This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
HACking centrochromatin: on the relationship between centromeres and repressive chromatin

Nuno Miguel Marques Vitória Cabrita Martins

Thesis submitted for the Degree of Doctor of Philosophy
The University of Edinburgh

December 2014
Declaration

I declare that this thesis was composed entirely by myself, and that the work documented within is my own, except where contributions by others were made (which are clearly stated).

This work has not been submitted for any other degree or professional qualification.

Nuno Miguel Marques Vitória Cabrita Martins

Edinburgh

December 2014
I would like to thank the following people for their significant contributions to my PhD work and this thesis:

My supervisor Bill Earnshaw, for allowing me the honour of joining his lab and become part of his research team. I have learned so much here, and had the pleasure to engage, learn from and work with some of the most experienced people in this field of science, as well as having access to the latest state-of-the-art technology. I feel I have truly been fortunate for the chance to work with him. I thank him for being always on the lookout for the well-being of my career progress, and his valuable advice on how to achieve success. I do not know what he saw in me, to offer me this opportunity, but I am very happy he put his trust in me.

All the members of the Earnshaw lab, for sharing their knowledge, expertise and advice on how to tackle all sorts of problems. A special mention goes to Jan Bergmann, for his patience and diligence, in teaching me all he knew, and making sure I was prepared to meet the challenges ahead.

Hiroshi Masumoto, and all the members of his lab, for receiving me in their lab, showing me the beautiful country of Japan, and having the patience to teach me all the secrets of the HAC. A special thanks goes to Masumoto-sensei, for his faith in me and many stimulating discussions.

The FCT, for awarding me my PhD scholarship, and allowing me to pursue my career interests abroad, in the wonderful city of Edinburgh.

Lars Jansen, for teaching me how to be a scientist, and believing in me. For the great professional advice he gave me, and pushing me out of my comfort zone. My experience in his lab at the IGC is what really opened me up to the world of science. I would not be here now if it hadn't happened.

And lastly, to my girlfriend Nadine, who has been my pillar for all those difficult and stressful times. For all the support, the scientific discussions, and for helping me find the way when things looked grim. I am extremely proud of the scientist she is, and is a privilege to have her by my side. I would not be here if it wasn't for her.
Lay Summary

The genetic information that makes up the cells of every living organism is contained within long threads of DNA, called chromosomes. These chromosomes are held within each cell in a structure called the nucleus. Whenever the cell needs to multiply, it must divide into two identical cells. To ensure both daughter cells possess the same genetic information (to possess a full encyclopaedia on how to make everything in an organism), each chromosome has to be duplicated individually and then equally distributed between the two dividing cells. During cell division, the cell relies on several molecular mechanisms that orient and separate chromosomes so that each daughter cell receives the exact same number of chromosomes. Flaws in this process can lead to a variety of problems such as the onset of cancer and development issues, which can cause miscarriages and birth defects.

Chromosomes encode information in the form of chemical sequences within the DNA thread, called genes. To control the flow of information from these genes the cell uses chemical 'bookmarks', called chromatin marks. These chromatin marks effectively control the chromosome: whether genes are read by the cell or not, which genes should be active in each cell type of the body, orchestrate the replication of chromosomes before division, control repair of damaged areas on the chromosome, and ultimately to organize the chromosome, both within the nucleus and during the delicate process of cell division. Disruption of chromatin regulation is at the heart of many genetic diseases and cancer.

The molecular mechanism of cell division relies on complex molecular machines, which allow chromosome copies to orient parallel to the plane of division. This allows each copy of a chromosome to be pulled into its corresponding daughter cell. To pull the two copies of a chromosome into the two daughter cells, they are attached to threads that spread out from each pole of the division axis. These threads bind to the chromosome on a region called the centromere, easily recognized as the centre of the 'X' shape of chromosomes. It is a special region of the chromosome, in that its location is not dictated by information on the DNA under it. Instead, it is the chromatin signature of centromeres that allows them to be maintained at their locations, because that signature is passed on through successive generations and throughout millions of years. This is called an 'epigenetic mechanism'. But because their location does not necessarily require a certain DNA sequence, centromeres can move to several locations along the chromosome and have been shown to do so throughout evolution. The mechanisms that allow this to happen are still largely unknown, as well as
how the centromeres response to large-scale chromatin mis-regulation in cancer cells. Current technology allows for high-resolution mapping of chromatin marks across the length of chromosomes. However, mapping of these marks at centromeres is complicated because centromeric regions are very difficult to ‘read’ by most current methods.

In this work I sought to further understand the chromatin marks involved in granting functionality to centromeres and also which ones can support formation of new centromeres. To that end, I used chicken cells that have 3 centromeres that can be mapped at high accuracy, and a Human Artificial Chromosome (HAC) in human cells, whose chromatin can be independently manipulated. I confirmed earlier observations and found new surprising connections of centromeres with chromatin marks that silence genes, which might be a defining criteria to where they prefer to form. On the other hand, I verified previous results that centromeres also are 'active' and 'read' by the cell (like genes), but I found that they manage to keep the required 'active' chromatin marks even when surrounded by different types of 'silencing' marks, like islands in a quiet sea.
Abstract

The centromere is a chromosomal locus required for accurate segregation of sister chromatids during cell division. They are maintained epigenetically in most eukaryotes, by incorporating the H3 variant CENP-A, and can, in rare instances, change location on the chromosome throughout generations. Centromeres are transcribed, and an active transcription chromatin signature is required for centromere maintenance. For this reason, insight into the nature of this so-called “centrochromatin” is essential for understanding a centromere’s place in the chromosome. The body of work contained in this thesis shows my efforts to understand the centromere in the context of chromatin, revealing interactions and new evidence for repressive chromatin domains with centromere activity, in two different vertebrate models: chicken DT40 cells and human HeLa cells.

Centromeres are generally embedded within large domains of heterochromatic repetitive sequences in most eukaryotes, and mapping “centrochromatin” to high-resolution has proven difficult. However, chromosomes 5, 27 and Z of Gallus gallus are not located within repeat arrays, and are fully sequenced. CENP-A distribution on these centromeres has been mapped by ChIP-seq, and I have performed ChIP against selected histone modifications as part of a collaboration. While levels of heterochromatin are naturally quite low in these centromeres, I have shown that repressive polycomb chromatin instead is enriched in these non-repetitive centromeres, suggesting a replacement of one silenced chromatin state with another. Additional mapping of these centromeres showed a pattern of active chromatin marks distinct from that reported for human cells, which exhibited dynamic distribution throughout the cell cycle. Furthermore, conditionally generated neocentromeres in DT40 cells revealed that centrochromatin formation lowers, but does not eliminate, active transcription.

