cell line. **b** Analysis of cell cycles of wild-type (top) and ΔH3.VΔH4.V cells (bottom) by flow cytometry. **c** fragment counts mapping to *H3.V* and *H4.V* CDS in wild-type and ΔH3.VΔH4.V cells as determined by RNAseq. d. Visualization of fragment reads mapping to the *H3.V* and *H4.V* CDSs (grey) in wild-type, ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells (window size 1 bp, step size 1 bp).

Figure a.12: Characterization of cell lines with Ty1-tagged proteins.



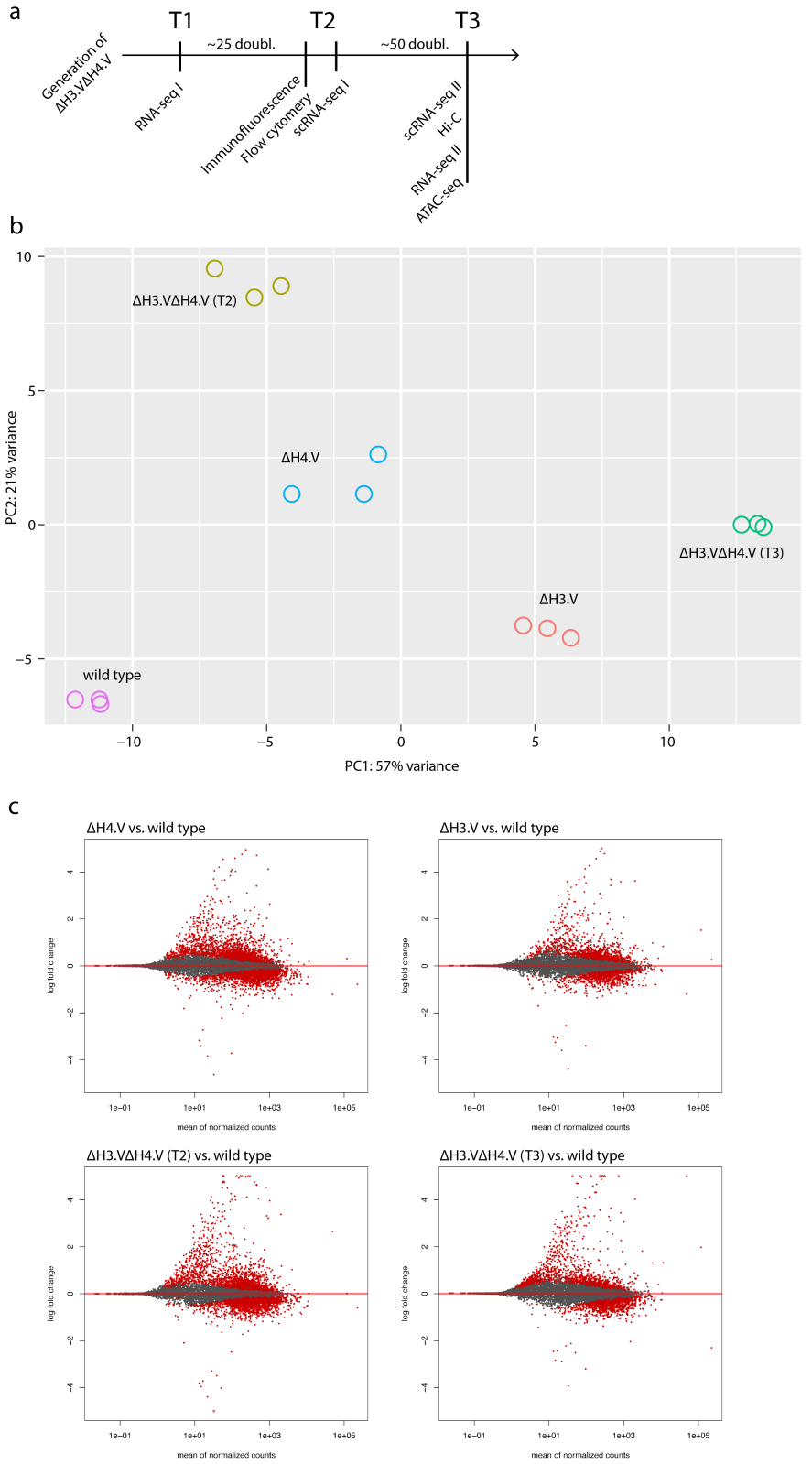
a Immunofluorescence imaging of Ty1-tagged H3.V, H4.V and Scc1 expression across the cell cycle using the BB2 antibody (magenta). DNA (blue) is stained with DAPI. Wild-type cells were treated in the same manner and used as negative controls. **b** Western blot agains Ty1-tagged versions of H3.V, H4.V and Scc1 in (5 million *T. brucei* cells). Imaging of α-tubulin (Tat1-antibody) was performed as sample loading control.

Figure a.13: Average H3.V, H4.V, Scc1 enrichment and accessibility profile at TTSs



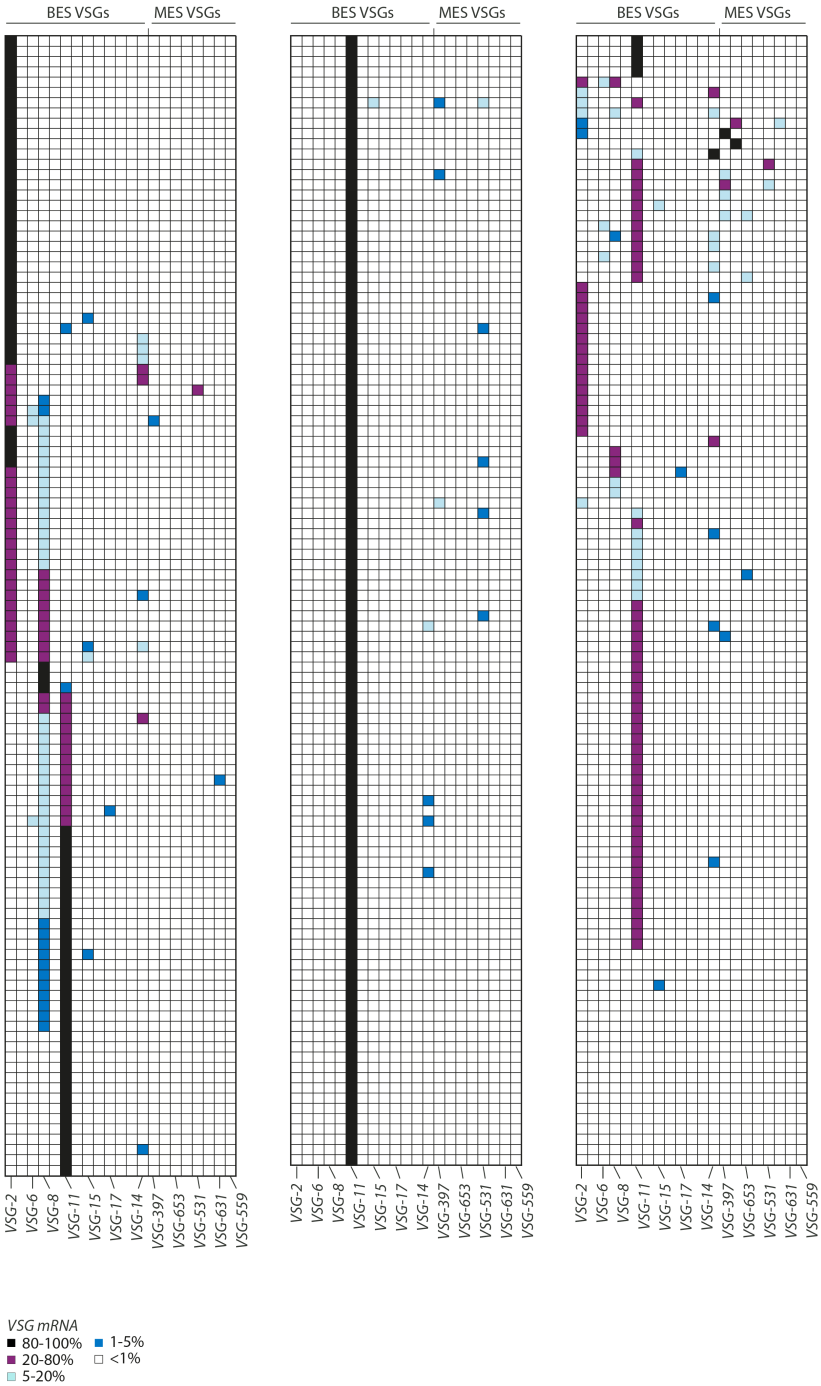
Fold-enrichment of H3.V (blue), H4.V (red) and Scc1 (green) ChIP signal and accessibility as determined by ATACseq (fold-change over gDNA sample) averaged across all annotated convergent transcription termination sites (cTTS) and unidirectional (single) transcription start sites (sTSS). TTS and TTS center are marked by 'o'; arrows indicate direction of transcription.