To directly study the interaction of polycomb and heterochromatin with centrochromatin, I used a synthetic Human Artificial Chromosome (HAC), which allows for specific conditional targeting of chromatin modification enzymes, allowing manipulation of the underlying chromatin. Enrichment of the polycomb chromatin state on the HAC centromere, by EZH2 tethering, reduced its active transcriptional chromatin signature, but did not impair its actual transcription or mitotic activity. However, direct tethering of polycomb secondary silencing effector PRC1 caused centromere loss, and this effect was
mimicked with homologous heterochromatin factors, indicating that centromeres can subsist within repressive chromatin domains, but are lost when direct repression is applied.

To understand the contribution of the local repressive heterochromatin to centromere stability, I erased heterochromatin marks from the HAC centromere by tethering JMJD2D (an H3K9me3 demethylase): long-term (but not short-term) heterochromatin loss impaired CENP-A assembly, perturbed mitotic behaviour, and resulted in significant HAC mis-segregation. These results strongly suggest that local heterochromatin is essential to maintain normal CENP-A dynamics and centromere function.

Together with previous observations, these data suggest that a repressive chromatin environment contributes to centromere stability, and that centromeres likely have natural mechanisms to maintain their transcriptional activity within such domains.
Abbreviations

α-Sat  Alpha-satellite
APC  Anaphase-promoting complex
BAC  Bacterial artificial chromosome
bp  Base pair(s)
BleoR  Bleomycin/Zeocin resistance gene
BSA  Bovine serum albumin
Bsr  Blasticidin S resistance
C. albicans  Candida albicans
CATD  Centromere targeting domain
CCAN  Constitutive centromere-associated network
cDNA  Complementary DNA
CENP  Centromere protein
Cdk  Cyclin-dependent kinase
ChIP  Chromatin immunoprecipitation
ChIP-seq.  ChIP massive parallel sequencing
Chr.  Chromosome
CIN  Chromosome instability
CIP  Calf intestinal phosphatase
cDMEM  Complete DMEM medium
CPC  Chromosome passenger complex
cRPMI  Complete RPMI 1640 medium
D. melanogaster  Drosophila melanogaster
DNA  Deoxyribonucleic acid
dNTP  Deoxynucleotide triphosphate
Dox  doxycycline
dsDNA  Double-stranded DNA
E. coli  Escherichia coli
EDTA  Ethylenediaminetetraacetic acid
EYFP  Enhanced yellow fluorescent protein
FIAU  (5-iodo-2'-fluoroarauracil)
FISH  Fluorescent in situ hybridization
FITC  Fluorescein isothiocyanate
G. gallus  Gallus gallus
HAC  Human artificial chromosome
HAT  Histone acetyl transferase
HDAC  Histone deacetylase
HSV-tk  Herpes Simplex Virus thymidine kinase
kb  Kilobase(s)
KMN  KNL1-Mis12-Ndc80
KMT  Lysine methyltransferase
KT  Kinetochore
LB  Luria Bertani broth
LAD  Lamin-associated domain
LMN  Lamina
M  molar
Mb  Megabase(s)
mg  milligram
ml  millilitre
mM  millimolar
MT  Microtubule
N. crassa  Neurospora crassa
NEB  New England Biolabs
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nM</td>
<td>nanomolar</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>O. sativa</td>
<td>Oryza sativa</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PBS-T</td>
<td>PBS triton X-100</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative PCR</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>PFA</td>
<td>paraformaldehyde</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse transcription quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium-dodecyl sulphate</td>
</tr>
<tr>
<td>sec</td>
<td>Second(s)</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>S. pombe</td>
<td>Schizosaccharomyces pombe</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>TetO</td>
<td>Tetracycline operator</td>
</tr>
<tr>
<td>TetR</td>
<td>Tetracycline repressor</td>
</tr>
<tr>
<td>TN-16</td>
<td>3-(1-anilinoethyldiene)-5-benzylpyrrolidine-2,4-dione</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>U</td>
<td>(enzyme) units</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra violet</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µl</td>
<td>microlitre</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>YAC</td>
<td>Yeast artificial chromosome</td>
</tr>
</tbody>
</table>
# Table of contents

Declaration ............................................................................................................. II
Acknowledgements .......................................................................................... III
Lay Summary ...................................................................................................... IV
Abstract ........................................................................................................... VI
Abbreviations ..................................................................................................... VIII
Table of contents .............................................................................................. X

## 1. Introduction .................................................................................................

### 1.1. Chromosomes: an overview ................................................................. 1

#### 1.1.1. The structural composition of chromosomes .................................. 1
#### 1.1.2. Chromosome structural changes across the cell cycle ................... 1
#### 1.1.3. Consequences of chromosome mis-segregation ............................. 2

### 1.2. Mitosis .................................................................................................... 3

#### 1.2.1. The stages of mitosis ..................................................................... 3
#### 1.2.2. The centromere and kinetochore .................................................. 5
#### 1.2.3. Error correction at the kinetochore .............................................. 6
#### 1.2.4. The mitotic checkpoint .................................................................. 7
#### 1.2.5. Centromeric cohesion .................................................................... 8

### 1.3. Chromatin regulation ............................................................................ 9

#### 1.3.1. Chromatin marks .......................................................................... 9
#### 1.3.2. Epigenetic memory ....................................................................... 11
#### 1.3.3. Nucleosome positioning and structural RNA ............................... 11
#### 1.3.4. Chromatin states: overview ......................................................... 12
#### 1.3.5. Euchromatin states ..................................................................... 13
#### 1.3.6. Heterochromatin states ............................................................... 16
#### 1.3.7. Chromatin modulation during genomic maintenance processes .... 20

### 1.4. The Centromere .................................................................................... 21

#### 1.4.1. CENP-A ..................................................................................... 21
#### 1.4.2. The CCAN: the foundation of kinetochores .................................. 22
#### 1.4.3. Cell cycle regulation of centromeres ............................................ 26
#### 1.4.4. Centromeres and disease ............................................................. 27

### 1.5. Centromatin .......................................................................................... 28

#### 1.5.1. The epigenetic nature of centromeres ......................................... 28
#### 1.5.2. Conservation of centromere organization in eukaryotes .............. 29
#### 1.5.3. Implication of chromatin state in centromere function .............. 36
#### 1.5.4. Artificial chromosomes and de novo nucleation of centromeres .... 39

### 1.6. Aims of this work ................................................................................ 42

#### 1.6.1. Chromatin mapping of non-repetitive chicken centromeres .......... 43
#### 1.6.2. Centromere activity can coexist with polycomb chromatin .......... 43
#### 1.6.3. Heterochromatin stabilizes the core centromere and prevents DNA damage 44

## 2. Material & Methods .................................................................................

### 2.1. Materials ............................................................................................... 45

#### 2.1.1. Buffers, solutions & reagents ....................................................... 45
#### 2.1.2. Bacterial and eukaryotic cell culture ........................................... 46
#### 2.1.3. Oligonucleotide primers ............................................................... 48
#### 2.1.4. Primary antibodies ..................................................................... 50
#### 2.1.5. Secondary antibodies ................................................................. 51
#### 2.1.6. Commercial kits ......................................................................... 51

### 2.2. Molecular biology techniques .............................................................. 51

#### 2.2.1. Preparation of bacterial plasmid DNA ......................................... 51
#### 2.2.2. Molecular cloning of plasmid constructs ...................................... 52
#### 2.2.3. Sequencing of plasmids ............................................................... 53
#### 2.2.4. Reverse transcriptase reaction ..................................................... 53
1. Introduction

1.1. Chromosomes: an overview

1.1.1. The structural composition of chromosomes

The genetic information carried by DNA, within each living cell, is contained within chromosomes. They are the units by which genetic information is partitioned between dividing cells, and correct regulation of that process is crucial to ensure maintenance of genetic information across generations.

Chromosomes are macromolecular composites, consisting of DNA, a complex multi-protein packaging structure, accessory RNAs and a myriad of enzymes that help regulate chromosome function. This dynamic and complex molecular aggregate that makes up chromosomes is termed chromatin. The basic unit of chromosome organization is the nucleosome, composed of a positively-charged core octamer of histone proteins (H2A, H2B, H3 and H4, two copies of each), around which 146 bp of negatively-charged DNA is wrapped. To successfully enclose several hundred centimetres of DNA within a cell nucleus a few microns wide, eukaryote genomes are highly folded in higher-order arrangements, in addition to packaging via nucleosomes. The regulation of chromosome dynamics and structure is invariably involved in most processes that concern DNA in eukaryotic cells.

1.1.2. Chromosome structural changes across the cell cycle

To partition chromosomes equally to each daughter cell, a highly complex set of specific machinery is required, as well as adequate regulation of chromosome structure itself, in the stages leading up to and throughout cell division.

During DNA replication, each chromosome is duplicated into two sister chromatids. Upon passage of the replication fork, the cohesin complex is thought to form a ring around
both chromatids, tethering them together. This physical linkage between both copies of each chromosome ensures they can both be oriented properly and equally partitioned to separate daughter cells upon the oncoming cell division.

Chromosomes undergo higher-order mitotic condensation when the cell division process starts. The mechanism is unknown, but condensin and topoisomerase II play a significant role. This physically compacts and resolves chromosomes from each other, which allows for their individual spatial reorientation so that each sister chromatid will face an opposite pole of the cell. This process also causes chromosomes to become in their vast majority transcriptionally inert.

Subsequent breakage of cohesin linkages between chromatids allows them to be separated to opposite poles, distributing equal copies of genetic information to the two forming daughter cells. These cells now decondense their chromosomes and resume interphase genomic functions.

1.1.3. Consequences of chromosome mis-segregation

The process of segregating genetic information is crucial to ensure healthy cellular functions. Disruption of this process most often leads to cell death or impaired growth, but it can also be a driver of evolution: mis-segregation of a single chromosome can render one of the daughter cells with surplus genetic information, which can potentially grant it beneficial (as well as deleterious) effects. In multicellular organisms, however, such genetic imbalance can eventually allow cells to escape checkpoint controls, to grow and divide unchecked, possibly developing into a tumour. Malfunctioning checkpoints and/or mitotic control mechanisms can lead to high rates of chromosome breakage, a highly mutagenic event. Cancerous cells, which normally have suffered many rounds of mis-segregation and acquired mutations that allow them to bypass checkpoint mechanisms, often have as a result...
greatly rearranged chromosomes. Therefore, it is essential for cell and organismal survival to ensure that chromosome segregation fidelity is maintained.

1.2. Mitosis

1.2.1. The stages of mitosis

Mitosis consists of four phases: prophase, metaphase, anaphase and telophase. Progression and fidelity of mitosis is controlled by several mechanisms, mainly through phosphorylation, through control of protein localization and/or degradation events.27,28

For the purposes of this work, I will focus primarily on the processes that concern the control of chromosomal movement and attachment to the mitotic spindle.

Prophase starts with chromosome condensation, as explained previously. In higher eukaryotes, the nuclear envelope subsequently breaks down, releasing chromosomes into the cytoplasm.29 This allows chromosomes to contact the mitotic microtubule network, which takes a characteristic bipolar spindle shape.30 Microtubules polymerizing from each pole contact the chromatids from each chromosome, through a protein super-structure called the kinetochore. Kinetochore are multi-protein complexes that are assembled on each chromatid at the centromere, which defines the primary constriction zone connecting both chromatids.31

Metaphase is characterized by the successful congression of chromosomes to the midzone of the cell, and fine-tuning the alignment of sister kinetochores to opposite spindle poles. The transitional stage of "prometaphase" is sometimes used to describe the progressive attachment and congression of chromosomes to the metaphase plate at the cell equator. Orientation of chromatids is controlled locally at the kinetochore, via sensing
mechanisms that detect both the kinetochore-microtubule attachment and the tension force across centromeres, which ultimately promote bi-orientation of chromatids in relation to spindle poles. Once this process is complete, activation of the APC (anaphase promoting complex) triggers progression into the next phase of mitosis $^{32,33}$.

Anaphase onset starts the process of sister chromatid separation and migration to opposite poles. The APC starts a regulatory cascade, by promoting degradation of several protein targets, chief among them securin and cyclin B $^{28}$. The former elicits separation of sister chromatids, and the latter triggers spindle elongation and microtubule catastrophe at kinetochores, which nonetheless remain bound to the depolymerizing microtubules. This effectively separates chromatids into opposite poles of the cell $^{34}$.

Telophase is the final stage of mitosis, where chromosomes decondense and re-start interphase genomic functions, and the two daughter cells are partitioned from each other. In higher eukaryotes, the reconstitution of the nuclear envelope around chromosomes takes place also in telophase $^{23}$.

Cytokinesis is the process that partitions the two forming daughter cells. This is achieved through a cleavage furrow that starts along the midzone, and eventually severs off each daughter cell's cytoplasm, effectively generating two new independent cells $^{35}$. The remnants of the central spindle accumulate at the intercellular bridging point, forming a transient structure called the midbody, which mediates final abscission and separation of the two cells $^{36}$.
1.2.2. The centromere and kinetochore

Centromeres are located at the primary constriction of chromosomes, connecting the two sister chromatids together, and are essential for chromosome segregation. They are particular genomic regions whose purpose is to control chromatid separation and grant directed mitotic mobility to chromosomes. The majority of eukaryotes possess "monocentric" chromosomes, that is, one single centromere on each chromosome. However, some species possess diffuse centromeres throughout the length of chromosomes, termed "holocentric" chromosomes, which show no evidence of a primary constriction, and microtubule binding occurs across the whole chromatid \(^{37}\). For the purpose of this work, however, I will focus solely on monocentric chromosomes.