Figure a.14: RNAseq experiments in ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells.



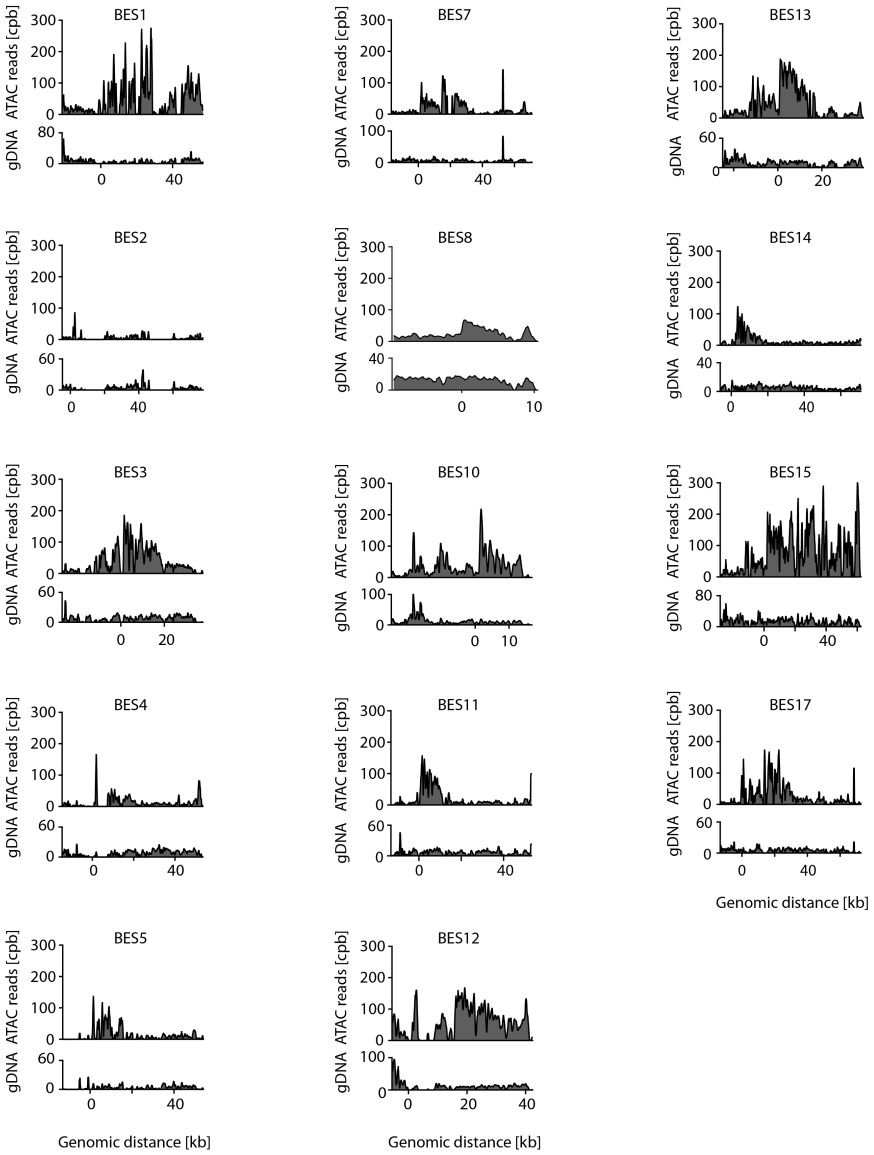
a Experimental timeline. **b** PCA analysis of RNAseq experiments. **c** MA-plots of RNAseq in ΔH3.V, ΔH4.V and ΔH3.VΔH4.V cells showing the log2 fold-change on in transcript levels in comparison to wild type. Significantly regulated genes are shown in red.

Figure a.15: Single-cell RNAseq in ΔH3.VΔH4.V cells to determine *VSG* expression.



a Each row corresponds to a single cell (n=331), columns correspond to selected *VSG* isoforms. Given is the percentage of a distinct *VSG* mRNA of all *VSG* mRNAs per single cell (considering uniquely mapped reads only). Sequencing reads were normalized for sequencing depth, library size, *VSG* gene length and uniqueness.

Figure a.16: ATACseq across all BESs in ΔH3.VΔH4.V cells.



cpb, counts per billion. The BES promoter region is demarked by 'o'. gDNA reads illustrate the degree of uniqueness ('mappability') across each expression site.

1. Acemel, R., Maeso, I & Gómez-Skarmeta, J. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdiscip Rev Dev Biol* **6** (2017).
2. Agmon, N, Liefshitz, B, Zimmer, C, Fabre, E & Kupiec, M. Effect of nuclear architecture on the efficiency of double-strand break repair. *Nat Cell Biol* **15,** 694–699 (2013).
3. Ahmad, K & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9,** 1191–1200 (2002).
4. Akiyoshi, B & Gull, K. Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156,** 1247–1258 (2014).
5. Alsford, S & Horn, D. RNA polymerase I transcription stimulates homologous recombination in Trypanosoma brucei. *Mol Biochem Parasitol* **153,** 77–79 (2007).
6. Alsford, S & Horn, D. Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. *Nucleic Acids Res* **40,** 10150–10160 (2012).
7. Alsford, S, Kawahara, T, Isamah, C & Horn, D. A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. *Mol Microbiol* **63,** 724–736 (2007).
8. Amaral, N, Ryu, T, Li, X & Chiolo, I. Nuclear Dynamics of Heterochromatin Repair. *Trends Genet* **33,** 86–100 (2017).
9. Amiguet-Vercher, A, Pérez-Morga, D, Pays, A, Poelvoorde, P, Van Xong, H, Tebabi, P, Vanhamme, L & Pays, E. Loss of the mono-allelic control of the VSG expression sites during the development of Trypanosoma brucei in the bloodstream. *Mol Microbiol* **51,** 1577–1588 (2004).
10. Archer, S., Inchaustegui, D, Queiroz, R & Clayton, C. The cell cycle regulated transcriptome of Trypanosoma brucei. *PLoS One* **6,** e18425 (2011).
11. Armelin-Correa, L., Gutiyama, L., Brandt, D. & Malnic, B. Nuclear compartmentalization of odorant receptor genes. *Proc Natl Acad Sci U S A* **111,** 2782–2787 (2014).
12. Ay, F & Noble, W. Analysis methods for studying the 3D architecture of the genome. *Genome Biol* **16,** 183 (2015).