The kinetochore, a complex multi-protein structure that assembles at centromeres, allows for chromosomes to interact with the mitotic segregation machinery. Kinetochore-microtubule attachment is a highly regulated, dynamic process, with an interplay between several kinetochore sub-complexes and microtubule-end binding proteins. Spindle microtubules are highly dynamic, alternating between polymerization and depolymerization, and the current interpretation is that this allows search-and-capture interactions with kinetochores on chromosomes \(^{38,39}\).
The outer proteins of the kinetochore are responsible for capture and attachment to microtubules. The major microtubule-binding protein complex, named the KMN network (for its components KNL-1, Mis12, Ndc80), binds directly to microtubule polymerizing ends, in an end-on attachment.

1.2.3. Error correction at the kinetochore

![Diagram of kinetochore attachment and error correction](image)

**Fig. 1.1: Attachment and error correction at the kinetochore**

Schematic comparing positional and regulatory differences in centromeres under tension or otherwise. KT-MT attachments which are not bipolar in orientation do not create tension, causing the kinetochores to remain proximal to Aurora B at the inner centromere. Aurora B-dependent phosphorylation of KT subunits causes attachments to be dynamic and unstable. When attachments create tension, KTs are pulled away from Aurora B and attachments become stable. Adapted from Tanaka *et al*, 2010.

The kinetochores of each sister chromatid possess the ability to detect if they are not properly oriented towards opposite poles. Sister kinetochores can perform four types of attachment orientation relative to the spindle poles: amphitelic (each kinetochore is attached to a single, opposite pole), monotelic (only one kinetochore is attached), syntelic (both...
kinetochores attached to the same pole), and merotelic (both kinetochores are attached to opposite poles, but one or both are attached to microtubules emanating from the opposite pole as well) \(^ {43}\).

To sense incorrect attachments, kinetochores couple their efficiency to bind microtubules to the tension between centromeres, so that attachments can only be stabilized in correct, amphitelic orientations (fig. 1.1). The current model holds that activity of the KMN network is regulated through phosphorylation by Aurora B kinase, which reduces its ability to bind microtubules \(^ {44-46}\). Aurora B is the enzymatic component of the CPC (chromosome passenger complex): a multi-subunit complex that regulates many diverse targets and has a dynamic localization during mitosis. The CPC ensures mitotic fidelity, by regulating the correct attachment of chromatids to opposite spindle poles, and is required for mediating the final steps that divide the two daughter cells \(^ {47-50}\). The CPC localizes to the inner region of centromeres, between sister kinetochores, in prometaphase. So at rest, when kinetochores are close to Aurora B, the net result will be only transient kinetochore-microtubule interactions, which facilitates their release from microtubules.

Once tension starts to be established between sister kinetochores, however, centromeres physically stretch polewards, and the outer kinetochore moves away from Aurora B's field of activity, causing the KMN proteins to be dephosphorylated \(^ {51}\). PP1, which binds unphosphorylated KNL1, seems to contribute by further decreasing phosphorylation levels \(^ {52}\), which efficiently increases microtubule-binding ability and stabilizes attachments under tension.

1.2.4. The mitotic checkpoint

To ensure chromosomes are properly aligned before anaphase onset, a "wait" molecular signal is generated which postpones APC activation and thus progression into anaphase. Unattached kinetochores produce a checkpoint signal, in the form of the MCC
(mitotic checkpoint complex), which binds and inhibits the APC, thereby preventing mitotic progression until all kinetochores are successfully attached\textsuperscript{46,53-56}.

Through KNL1, the kinetochore recruits the Bub1 kinase, which forms the platform of the checkpoint machinery, by recruiting other components, such as BubR1, Bub3, Mad1 and Mad2, for as long as the kinetochore remains unattached. Current understanding of the mitotic checkpoint posits that interaction of Mad2 with the checkpoint machinery at unattached kinetochores induces a conformational change in the protein. Joined by BubR1 and Bub3, this pool of Mad2 sequesters the APC activator Cdc20, thus forming the complete cytosolic MCC, and preventing APC activity and anaphase onset\textsuperscript{57,58}.

Once successfully attached, kinetochores are "stripped" of the mitotic checkpoint machinery, through relocation of Mad1-Mad2 away from kinetochores by the RZZ complex and dynein\textsuperscript{59}. In this way, as long as one single kinetochore is still unattached, functional MCC will be continuously generated, effectively preventing the APC from inducing irreversible entry into anaphase.

As soon as all kinetochores are attached to microtubules, the mitotic checkpoint is satisfied, and the MCC is no longer produced\textsuperscript{60-62}. The APC is therefore no longer inhibited, and proceeds to irreversibly trigger mitotic progression into anaphase.

1.2.5. Centromeric cohesion

Another function of centromeres is to provide a connective point between the two sister chromatids, via accumulation of cohesin. By restraining their separation, this cohesion couples the pulling forces from opposite spindle poles to generate tension between the centromeres of sister chromatids and thus promote their correct alignment to the future daughter cells.

In vertebrate mitosis, most cohesin is lost from chromosome arms in prophase, being retained only at centromeres\textsuperscript{63}. The removal of cohesin from chromosome arms is triggered
by Aurora B and Plk1-dependent phosphorylations. Cohesin is retained only around the centromere, due to the protective action of centromere-associated Sgo1-PP2A, which removes the phosphorylations mentioned above. Sgo1 itself is recruited to centromeres due to the actions of the Bub1 kinase, which is part of the kinetochore. The resulting cohesin protection extends throughout the centromere region.

Full release of sister chromatids from each other occurs in anaphase, primarily via cohesin degradation, caused by the enzyme separase. Separase is normally held inactive until anaphase onset, at which point its inhibitor, securin, is targeted for degradation by the now-active APC.

1.3. Chromatin regulation

1.3.1. Chromatin marks

The expression of genetic information encoded in DNA, as well as DNA replication, recombination and repair, all rely on physical access to DNA. As such, these processes must interact with and navigate the building blocks of chromatin, the nucleosomes. Nucleosomes are not simply structural packaging proteins for DNA: they are at the center of most processes that interact with the genome.