13. Ay, F, Bunnik, E., Varoquaux, N, Bol, S., Prudhomme, J, Vert, J., Noble, W. & Le Roch, K. Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res* **24,** 974–988 (2014).

14. Baltimore, D. Our genome unveiled. *Nature* **409,** 814–816 (2001).

15. Banaszynski, L. e. a. Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. *Cell* **155,** 107–120 (2013).

16. Barry, J., Ginger, M., Burton, P & McCulloch, R. Why are parasite contingency genes often associated with telomeres. *Int J Parasitol* **33,** 29–45 (2003).

17. Barry, J., Hall, J. & Plenderleith, L. Genome hyperevolution and the success of a parasite. *Ann N Y Acad Sci* **1267,** 11–17 (2012).

18. Bastin, P, Bagherzadeh, Z, Matthews, K. & Gull, K. A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei. *Mol Biochem Parasitol* **77,** 235–239 (1996).

19. Batram, C, Jones, N., Janzen, C., Markert, S. & Engstler, M. Expression site attenuation mechanistically links antigenic variation and development in Trypanosoma brucei. *Elife* **3,** e02324 (2014).

20. Batté, A, Brocas, C, Bordelet, H, Hocher, A, Ruault, M, Adjiri, A, Taddei, A & Dubrana, K. Recombination at subtelomeres is regulated by physical distance, double-strand break resection and chromatin status. *EMBO J* **36,** 2609–2625 (2017).

21. Beitel, C., Froenicke, L, Lang, J., Korf, I., Michelmore, R., Eisen, J. & Darling, A. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* **2,** e415 (2014).

22. Belaghzal, H, Dekker, J & Gibcus, J. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123,** 56–65 (2017).

23. Belton, J., McCord, R., Gibcus, J., Naumova, N, Zhan, Y & Dekker, J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58,** 268–276 (2012).

24. Benjamini, Y, Drai, D, Elmer, G, Kafkafi, N & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125,** 279–284 (2001).

25. Berriman, M. e. a. The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Mol Biochem Parasitol* **122,** 131–140 (2002).

26. Bickmore, W. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* **14,** 67–84 (2013).

27. Bitter, W, Gerrits, H, Kieft, R & Borst, P. The role of transferrin-receptor variation in the host range of Trypanosoma brucei. *Nature* **391,** 499–502 (1998).

28. Black, B. & Cleveland, D. Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell* **144,** 471–479 (2011).

29. Borst, P & Cross, G. Molecular basis for trypanosome antigenic variation. *Cell* **29,** 291–303 (1982).

30. Borst, P & Sabatini, R. Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* **62,** 235–251 (2008).

31. Borst, P, van der Ploeg, M, van Hoek, J., Tas, J & James, J. On the DNA content and ploidy of trypanosomes. *Mol Biochem Parasitol* **6,** 13–23 (1982).

32. Boyle, S, Gilchrist, S, Bridger, J., Mahy, N., Ellis, J. & Bickmore, W. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* **10,** 211–219 (2001).

33. Branco, M. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4,** e138 (2006).

34. Brown, J. e. a. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol* **182,** 1083–1097 (2008).

35. Brun, R & Schönenberger. Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. *Acta Trop* **36,** 289–292 (1979).

36. Buck, L & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65,** 175–187 (1991).

37. Buenrostro, J., Giresi, P., Zaba, L., Chang, H. & Greenleaf, W. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10,** 1213–1218 (2013).

38. Bunnik, E. e. a. Changes in genome organization of parasite-specific gene families during the Plasmodium transmission stages. *Nat Commun* **9,** 1910 (2018).

39. Bunnik, E. e. a. Comparative 3D genome organization in apicomplexan parasites. *Proc Natl Acad Sci U S A* (2019).

40. Burton, J., Adey, A, Patwardhan, R., Qiu, R, Kitzman, J. & Shendure, J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* **31,** 1119–1125 (2013).

41. Bäumer, C, Fisch, E, Wedler, H, Reinecke, F & Korfhage, C. Exploring DNA quality of single cells for genome analysis with simultaneous whole-genome amplification. *Sci Rep* **8,** 7476 (2018).

42. Bönisch, C & Hake, S. Histone H2A variants in nucleosomes and chromatin: more or less stable. *Nucleic Acids Res* **40,** 10719–10741 (2012).

43. Cavalier-Smith, T. Origin of the cell nucleus. *Bioessays* **9,** 72–78 (1988).

44. Chambeyron, S & Bickmore, W. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* **18,** 1119–1130 (2004).

45. Chaves, I, Zomerdijk, J, Dirks-Mulder, A, Dirks, R., Raap, A. & Borst, P. Subnuclear localization of the active variant surface glycoprotein gene expression site in Trypanosoma brucei. *Proc Natl Acad Sci U S A* **95,** 12328–12333 (1998).

46. Chiolo, I, Minoda, A, Colmenares, S., Polyzos, A, Costes, S. & Karpen, G. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144,** 732–744 (2011).

47. Clayton, C. Life without transcriptional control? From fly to man and back again. *EMBO J* **21,** 1881–1888 (2002).

48. Cook, P. & Marenduzzo, D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res* **46,** 9895–9906 (2018).

49. Cramer, P. e. a. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37,** 337–352 (2008).

50. Crane, E, Bian, Q, McCord, R., Lajoie, B., Wheeler, B., Ralston, E., Uzawa, S, Dekker, J & Meyer, B. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523,** 240–244 (2015).

51. Cremer, T & Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol* **2,** a003889 (2010).

52. Cremer, T, Kreth, G, Koester, H, Fink, R., Heintzmann, R, Cremer, M, Solovei, I, Zink, D & Cremer, C. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr* **10,** 179–212 (2000).

53. Cremer, T, Cremer, M, Hübner, B, Strickfaden, H, Smeets, D, Popken, J, Sterr, M, Markaki, Y, Rippe, K & Cremer, C. The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett* **589,** 2931–2943 (2015).

54. Cremer, T, Cremer, M & Cremer, C. The 4D Nucleome: Genome Compartmentalization in an Evolutionary Context. *Biochemistry (Mosc)* **83,** 313–325 (2018).

55. Cross, G. Antigenic variation in trypanosomes: secrets surface slowly. *Bioessays* **18,** 283–291 (1996).

56. Cross, G., Kim, H. & Wickstead, B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Mol Biochem Parasitol* **195,** 59–73 (2014).