Control of nucleosome dynamics and chromatin processes hinges greatly on histone modifications. These are post-translational modifications to aminoacid residues, most often on the flexible N-terminal tails which protrude from the nucleosome core, with a few also possible on the histone core itself. The most common are methylation (in mono-, di- or trimethylated state), acetylation, phosphorylation and mono-ubiquitination. Several enzymes have been identified that catalyze either the formation or removal of such marks.

Some modifications, such as acetylation and phosphorylation, directly affect chromatin structure by reducing the positive charge of histones and thus disrupting electrostatic interactions between them and DNA. This has the effect of creating a less
compact and more accessible chromatin region. There are many acetylation sites in histone tails, and genomic regions with hyper-acetylated histones are generally highly expressed or otherwise very "open" in structure \(^71\).

Histone modifications also provide docking sites for particular structural proteins and enzymes ("readers" of marks) \(^72,73\). The interactions depend on the specific residue and specific modification. This permits recruitment in \(c\text{is}\) of multiple specific functions, depending on the combination of marks on the nucleosome. There is a high degree of cross-talk between different modifications, in at least three ways: 1) the same residue can be modified in different ways, such as methylation, acetylation or ubiquitylation; 2) one modification can influence the function of others, whether directly through molecular interaction or indirectly by recruiting an effector protein; 3) modifications on one residue can affect binding of factors to adjacent residues.

In addition to post-translational modifications, another level of histone-based regulation exists: histone variants \(^74\). Unlike their canonical counterparts, which constitute the majority of chromatin and are incorporated unspecifically during replication, histone variants are recruited and assembled into chromatin by particular processes. Their structural differences from canonical histones grant them the status of an effective chromatin mark, and like histone modifications, they can instigate unusual nucleosome dynamics or recruit downstream factors to the locus they are inserted in.

One more level of chromatin control exists, directly at the DNA level: methylation of DNA bases \(^75\). This mark is generally associated with local transcriptional repression, as it can both interfere with binding of a transcription factor to its recognized DNA sequence, and also recruit repressor proteins to effect transcription silencing in \(c\text{is}\) \(^76\).

The accumulation of correlations between chromatin marks and genomic processes has led to the hypothesis of a "chromatin code", in which combinatorial patterns of marks, and the factors they recruit, determine the state of each genomic region \(^77\).
1.3.2. Epigenetic memory

Epigenetics is generally defined as "changes in gene function that are heritable upon cell division and that cannot be explained by changes in DNA sequence" 78. It was first coined by C. H. Waddington to explain how, in the process of development in multicellular organisms, the same genotype can give rise to different phenotypes in the form of cell lineages, implicating a degree of "memory" in these differentiating cells.

In the recent decades, the study of chromatin processes has given insight into how organisms can effect "epigenetic memory". Many chromatin marks can, through the factors they recruit, establish feedback loops that essentially maintain those marks, giving rise to stable chromatin states, which can comprise several marks in combination. It is still a matter of debate, for many stable chromatin states, which marks/factors are cause or consequence of the other, due to their consistent association with each other throughout a genome 79.

Many chromatin states can be perpetuated across cell division and even across whole-organism generations 80. The stability of each mark or chromatin state can differ depending on the type of process it is associated with, and may be more or less susceptible to stimuli response. It should be said that, although "memory" can be mediated by chromatin without changes in the DNA, sequence features such as transcription factor binding sites or repetitive DNA can severely affect local chromatin environment, thus dictating their own chromatin environment in a sequence-based manner.

The current definition of epigenetics, from a chromatin perspective, is "a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" 81, specifically implying the heritable memory is still carried on chromosomes.

1.3.3. Nucleosome positioning and structural RNA

For many chromatin processes, changing the arrangement and position of nucleosomes is paramount, given that they can physically inhibit access to DNA by other
DNA-binding proteins. Nucleosome positioning can be strongly affected by underlying DNA sequences, such as how more or less-bendable stretches of DNA bases may favour wrapping around histones\(^8\). Additionally, chromatin factors can also induce active changes or maintenance of positioning in nucleosomes. This is mediated by so-called chromatin remodellers, which can slide nucleosomes along DNA or even promote their eviction\(^8\).

RNA can also play a role in chromatin regulation, especially non-coding RNAs\(^8\). Long non-coding RNAs can act as structural scaffolds or provide targeting activities to other chromatin enzymes. For instance, the RNA interference machinery in *S. pombe*, which detects double stranded RNAs, resulting from expression of repetitive elements or convergent transcription, processes those RNAs and utilizes them as a sequence-based targeting mechanism to DNA, to effect directed transcriptional repression\(^8\). Another example is X-inactivation in female mammals, which exists to cope with the presence of an extra copy of the X chromosome in the cell, as a means of transcriptional dosage compensation\(^8\). The long non-coding RNA *Xist*, expressed from the X chromosome copy that will become inactive, leads to its transcriptional inactivation. The *Xist* RNA physically coats that copy of the X chromosome, and recruits several silencing factors in *cis*, eventually causing transcriptional repression across the chromosome.

### 1.3.4. Chromatin states: overview

The explosion of sequencing technology and genome-wide studies of chromatin marks have led to vital insights in understanding the functions and interplay of chromatin signatures with the information contained in the DNA sequence. This has allowed the mapping and association of chromatin marks, or combinations thereof, with certain transcriptional states within the genome (fig.1.2)\(^8\)–\(^9\). See section 6.1. "Appendix: Chromatin state signatures" for a detailed quantification of chromatin mark patterns across chromatin states.
Two main chromatin states were classically described before: euchromatin and heterochromatin, corresponding to actively transcribed and repressed chromatin, respectively. Since then, pioneering work in *D. melanogaster* and further studies performed in human cells, have expanded this view, allowing the identification and dissection of a number of distinct chromatin states and sub-states.

**Fig. 1.2: Chromatin states in higher eukaryotes**  
Summary of studies on chromatin states in *D. melanogaster* and human cells. States are separated by colour, yellow-red composing actively transcribing, euchromatin states, and blue, green and black composing repressed, heterochromatin states. Key regulatory chromatin proteins of each are listed as well as the more well-characterized histone modifications and general characteristics. Note that there is some degree of variation within each state concerning mark/regulator density. HATs: histone acetyltransferases; KMTs: lysine methyltransferases; HDACs: histone deacetylases; LMN: nuclear lamina.

### 1.3.5. Euchromatin states

Euchromatin corresponds to actively transcribed chromatin. As classically described, through microscopy, euchromatin represented the fraction of the nuclear material that shows
less compaction and clustering, as opposed to heterochromatin. Euchromatin regions are more gene-rich and easily transcribed.

As it is now believed, the broad chromatin state described as euchromatin represents a nuanced range of various sub-states, depending on level of transcription, amount of introns in the gene body, dosage compensation, topological looping with nearby genomic regions, among others.

One key characteristic of actively transcribed chromatin is higher levels of overall histone acetylation. This is promoted by HATs (histone acetyl-transferases), which are essential for chromatin decompaction, allowing other members of the chromatin machinery access to DNA. Several residues in both the tails of H3 and H4 can be acetylated, and the vast majority is correlated with promotion of transcription, or at the very least, resistance to transcriptional repression. Histone acetylations are reversible and also highly dynamic, thanks to the actions of HDACs (histone deacetylases), resulting in a balanced fine-tuning of the transcription process.

1.3.5.1. **Transcription start site**

Transcription activation starts with events at the TSS. It is located immediately downstream of the promoter of genes, and it is where RNA polymerase assembles onto DNA. In order to do so, the nucleosomes at the TSS must be decompacted through acetylation, and remodelled via sliding or eviction to provide access to DNA. As a result, TSSs show the highest levels of histone acetylation and have a very low histone density, and are generally dubbed a "nucleosome-free region". To achieve this, a great interplay of chromatin marks occurs at the TSS, but for the purpose of this work, I will focus on the most prevalent and well described: high levels of H3K4me3, nucleosomes containing an unstable H2A variant, and multiple acetylations on H3 and H4, together with...
absence of repressive marks (H3K9me3, H4K20me3, H3K27me3 and DNA methylation).

H3K4me3, catalyzed by the MLL/trithorax complexes, acts as a binding site for many HAT complexes, and has been considered a memory and initiator mark for the transcription activation cascade, at the chromatin level. High level of local HAT recruitment leads to a very open chromatin state. Acetylation of the aforementioned H2A variant at the TSS nucleosomes (called H2A.Z in vertebrates) causes its nucleosome to become highly unstable, facilitating displacement or eviction by SWI/SNF and ISWI chromatin remodellers. This allows better access of activating transcription factors and later binding of RNA polymerase to the locus. Once bound to the TSS, the RNA polymerase initiation complex will then assemble, and if activation conditions are adequate, convert into an elongating form which progresses downstream through the gene, catalyzing formation of a RNA strand complementary to the underlying DNA.

1.3.5.2. Gene body

The downstream body of transcribed regions presents its own combination of marks, in some ways similar to the TSS-proximal 5' region. Many are established by factors associated with the passing RNA polymerase, such as histone acetylation, H3K4me2, H3K79me2/3 and H3K36me2/3. The latter recruits the Rpd3 HDAC, whose function seems to be to deacetylate and "close up" the chromatin after RNA polymerase passage, thus preventing deleterious transcription initiation from cryptic promoters within the gene body.

Transcription can induce significant remodelling of nucleosomes, as they represent a physical barrier to this process. The RNA Pol. II-associated FACT complex can both destabilize and re-assemble the H2A-H2B dimers in nucleosomes. This is believed to result in a more flexible nucleosome structure, or even eviction and re-construction of nucleosomes.
in the wake of RNA polymerase passage \textsuperscript{113}. In addition to this, transcription-dependent histone eviction is accompanied by compensatory deposition of new histones, in a replication-independent manner. This leads to deposition of an H3 variant, called H3.3 in vertebrates, in actively transcribing regions. H3.3 has been suggested to have a role in the epigenetic memory of this chromatin state \textsuperscript{114,115}. Additionally, in \textit{S. pombe}, a low nucleosome turnover has been shown to be required for maintenance of repressive chromatin\textsuperscript{116}. In this way, eviction of the nucleosomes carrying repressive histone modifications in actively transcribed regions may constitutively protect them from silencing.

A somewhat counter-intuitive aspect of the body of genes is that they can be enriched, to varying degrees, in marks associated with transcriptional repression, such as H3K9me3 and DNA methylation. Surprisingly, for genes that harbour them, often these marks actually seem to be required for normal processing of transcription, so long as they do not spread to the TSS \textsuperscript{117–120}. Thus, these marks can cover active genes without inhibiting their expression, as long as the TSS is free of their repressive effect.

Another important facet of elongation regulated by chromatin is co-transcriptional RNA processing \textsuperscript{121,122}. H3K36me3 is highly associated with exons and is involved in regulation of splicing and exon inclusion. Additionally, it has been shown that repressive marks on the gene body also slow down elongation speed, which can affect exon choice in genes with alternative splicing.

\subsection*{1.3.6. Heterochromatin states}

Heterochromatin is the term used for those genomic regions that are transcriptionally repressed. It was classically described as denser regions within the nuclear material, as observed by microscopy, corresponding in majority to repetitive DNA \textsuperscript{95,123}. Overall, heterochromatin is characterized by having a low or negligible level of transcriptional
activity, which is actively silenced. This is achieved by active recruitment of HDACs, leading to reduction of nucleosome accessibility, and other silencing mechanisms.

Several types of heterochromatin can be differentiated. Some lower eukaryotes, such as *S. cerevisiae*, possess only heterochromatin based on HDAC recruitment mechanisms. *S. pombe*, on the other hand, possesses heterochromatin much more similar to that of metazoa. Animals have two types of heterochromatin: the (classically described) constitutive heterochromatin and polycomb facultative heterochromatin. For the remainder of this work, and for ease of distinction, I will refer to them as heterochromatin and polycomb chromatin, respectively.

### 1.3.6.1. Constitutive (classic) heterochromatin

Regions of constitutive heterochromatin are gene-poor and often associated with repetitive elements, including some rDNA arrays, telomeres and centromeres in most eukaryotes. Constitutive heterochromatin, as the title implies, is found at genomic loci which are under transcriptional repression throughout the cell cycle (and in metazoans, in the vast majority of cell lineages). Heterochromatin regions, including those from different chromosomes, tend to cluster with one another, and in some organisms this clustering is evident as dense "chromocenters" of DNA. Heterochromatin is broadly characterized by the chromatin marks H3K9me3, H4K20me3, DNA methylation and hypoacetylation of overall histone residues.

The foundation chromatin mark for heterochromatin is H3K9me3, which is catalyzed and maintained mostly by the Suv39h1/h2 enzymes in mammals, leading to recruitment of a cascade of factors that create the heterochromatin environment. H3K9me3 is recognized by HP1 (heterochromatin protein 1), which recruits to the locus DNA methyltransferases and HDACs, leading to an overall "closed" chromatin conformation with reduced transcriptional accessibility. Additionally, HP1 recruits the Suv420h1/h2
enzymes, which catalyze establishment of H4K20me3, a mark reported to be required for proper heterochromatin compaction \(^{128}\).