57. Dacks, J., Field, M., Buick, R, Eme, L, Gribaldo, S, Roger, A., Brochier-Armanet, C & Devos, D. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *J Cell Sci* **129,** 3695–3703 (2016).

58. Dahan-Pasternak, N. e. a. PfSec13 is an unusual chromatin-associated nucleoporin of Plasmodium falciparum that is essential for parasite proliferation in human erythrocytes. *J Cell Sci* **126,** 3055–3069 (2013).

59. Daniels, J., Gull, K & Wickstead, B. Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev* **74,** 552–569 (2010).

60. De Wit, E & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26,** 11–24 (2012).

61. Dean, S, Sunter, J, Wheeler, R., Hodkinson, I, Gluenz, E & Gull, K. A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol* **5,** 140197 (2015).

62. Deitsch, K., Moxon, E. & Wellems, T. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* **61,** 281–293 (1997).

63. Deitsch, K., Lukehart, S. & Stringer, J. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7,** 493–503 (2009).

64. Dekker, J & Heard, E. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett* **589,** 2877–2884 (2015).

65. Dekker, J, Rippe, K, Dekker, M & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).

66. Denninger, V, Fullbrook, A, Bessat, M, Ersfeld, K & Rudenko, G. The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in Trypanosoma brucei. *Mol Microbiol* **78,** 459–474 (2010).

67. Dietzel, S. e. a. Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. *Chromosome Res* **6,** 25–33 (1998).

68. Dion, V & Gasser, S. Chromatin movement in the maintenance of genome stability. *Cell* **152,** 1355–1364 (2013).

69. Dixon, J., Selvaraj, S, Yue, F, Kim, A, Li, Y, Shen, Y, Hu, M, Liu, J. & Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

70. Dobin, A, Davis, C., Schlesinger, F, Drenkow, J, Zaleski, C, Jha, S, Batut, P, Chaisson, M & Gingeras, T. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

71. Dostie, J. e. a. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* **16,** 1299–1309 (2006).

72. Doyle, B, Fudenberg, G, Imakaev, M & Mirny, L. Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput Biol* **10,** e1003867 (2014).

73. DuBois, K. e. a. NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. *PLoS Biol* **10,** e1001287 (2012).

74. Duan, Z, Andronescu, M, Schutz, K, McIlwain, S, Kim, Y., Lee, C, Shendure, J, Fields, S, Blau, C. & Noble, W. A three-dimensional model of the yeast genome. *Nature* **465,** 363–367 (2010).

75. Edgar, R, Domrachev, M & Lash, A. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30,** 207–210 (2002).

76. Eid, J. e. a. Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).

77. Eid, R, Demattei, M., Episkopou, H, Augé-Gouillou, C, Decottignies, A, Grandin, N & Charbonneau, M. Genetic Inactivation of ATRX Leads to a Decrease in the Amount of Telomeric Cohesin and Level of Telomere Transcription in Human Glioma Cells. *Mol Cell Biol* **35,** 2818–2830 (2015).

78. Elsässer, S., Noh, K., Diaz, N, Allis, C. & Banaszynski, L. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522,** 240–244 (2015).

79. Ersfeld, K. Nuclear architecture, genome and chromatin organisation in Trypanosoma brucei. *Res Microbiol* **162,** 626–636 (2011).

80. Ersfeld, K & Gull, K. Partitioning of large and minichromosomes in Trypanosoma brucei. *Science* **276,** 611–614 (1997).

81. Fakan, S & van Driel, R. The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding. *Semin Cell Dev Biol* **18,** 676–681 (2007).

82. Figueiredo, L. & Cross, G. Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. *Eukaryot Cell* **9,** 148–154 (2010).

83. Figueiredo, L., Janzen, C. & Cross, G. A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biol* **6,** e161 (2008).

84. Fraschka, S., Henderson, R. & Bártfai, R. H3.3 demarcates GC-rich coding and subtelomeric regions and serves as potential memory mark for virulence gene expression in Plasmodium falciparum. *Sci Rep* **6,** 31965 (2016).

85. Fraser, P & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447,** 413–417 (2007).

86. Freitas-Junior, L. e. a. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* **121,** 25–36 (2005).

87. Freitas-Junior, L., Bottius, E, Pirrit, L., Deitsch, K., Scheidig, C, Guinet, F, Nehrbass, U, Wellems, T. & Scherf, A. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature* **407,** 1018–1022 (2000).

88. Fudenberg, G, Belton, J., Goloborodko, A, Imakaev, M, Dekker, J & Mirny, L. Polymer models of yeast S. cerevisiae genome organization. *Epigentics & Chromatin* **6,** 128 (2013).

89. Fullwood, M. e. a. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).

90. Gandhi, R, Gillespie, P. & Hirano, T. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol* **16,** 2406–2417 (2006).

91. Gerbi, S. Bundling up DNA. *Elife* **7** (2018).

92. Glover, L, Hutchinson, S, Alsford, S, McCulloch, R, Field, M. & Horn, D. Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. *Cell Microbiol* **15,** 1984–1993 (2013).

93. Glover, L, Alsford, S & Horn, D. DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathog* **9,** e1003260 (2013).

94. Glover, L, Hutchinson, S, Alsford, S & Horn, D. VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proc Natl Acad Sci U S A* **113,** 7225–7230 (2016).

95. Gluenz, E, Sharma, R, Carrington, M & Gull, K. Functional characterization of cohesin subunit SCC1 in Trypanosoma brucei and dissection of mutant phenotypes in two life cycle stages. *Mol Microbiol* **69,** 666–680 (2008).

96. Gogarten, J., Kibak, H, Dittrich, P, Taiz, L, Bowman, E., Bowman, B., Manolson, M., Poole, R., Date, T & Oshima, T. Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci U S A* **86,** 6661–6665 (1989).

97. Goldberg, A. e. a. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140,** 678–691 (2010).

98. Grasser, F, Neusser, M, Fiegler, H, Thormeyer, T, Cremer, M, Carter, N., Cremer, T & Müller, S. Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J Cell Sci* **121,** 1876–1886 (2008).

99. Grosberg, A., Nechaev, S. & Shakhnovich, E. [The role of topological limitations in the kinetics of homopolymer collapse and self-assembly of biopolymers]. *Biofizika* **33,** 247–253 (1988).

100. Gruber, S, Arumugam, P, Katou, Y, Kuglitsch, D, Helmhart, W, Shirahige, K & Nasmyth, K. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* **127,** 523–537 (2006).

101. Guelen, L. e. a. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951 (2008).

102. Hadjur, S, Williams, L., Ryan, N., Cobb, B., Sexton, T, Fraser, P, Fisher, A. & Merkenschlager, M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460,** 410–413 (2009).

103. Haeusler, R. & Engelke, D. Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Res* **34,** 4826–4836 (2006).

104. Hake, S., Garcia, B., Duncan, E., Kauer, M, Dellaire, G, Shabanowitz, J, Bazett-Jones, D., Allis, C. & Hunt, D. Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* **281,** 559–568 (2006).

105. Hall, J., Wang, H & Barry, J. Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. *PLoS Pathog* **9,** e1003502 (2013).

106. Harmston, N, Ing-Simmons, E, Tan, G, Perry, M, Merkenschlager, M & Lenhard, B. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* **8,** 441 (2017).