Heterochromatin also possesses the ability to spread across chromatin domains, in a phenomenon termed position-effect variegation. It was first discovered when genes close to heterochromatin regions could become inactivated, in a random fashion for each cell throughout an organism's tissue, leading to a mosaic, variegated phenotype \(^{129,130}\). Spreading seems to be in part promoted via direct binding of Suv39h1/h2 to its own catalytic product, H3K9me3, thus eventually nucleating the same mark in neighbouring nucleosomes. Coupled with the ability of Suv39h1/h2 to directly recruit HP1 and DNA methyltransferases \(^{131,132}\), progressive expansion of H3K9me3 domains allows repressive activity to cover adjacent chromatin and affect transcription of nearby loci very efficiently.

Heterochromatin also appears to use a *de novo* targeting mechanism to chromatin, called RNA interference (RNAi), which also complements its maintenance \(^{133,134}\). It relies on the existence of double-stranded RNAs, formed either by annealing of hairpin structures in an RNA sequence, but also from transcripts of repetitive regions or convergent transcription \(^{135}\), which allows two complementary RNA molecules to anneal. In heterochromatin-silenced regions, low-level transcription of underlying DNA (often repetitive in nature) gives rise to a sufficient amount of double-stranded RNA to trigger the process. Double-stranded RNA is detected by the RNAi machinery, processed and then used as a guide for sequence-targeted silencing, by recruiting HP1 and possibly other heterochromatin factors. This pathway has been extensively characterized in *D. melanogaster* and *S. pombe*, but it is still not clear if in vertebrates the process occurs in exactly the same fashion \(^{136}\).

The heterochromatin environment also seems to prevent recombination \(^{137–139}\). Unequal exchange between repetitive DNA regions during recombination can lead to expansion or excision of chromosomal domains. Heterochromatin regions are refractory to homologous recombination between chromosomes, apparently by recruiting the SMC5-6 complex, which promotes sister chromatid exchange instead \(^{140–144}\). This occurs whether
recombination is triggered by DNA damage in order to promote repair \(^{145}\), or during programmed DNA breaks in meiosis \(^{146}\). Many repetitive mobile genetic elements (or relics thereof) are scattered throughout chromosomes, and could give rise to such deleterious rearrangements. But more importantly, this process must be prevented from occurring on centromeres, telomeres and rDNA arrays, which are all repetitive in nature\(^{147,148}\), and are essential for cell viability: loss of large tracts by recombination-derived excision could have consequences such as birth defects or cancer.

1.3.6.2. Polycomb/facultative heterochromatin

The polycomb repressive pathway is thought to be responsible for maintaining the identity of cell lineages, thus controlling differentiation during development in metazoans \(^{149-151}\). The genomic regions it covers are rich in developmentally regulated or cell-type specific genes. Their facultative moniker comes from the fact that different lineages will have different subsets of genes inactivated by polycomb, and inactivation of different genes is performed in a step-wise manner to orchestrate metazoan development. Polycomb repressed regions are broadly characterized by H3K27me3, H2AK119 ubiquitination, DNA methylation and histone hypoacetylation. Polycomb chromatin occupies a different niche to that of heterochromatin, and the two signatures do not overlap throughout the genome.

Polycomb silencing occurs in a step-wise manner, similar to constitutive heterochromatin. First, the PRC2 complex catalyzes the establishment of the polycomb foundation mark, H3K27me3 \(^{152}\). PRC2 can also recruit histone deacetylases \(^{153-155}\) and DNA methyltransferases \(^{156}\), although the latter is still contentious. The H3K27me3 mark then serves as a docking site for the PRC1 complex, which establishes H2AK119 ubiquitination and chromatin compaction \(^{157-161}\), which are thought to block RNA polymerase elongation and prevent access to chromatin, respectively. Recent findings in mouse and human cells also point to PRC1 having a role in global steady-state transcriptional control, by binding to
promoters throughout the genome, in a H3K27me3-independent manner. This has been suggested to have a function in fine-tuning gene expression by monitoring promoters of regulated genes, in order to enforce their inactivation in the absence of the triggering stimulus.

### 1.3.6.3. Lamin associated domains

Apart from heterochromatin and polycomb chromatin, another less well-characterized state has been reported: the LADs (lamin-associated domains), located at the nuclear periphery and are generally transcriptionally silenced. It has been shown in some species that dynamic relocation to the nuclear periphery can be a means of facultative silencing during development or regulated gene activation. In some organisms, links between it and heterochromatin- and polycomb-like signatures were also made, but the mechanisms are still unclear.

### 1.3.7. Chromatin modulation during genomic maintenance processes

Other important chromosomal processes, such as DNA replication and repair, have to invariably contend with chromatin itself. As such, these processes, just like transcription, recruit chromatin remodellers and HATs that "open" chromatin, in order to access DNA.

Replication involves incorporation of new histones but also redistribution of existing nucleosomes between the newly-formed sister chromatids, with the FACT complex playing a significant role. The latter allows chromatin marks to be segregated and their respective states maintained.

The DNA double-strand break repair machinery transduces detection of a break into a chromatin signal, by phosphorylation of H2A (or in the case of mammals, the
specific variant H2A.X) \(^{185}\). This mark now provides a docking site for additional repair machinery, which includes both HATs and HDACs (which is believed to help modulate "openness" of the locus) \(^{168,186}\) but also the polycomb PRC1 repressor complex, which helps shut down local transcription \(^{187,188}\). Nucleosome turnover aided by H2A.Z incorporation is proposed to help the locus erase these chromatin marks and restore itself to its former state \(^{189,190}\).

1.4. The Centromere

1.4.1. CENP-A

The centromeric domain is defined on chromosomes via incorporation of an H3 variant: CENP-A \(^{191}\). CENP-A replaces some of the canonical H3 in nucleosomes at centromeres, in an interspersed fashion \(^{192-196}\). In *Drosophila melanogaster*, its introduction into chromatin seems to be sufficient to confer epigenetic memory of centromere specification and recruit other centromeric proteins \(^{197-199}\). However, in human cells this did not seem to be the case \(^{198}\), indicating other factors may be required.

CENP-A is highly diverged from canonical H3 and also its sequence is poorly conserved throughout eukaryotes, in stark contrast to H3. This is primarily true of its N-terminal tail, apparently rendering CENP-A unable to host many of the important H3 modifications that grant many chromatin states their different properties. Some CENP-A tail post-translational modifications have been implied to significantly alter nucleosome positioning and compaction \(^{200}\). The C-terminal tail of CENP-A is in turn required to recruit other centromeric proteins \(^{198,201,202}\).

A unique domain within the protein, the CATD (CENP-A targeting domain), grants CENP-A its specific targeting, via the chaperone HJURP, which catalyzes its incorporation into chromatin \(^{203-206}\). Furthermore, CENP-A-containing nucleosomes have been shown to have a distinct shape from canonical H3 nucleosomes, being flatter and more compact \(^{204,207-}\).
As it stands, all the evidence suggests that introduction of CENP-A into a genomic locus drastically changes the local physical behaviour and signature state of chromatin.