107. Harr, J., Luperchio, T., Wong, X, Cohen, E, Wheelan, S. & Reddy, K. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J Cell Biol* **208,** 33–52 (2015).

108. Hayashi, H, Nomoto, M & Iwai, K. Tetrahymena histone H4. Complete amino acid sequences of two variants. *J Biochem* **96,** 1449–1456 (1984).

109. Heger, P, Marin, B, Bartkuhn, M, Schierenberg, E & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A* **109,** 17507–17512 (2012).

110. Heinz, S, Benner, C, Spann, N, Bertolino, E, Lin, Y., Laslo, P, Cheng, J., Murre, C, Singh, H & Glass, C. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38,** 576–589 (2010).

111. Henikoff, S & Smith, M. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* **7,** a019364 (2015).

112. Hertz-Fowler, C. e. a. Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One* **3,** e3527 (2008).

113. Hewitt, S., Farmer, D, Marszalek, K, Cadera, E, Liang, H., Xu, Y, Schlissel, M. & Skok, J. Association between the Igk and Igh immunoglobulin loci mediated by the 3' Igk enhancer induces 'decontraction' of the Igh locus in pre-B cells. *Nat Immunol* **9,** 396–404 (2008).

114. Hirumi, H & Hirumi, K. Axenic culture of African trypanosome bloodstream forms. *Parasitol Today* **10,** 80–84 (1994).

115. Hnisz, D, Shrinivas, K, Young, R., Chakraborty, A. & Sharp, P. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).

116. Horn, D. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol* **195,** 123–129 (2014).

117. Horn, D & McCulloch, R. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol* **13,** 700–705 (2010).

118. Hovel-Miner, G, Mugnier, M., Goldwater, B, Cross, G. & Papavasiliou, F. A Conserved DNA Repeat Promotes Selection of a Diverse Repertoire of Trypanosoma brucei Surface Antigens from the Genomic Archive. *PLoS Genet* **12,** e1005994 (2016).

119. Hovel-Miner, G., Boothroyd, C., Mugnier, M, Dreesen, O, Cross, G. & Papavasiliou, F. Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. *PLoS Pathog* **8,** e1002900 (2012).

120. Hozumi, N & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* **73,** 3628–3632 (1976).

121. Hsieh, T., Weiner, A, Lajoie, B, Dekker, J, Friedman, N & Rando, O. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162,** 108–119 (2015).

122. Hughes, K, Wand, M, Foulston, L, Young, R, Harley, K, Terry, S, Ersfeld, K & Rudenko, G. A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *EMBO J* **26,** 2400–2410 (2007).

123. Hübner, B, Lomiento, M, Mammoli, F, Illner, D, Markaki, Y, Ferrari, S, Cremer, M & Cremer, T. Remodeling of nuclear landscapes during human myelopoietic cell differentiation maintains co-aligned active and inactive nuclear compartments. *Epigenetics Chromatin* **8,** 47 (2015).

124. Ikegami, K, Egelhofer, T., Strome, S & Lieb, J. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol* **11,** R120 (2010).

125. Imakaev, M, Fudenberg, G, McCord, R., Naumova, N, Goloborodko, A, Lajoie, B., Dekker, J & Mirny, L. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9,** 999–1003 (2012).

126. Imakaev, M., Tchourine, K., Nechaev, S. & Mirny, L. Effects of topological constraints on globular polymers. *Soft Matter* **11,** 665–671 (2015).

127. Janssen, A, Breuer, G., Brinkman, E., van der Meulen, A., Borden, S., van Steensel, B, Bindra, R., LaRocque, J. & Karpen, G. A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* **30,** 1645–1657 (2016).

128. Jin, C, Zang, C, Wei, G, Cui, K, Peng, W, Zhao, K & Felsenfeld, G. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41,** 941–945 (2009).

129. Jin, D., Mata Martin, C, Sun, Z, Cagliero, C & Zhou, Y. Nucleolus-like compartmentalization of the transcription machinery in fast-growing bacterial cells. *Crit Rev Biochem Mol Biol* **52,** 96–106 (2017).

130. Johnston, R. & Desplan, C. Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343,** 661–665 (2014).

131. Kalashnikova, A., Porter-Goff, M., Muthurajan, U., Luger, K & Hansen, J. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface* **10,** 20121022 (2013).

132. Kaplan, N & Dekker, J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol* **31,** 1143–1147 (2013).

133. Kassem, A, Pays, E & Vanhamme, L. Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in Trypanosoma brucei. *Proc Natl Acad Sci U S A* **111,** 8943–8948 (2014).

134. Keel, B. & Snelling, W. Comparison of Burrows-Wheeler Transform-Based Mapping Algorithms Used in High-Throughput Whole-Genome Sequencing: Application to Illumina Data for Livestock Genomes. *Front Genet* **9,** 35 (2018).

135. Kind, J. e. a. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163,** 134–147 (2015).

136. Kind, J, Pagie, L, Ortabozkoyun, H, Boyle, S, de Vries, S., Janssen, H, Amendola, M, Nolen, L., Bickmore, W. & van Steensel, B. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153,** 178–192 (2013).

137. Kinney, N., Sharakhov, I. & Onufriev, A. Chromosome-nuclear envelope attachments affect interphase chromosome territories and entanglement. *Epigenetics Chromatin* **11,** 3 (2018).

138. Kirkland, J., Raab, J. & Kamakaka, R. TFIIIC bound DNA elements in nuclear organization and insulation. *Biochim Biophys Acta* **1829,** 418–424 (2013).

139. Koehler, D, Zakhartchenko, V, Froenicke, L, Stone, G, Stanyon, R, Wolf, E, Cremer, T & Brero, A. Changes of higher order chromatin arrangements during major genome activation in bovine preimplantation embryos. *Exp Cell Res* **315,** 2053–2063 (2009).

140. Kornberg, R. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–871 (1974).

141. Kosak, S., Scalzo, D, Alworth, S., Li, F, Palmer, S, Enver, T, Lee, J. & Groudine, M. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol* **5,** e309 (2007).

142. Kruithof, M, Chien, F., Routh, A, Logie, C, Rhodes, D & van Noort, J. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat Struct Mol Biol* **16,** 534–540 (2009).

143. Lajoie, B., Dekker, J & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72,** 65–75 (2015).

144. Landeira, D & Navarro, M. Nuclear repositioning of the VSG promoter during developmental silencing in Trypanosoma brucei. *J Cell Biol* **176,** 133–139 (2007).

145. Landeira, D, Bart, J., Van Tyne, D & Navarro, M. Cohesin regulates VSG monoallelic expression in trypanosomes. *J Cell Biol* **186,** 243–254 (2009).

146. Langmead, B & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).

147. Le Dily, F. e. a. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* **28,** 2151–2162 (2014).

148. Lee, C., Wang, R., Chang, H., Capurso, D, Segal, M. & Haber, J. Chromosome position determines the success of double-strand break repair. *Proc Natl Acad Sci U S A* **113,** E146–54 (2016).

149. Lemieux, J., Kyes, S., Otto, T., Feller, A., Eastman, R., Pinches, R., Berriman, M, Su, X. & Newbold, C. Genome-wide profiling of chromosome interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Mol Microbiol* **90,** 519–537 (2013).

150. Lengronne, A, Katou, Y, Mori, S, Yokobayashi, S, Kelly, G., Itoh, T, Watanabe, Y, Shirahige, K & Uhlmann, F. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430,** 573–578 (2004).

151. Lewis, P., Elsaesser, S., Noh, K., Stadler, S. & Allis, C. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* **107,** 14075–14080 (2010).

152. Li, H & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).

153. Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G, Abecasis, G, Durbin, R & 1000, G. P.D.P. S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

154. Li, Q, Tjong, H, Li, X, Gong, K, Zhou, X., Chiolo, I & Alber, F. The three-dimensional genome organization of Drosophila melanogaster through data integration. *Genome Biol* **18,** 145 (2017).

155. Liang, X., Haritan, A, Uliel, S & Michaeli, S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell* **2,** 830–840 (2003).

156. Lieberman-Aiden, E. e. a. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).

157. Litwin, I, Pilarczyk, E & Wysocki, R. The Emerging Role of Cohesin in the DNA Damage Response. *Genes (Basel)* **9** (2018).

158. Liu, A., Van der Ploeg, L., Rijsewijk, F. & Borst, P. The transposition unit of variant surface glycoprotein gene 118 of Trypanosoma brucei. Presence of repeated elements at its border and absence of promoter-associated sequences. *J Mol Biol* **167,** 57–75 (1983).

159. Lopez-Rubio, J., Mancio-Silva, L & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5,** 179–190 (2009).

160. Losada, A, Hirano, M & Hirano, T. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev* **12,** 1986–1997 (1998).

161. Love, M., Huber, W & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

162. Lowell, J. & Cross, G. A variant histone H3 is enriched at telomeres in Trypanosoma brucei. *J Cell Sci* **117,** 5937–5947 (2004).

163. Lowell, J., Kaiser, F, Janzen, C. & Cross, G. Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei. *J Cell Sci* **118,** 5721–5730 (2005).

164. Lua, R, Borovinsky, A. & Grosberg, A. Fractal and statistical properties of large compact polymers: a computational study. *Polymer* **45,** 717–731 (2004).

165. Luger, K, Mäder, A., Richmond, R., Sargent, D. & Richmond, T. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).

166. Lupiáñez, D. e. a. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161,** 1012–1025 (2015).

167. López-García, P & Moreira, D. Open Questions on the Origin of Eukaryotes. *Trends Ecol Evol* **30,** 697–708 (2015).

168. Ma, W. e. a. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* **12,** 71–78 (2015).

169. Malik, H. & Henikoff, S. Phylogenomics of the nucleosome. *Nat Struct Biol* **10,** 882–891 (2003).

170. Marchetti, M., Tschudi, C, Silva, E & Ullu, E. Physical and transcriptional analysis of the Trypanosoma brucei genome reveals a typical eukaryotic arrangement with close interspersionof RNA polymerase II- and III-transcribed genes. *Nucleic Acids Res* **26,** 3591–3598 (1998).

171. Markenscoff-Papadimitriou, E, Allen, W., Colquitt, B., Goh, T, Murphy, K., Monahan, K, Mosley, C., Ahituv, N & Lomvardas, S. Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* **159,** 543–557 (2014).

172. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

173. Martin, W. A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. *Proc. R. Soc. Lond. B* **266,** 1387–1395 (1999).

174. Martin, W & Koonin, E. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440,** 41–45 (2006).

175. Martin, W & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).

176. Martin, W., Garg, S & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci* **370,** 20140330 (2015).

177. Mattiroli, F, D'Arcy, S & Luger, K. The right place at the right time: chaperoning core histone variants. *EMBO Rep* **16,** 1454–1466 (2015).

178. McCulloch, R, Rudenko, G & Borst, P. Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol Cell Biol* **17,** 833–843 (1997).

179. McStay, B. Nucleolar organizer regions: genomic 'dark matter' requiring illumination. *Genes Dev* **30,** 1598–1610 (2016).

180. Meister, P & Taddei, A. Building silent compartments at the nuclear periphery: a recurrent theme. *Curr Opin Genet Dev* **23,** 96–103 (2013).

181. Melville, S., Leech, V, Navarro, M & Cross, G. The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. *Mol Biochem Parasitol* **111,** 261–273 (2000).

182. Mereschkowsky, K. Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. *Biologisches Centralblatt* **30,** 289–303 (1910).

183. Merrick, C., Jiang, R., Skillman, K., Samarakoon, U, Moore, R., Dzikowski, R, Ferdig, M. & Duraisingh, M. Functional analysis of sirtuin genes in multiple Plasmodium falciparum strains. *PLoS One* **10,** e0118865 (2015).

184. Millau, J. & Gaudreau, L. CTCF, cohesin, and histone variants: connecting the genome. *Biochem Cell Biol* **89,** 505–513 (2011).

185. Mirny, L. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* **19,** 37–51 (2011).

186. Misteli, T, Cáceres, J. & Spector, D. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387,** 523–527 (1997).

187. Mitchell, J. & Fraser, P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* **22,** 20–25 (2008).

188. Mizuguchi, T. e. a. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* **516,** 432–435 (2014).

189. Monahan, K, Schieren, I, Cheung, J, Mumbey-Wafula, A, Monuki, E. & Lomvardas, S. Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *Elife* **6** (2017).

190. Monahan, K, Horta, A & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* (2019).

191. Moreira & Lopez-Garcia. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* **47,** 517–530 (1998).

192. Morelle, C, Sterkers, Y, Crobu, L, MBang-Benet, D., Kuk, N, Portalès, P, Bastien, P, Pagès, M & Lachaud, L. The nucleoporin Mlp2 is involved in chromosomal distribution during mitosis in trypanosomatids. *Nucleic Acids Res* **43,** 4013–4027 (2015).

193. Mugnier, M., Cross, G. & Papavasiliou, F. The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science* **347,** 1470–1473 (2015).

194. Müller, L. e. a. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563,** 121–125 (2018).

195. Nagano, T, Lubling, Y, Stevens, T., Schoenfelder, S, Yaffe, E, Dean, W, Laue, E., Tanay, A & Fraser, P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59–64 (2013).

196. Nasmyth, K & Haering, C. Cohesin: its roles and mechanisms. *Annu Rev Genet* **43,** 525–558 (2009).

197. Naumova, N, Imakaev, M, Fudenberg, G, Zhan, Y, Lajoie, B., Mirny, L. & Dekker, J. Organization of the mitotic chromosome. *Science* **342,** 948–953 (2013).

198. Navarro, M & Cross, G. In situ analysis of a variant surface glycoprotein expression-site promoter region in Trypanosoma brucei. *Mol Biochem Parasitol* **94,** 53–66 (1998).

199. Navarro, M & Gull, K. A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature* **414,** 759–763 (2001).

200. Navarro, M, Peñate, X & Landeira, D. Nuclear architecture underlying gene expression in Trypanosoma brucei. *Trends Microbiol* **15,** 263–270 (2007).

201. Noma, K, Cam, H., Maraia, R. & Grewal, S. A role for TFIIIC transcription factor complex in genome organization. *Cell* **125,** 859–872 (2006).

202. Nora, E. e. a. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

203. Nora, E., Goloborodko, A, Valton, A., Gibcus, J., Uebersohn, A, Abdennur, N, Dekker, J, Mirny, L. & Bruneau, B. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–944.e22 (2017).

204. Németh, A, Guibert, S, Tiwari, V., Ohlsson, R & Längst, G. Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J* **27,** 1255–1265 (2008).

205. Németh, A, Conesa, A, Santoyo-Lopez, J, Medina, I, Montaner, D, Péterfia, B, Solovei, I, Cremer, T, Dopazo, J & Längst, G. Initial genomics of the human nucleolus. *PLoS Genet* **6,** e1000889 (2010).

206. Oddes, S, Zelig, A & Kaplan, N. Three invariant Hi-C interaction patterns: Applications to genome assembly. *Methods* **142,** 89–99 (2018).

207. Ogbadoyi, E, Ersfeld, K, Robinson, D, Sherwin, T & Gull, K. Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. *Chromosoma* **108,** 501–513 (2000).

208. Olins, A. & Olins, D. Spheroid chromatin units (v bodies). *Science* **183,** 330–332 (1974).

209. Osborne, C. e. a. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36,** 1065–1071 (2004).

210. Osborne, C., Chakalova, L, Mitchell, J., Horton, A, Wood, A., Bolland, D., Corcoran, A. & Fraser, P. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* **5,** e192 (2007).

211. Parelho, V. e. a. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132,** 422–433 (2008).

212. Pays, E. Regulation of antigen gene expression in Trypanosoma brucei. *Trends Parasitol* **21,** 517–520 (2005).

213. Peters, J., Tedeschi, A & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev* **22,** 3089–3114 (2008).

214. Phillips-Cremins, J. e. a. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153,** 1281–1295 (2013).

215. Pickersgill, H, Kalverda, B, de Wit, E, Talhout, W, Fornerod, M & van Steensel, B. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet* **38,** 1005–1014 (2006).

216. Pinger, J. e. a. African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. *Nat Microbiol* **3,** 932–938 (2018).

217. Pombo, A, Jackson, D., Hollinshead, M, Wang, Z, Roeder, R. & Cook, P. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J* **18,** 2241–2253 (1999).

218. Povelones, M., Gluenz, E, Dembek, M, Gull, K & Rudenko, G. Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in Trypanosoma brucei. *PLoS Pathog* **8,** e1003010 (2012).

219. Pérez-Morga, D, Amiguet-Vercher, A, Vermijlen, D & Pays, E. Organization of telomeres during the cell and life cycles of Trypanosoma brucei. *J Eukaryot Microbiol* **48,** 221–226 (2001).

220. Quinlan, A. & Hall, I. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

221. Ragoczy, T, Bender, M., Telling, A, Byron, R & Groudine, M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev* **20,** 1447–1457 (2006).

222. Rao, S. e. a. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

223. Rao, S. e. a. Cohesin Loss Eliminates All Loop Domains. *Cell* **171,** 305–320.e24 (2017).

224. Raska, I, Shaw, P. & Cmarko, D. Structure and function of the nucleolus in the spotlight. *Curr Opin Cell Biol* **18,** 325–334 (2006).

225. Razin, S., Gavrilov, A., Pichugin, A, Lipinski, M, Iarovaia, O. & Vassetzky, Y. Transcription factories in the context of the nuclear and genome organization. *Nucleic Acids Res* **39,** 9085–9092 (2011).

226. Reynolds, D, Hofmeister, B., Cliffe, L, Alabady, M, Siegel, T., Schmitz, R. & Sabatini, R. Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma brucei. *PLoS Genet* **12,** e1005758 (2016).

227. Rieder, D, Trajanoski, Z & McNally, J. Transcription factories. *Front Genet* **3,** 221 (2012).

228. Ritchie, K, Seah, C, Moulin, J, Isaac, C, Dick, F & Bérubé, N. Loss of ATRX leads to chromosome cohesion and congression defects. *J Cell Biol* **180,** 315–324 (2008).

229. Roditi, I, Carrington, M & Turner, M. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of Trypanosoma brucei. *Nature* **325,** 272–274 (1987).

230. Roukos, V & Misteli, T. The biogenesis of chromosome translocations. *Nat Cell Biol* **16,** 293–300 (2014).

231. Routh, A, Sandin, S & Rhodes, D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A* **105,** 8872–8877 (2008).

232. Sanborn, A. e. a. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112,** E6456–65 (2015).

233. Scahill, M., Pastar, I & Cross, G. CRE recombinase-based positive-negative selection systems for genetic manipulation in Trypanosoma brucei. *Mol Biochem Parasitol* **157,** 73–82 (2008).

234. Schalch, T, Duda, S, Sargent, D. & Richmond, T. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436,** 138–141 (2005).

235. Schenk, R, Jenke, A, Zilbauer, M, Wirth, S & Postberg, J. H3.5 is a novel hominid-specific histone H3 variant that is specifically expressed in the seminiferous tubules of human testes. *Chromosoma* **120,** 275–285 (2011).

236. Schoenfelder, S. e. a. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42,** 53–61 (2010).

237. Schulz, D, Mugnier, M., Paulsen, E., Kim, H., Chung, C., Tough, D., Rioja, I, Prinjha, R., Papavasiliou, F. & Debler, E. Bromodomain Proteins Contribute to Maintenance of Bloodstream Form Stage Identity in the African Trypanosome. *PLoS Biol* **13,** e1002316 (2015).

238. Schulz, D, Zaringhalam, M, Papavasiliou, F. & Kim, H. Base J and H3.V Regulate Transcriptional Termination in Trypanosoma brucei. *PLoS Genet* **12,** e1005762 (2016).

239. Selvaraj, S, R Dixon, J, Bansal, V & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat Biotechnol* **31,** 1111–1118 (2013).

240. Servant, N, Varoquaux, N, Lajoie, B., Viara, E, Chen, C., Vert, J., Heard, E, Dekker, J & Barillot, E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16,** 259 (2015).

241. Sexton, T, Yaffe, E, Kenigsberg, E, Bantignies, F, Leblanc, B, Hoichman, M, Parrinello, H, Tanay, A & Cavalli, G. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–472 (2012).

242. Shachar, S & Misteli, T. Causes and consequences of nuclear gene positioning. *J Cell Sci* **130,** 1501–1508 (2017).

243. Shaltiel, I., Krenning, L, Bruinsma, W & Medema, R. The same, only different - DNA damage checkpoints and their reversal throughout the cell cycle. *J Cell Sci* **128,** 607–620 (2015).

244. Siegel, T., Kawahara, T, Degrasse, J., Janzen, C., Horn, D & Cross, G. Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in Trypanosoma brucei. *Mol Microbiol* **67,** 762–771 (2008).

245. Siegel, T., Hekstra, D., Kemp, L., Figueiredo, L., Lowell, J., Fenyo, D, Wang, X, Dewell, S & Cross, G. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev* **23,** 1063–1076 (2009).

246. Siegel, T., Hekstra, D., Wang, X, Dewell, S & Cross, G. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res* **38,** 4946–4957 (2010).

247. Simms, T., Dugas, S., Gremillion, J., Ibos, M., Dandurand, M., Toliver, T., Edwards, D. & Donze, D. TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. *Eukaryot Cell* **7,** 2078–2086 (2008).

248. Simonis, M, Klous, P, Splinter, E, Moshkin, Y, Willemsen, R, de Wit, E, van Steensel, B & de Laat, W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38,** 1348–1354 (2006).

249. Splinter, E. e. a. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* **25,** 1371–1383 (2011).

250. Stevens, J. Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite* **15,** 226–232 (2008).

251. Stevens, T. e. a. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544,** 59–64 (2017).

252. Tabata, T & Iwabuchi, M. Molecular cloning and nucleotide sequence of a variant wheat histone H4 gene. *Gene* **31,** 285–289 (1984).

253. Taddei, A, Schober, H & Gasser, S. The budding yeast nucleus. *Cold Spring Harb Perspect Biol* **2,** a000612 (2010).

254. Tagami, H, Ray-Gallet, D, Almouzni, G & Nakatani, Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116,** 51–61 (2004).

255. Talbert, P. & Henikoff, S. Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **18,** 115–126 (2017).

256. Thambirajah, A., Li, A, Ishibashi, T & Ausió, J. New developments in post-translational modifications and functions of histone H2A variants. *Biochem Cell Biol* **87,** 7–17 (2009).

257. Thatcher, T. & Gorovsky, M. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res* **22,** 174–179 (1994).

258. Thoma, F, Koller, T & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol* **83,** 403–427 (1979).

259. Thompson, M, Haeusler, R., Good, P. & Engelke, D. Nucleolar clustering of dispersed tRNA genes. *Science* **302,** 1399–1401 (2003).

260. Tremethick, D. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128,** 651–654 (2007).

261. Udugama, M, M Chang, F., Chan, F., Tang, M., Pickett, H., R McGhie, J., Mayne, L, Collas, P, Mann, J. & Wong, L. Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res* **43,** 10227–10237 (2015).

262. Ukaegbu, U., Kishore, S., Kwiatkowski, D., Pandarinath, C, Dahan-Pasternak, N, Dzikowski, R & Deitsch, K. Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in P. falciparum. *PLoS Pathog* **10,** e1003854 (2014).

263. Umaer, K, Ciganda, M & Williams, N. Ribosome biogenesis in african trypanosomes requires conserved and trypanosome-specific factors. *Eukaryot Cell* **13,** 727–737 (2014).

264. Van Bortle, K & Corces, V. tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription* **3,** 277–284 (2012).

265. Van Bortle, K, Phanstiel, D. & Snyder, M. Topological organization and dynamic regulation of human tRNA genes during macrophage differentiation. *Genome Biol* **18,** 180 (2017).

266. Van Leeuwen, F, Wijsman, E., Kieft, R, van der Marel, G., van Boom, J. & Borst, P. Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. *Genes Dev* **11,** 3232–3241 (1997).

267. Van Steensel, B & Belmont, A. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169,** 780–791 (2017).

268. Vanhamme, L, Berberof, M, Le Ray, D & Pays, E. Stimuli of differentiation regulate RNA elongation in the transcription units for the major stage-specific antigens of Trypanosoma brucei. *Nucleic Acids Res* **23,** 1862–1869 (1995).

269. Vanhamme, L, Pays, E, McCulloch, R & Barry, J. An update on antigenic variation in African trypanosomes. *Trends Parasitol* **17,** 338–343 (2001).

270. Veitch, N., Johnson, P., Trivedi, U, Terry, S, Wildridge, D & MacLeod, A. Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. *BMC Genomics* **11,** 124 (2010).

271. Vietri Rudan, M, Barrington, C, Henderson, S, Ernst, C, Odom, D., Tanay, A & Hadjur, S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10,** 1297–1309 (2015).

272. Voon, H. & Wong, L. New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone. *Nucleic Acids Res* **44,** 1496–1501 (2016).

273. Wang, Q., Kawahara, T & Horn, D. Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. *Mol Microbiol* **77,** 1237–1245 (2010).

274. Wang, Y, Nagarajan, M, Uhler, C & Shivashankar, G. Orientation and repositioning of chromosomes correlate with cell geometry-dependent gene expression. *Mol Biol Cell* **28,** 1997–2009 (2017).

275. Wedel, C & Siegel, T. Genome-wide analysis of chromatin structures in Trypanosoma brucei using high-resolution MNase-ChIP-seq. *Exp Parasitol* **180,** 2–12 (2017).

276. Wedel, C, Förstner, K., Derr, R & Siegel, T. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J* **36,** 2581–2594 (2017).

277. Wendt, K. e. a. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451,** 796–801 (2008).

278. Wiedemann, S. e. a. Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J Cell Biol* **190,** 777–791 (2010).

279. Williams, R. e. a. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* **119,** 132–140 (2006).

280. Wingett, S, Ewels, P, Furlan-Magaril, M, Nagano, T, Schoenfelder, S, Fraser, P & Andrews, S. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4,** 1310 (2015).

281. Witt, O, Albig, W & Doenecke, D. Testis-specific expression of a novel human H3 histone gene. *Exp Cell Res* **229,** 301–306 (1996).

282. Wong, L. e. a. Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Res* **19,** 404–414 (2009).

283. Woods, A, Sherwin, T, Sasse, R, MacRae, T., Baines, A. & Gull, K. Definition of individual components within the cytoskeleton of Trypanosoma brucei by a library of monoclonal antibodies. *J Cell Sci* **93,** 491–500 (1989).

284. Wright, J., Siegel, T. & Cross, G. Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in Trypanosoma brucei. *Mol Biochem Parasitol* **172,** 141–144 (2010).

285. Xong, H., Vanhamme, L, Chamekh, M, Chimfwembe, C., Van Den Abbeele, J, Pays, A, Van Meirvenne, N, Hamers, R, De Baetselier, P & Pays, E. A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell* **95,** 839–846 (1998).

286. Yaffe, E & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* **43,** 1059–1065 (2011).

287. Yang, X, Figueiredo, L., Espinal, A, Okubo, E & Li, B. RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. *Cell* **137,** 99–109 (2009).

288. Zaremba-Niedzwiedzka, K. e. a. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541,** 353–358 (2017).

289. Zeiner, G., Sturm, N. & Campbell, D. Exportin 1 mediates nuclear export of the kinetoplastid spliced leader RNA. *Eukaryot Cell* **2,** 222–230 (2003).

290. Zhao, Z. e. a. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38,** 1341–1347 (2006).

291. Zomerdijk, J., Ouellette, M, ten Asbroek, A., Kieft, R, Bommer, A., Clayton, C. & Borst, P. The promoter for a variant surface glycoprotein gene expression site in Trypanosoma brucei. *EMBO J* **9,** 2791–2801 (1990).

292. Zorn, C, Cremer, C, Cremer, T & Zimmer, J. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res* **124,** 111–119 (1979).

293. Zuin, J. e. a. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111,** 996–1001 (2014).

Affidavit

I hereby confirm that my thesis entitled 'The role of nuclear architecture in the context of antigenic variation in *Trypanosoma brucei*' 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 or similar form.

Würzburg, May 20, 2019

Laura Müller-Hübner

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation 'Über die Rolle der Zellkernarchitektur im Kontext von Antigenvariation in *Trypanosoma brucei*' eigenständig, d.h. insbesondere selbstständig 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.

Würzburg, May 20, 2019

Laura Müller-Hübner