**Fig. 1.3: The CCAN and kinetochore**

**1.4.2. The CCAN: the foundation of kinetochores**

The structural foundation of the kinetochore is formed by a network of proteins termed the CCAN (constitutive centromere associated network) (fig. 1.3). The majority of these proteins bind the centromere locus throughout the cell cycle, and many of them recruit and/or regulate the outer kinetochore proteins that assemble only upon the onset of mitosis.

The CCAN is composed of a series of CENP proteins in vertebrates. CENP-C and the CENP-T/W/S/X complex are the innermost structural components, and provide
structural attachment between CENP-A chromatin and outer kinetochore proteins \(^{218}\). Other sub-complexes, such as CENP-L/M/N \(^{212,214}\), -H/I/K \(^{219-221}\) and -P/O/R/Q/U \(^{212,213,222,223}\) seem to promote CCAN stability and interconnectivity, by reinforcing each other's localization \(^{213,224-227}\) and additionally contributing to regulation of kinetochore function \(^{212,223-225,228-231}\). Briefly, CENP-L/M/N directly recognizes CENP-A nucleosomes \(^{232,233}\) and together with CENP-H/I/K they mutually stabilize their centromeric localization \(^{213,224,226}\). Both complexes also contribute to outer kinetochore (Ndc80) recruitment \(^{233,234}\). Importantly, CENP-H/I/K has been suggested to be required for maintenance of CENP-C and CENP-A assembly at centromeres, thus closing the CCAN epigenetic feedback loop and ensuring centromere memory \(^{220,221}\). Finally, CENP-P/O/R/Q/U is recruited to the CCAN through CENP-H/I/K \(^{212,213,222,223}\) and is believed it may help regulate kinetochore activity in microtubule attachment \(^{223,225,228-230}\). Curiously, in \(S.\) \(cerevisiae\), this complex has the role of recruiting additional cohesin to the pericentromeric region \(^{235-237}\), but evidence for such a role in vertebrates is still lacking.

For the purpose of this work, I will focus on the basal components of the CCAN: CENP-C, CENP-T/W/S/X and CENP-B, which contact chromatin more closely.

### 1.4.2.1. The cornerstones: CENP-C and CENP-T/W/S/X

These two protein complexes bind centromeres constitutively at all cell cycle stages, and link centromeric chromatin to kinetochore proteins and therefore microtubules. Both are thought to recruit the basal kinetochore protein Mis12 through distinct interaction domains, anchoring it to the centromeric region.

CENP-C can bind the CENP-A nucleosome directly \(^{238-241}\), as well as the adjacent DNA \(^{215,242-245}\). CENP-C is required for recruitment of most other centromere proteins \(^{246}\), including the CENP-T/W/S/X complex \(^{239}\). Additionally, CENP-C seems to act as a crosslinker within the centromere, holding together stretches of CENP-A-containing
nucleosomes and effectively stabilizing the locus as a single structure during mitosis\(^{247}\). As such, CENP-C is one of the most crucial pieces of the CCAN, which recognizes CENP-A and recruits all other sub-complexes.

The CENP/T/W/S/X complex\(^{214,216,248}\) is composed of proteins which contain histone folds in their structure, and curiously interacts stably with H3-containing nucleosomes at the centromere instead of CENP-A nucleosomes\(^{215,247}\). It has been proposed that this complex wraps DNA around itself in an opposite way to that of canonical nucleosome wrapping, inducing positive supercoiling\(^{217,249}\). Additionally, CENP-S and -X\(^{216}\) have been shown have roles in DNA repair and homologous recombination, outside of centromeric loci\(^{250-253}\). So far, it is still unclear if these functions play a role in centromere assembly/structure.

CENP-C and the CENP-T/W/S/X complex seem to contribute to kinetochore anchoring in distinct pathways. CENP-T anchors the outer kinetochore through recruitment of microtubule-binding proteins Ndc80 and KNL1\(^{254,255}\) and this function is absolutely critical for kinetochore function\(^{215,256}\). While CENP-C is required to recruit the Mis12 complex, which together with Ndc80 and KNL1 forms the KMN network\(^{246,257-260}\), its depletion from centromeres does not cause such an immediate dramatic effect like that of CENP-T. However, CENP-C is the means by which the remainder of the CCAN and indeed CENP-A seems to be recruited, possibly contributing to maintenance of epigenetic memory of the centromere\(^{246,261,262}\). In this way, current evidence points to CENP-C having a more prominent role in maintaining proper recruitment of the remaining centromere proteins rather than an direct structural role in kinetochore anchoring.

**1.4.2.2. CENP-B**

CENP-B is a centromeric protein, present in mammals, whose exact function is still unclear\(^{191,263}\). Its DNA-binding domain recognizes a motif, the 17bp CENP-B box, present
only in kinetochore-proximal centromeric satellite repeats, typically interspersed throughout the repetitive array (which is longer than its CENP-A containing region) 264.

CENP-B binding to DNA is required for de novo formation of centromeres, on DNA templates in vivo 265,266. However, knockdown of this protein is tolerated, and mice that genetically do not express CENP-B are also healthy and viable for several generations 267. Additionally, the centromeric satellite repeats in the human Y chromosome do not possess CENP-B boxes 268. As it seemed, CENP-B was apparently dispensable for centromere function.

Recently, it was suggested that CENP-B interacts with the N-terminal tail of CENP-A, which is required to maintain a portion of total CENP-B at the centromere. This fraction of total CENP-B seemed to be required for long-term centromere maintenance when the N-terminal tail of CENP-A (which promotes CENP-C accumulation) is deleted 201. This suggests that CENP-B may contribute to a second pathway ensuring CCAN and centrochromatin stability.

This protein is evolutionarily related to transposases 269, and its three homologs in S. pombe, which are non-centromeric, have redundant functions in controlling DNA replication through repetitive regions and also promoting chromatin silencing 270,271. Interestingly, this latter function might also be present in human CENP-B 272. CENP-B binding to the CENP-B box has also been shown to induce phasing of nucleosomes 273.

Finally, it bears mentioning that the structure of the centromere appears to be quite flexible and dynamic. Knockdown or depletion of CENP-A expression only has visible effects after several cell divisions, as residual levels of previously incorporated CENP-A at centromeres are sufficient to confer efficient kinetochore activity until a critical threshold is reached (3~5 divisions in cancerous human and chicken cells) 40,201,274,275. This is also true for CENP-C to an extent 276, and most of the CCAN proteins are highly dynamic during most of
the cell cycle. The result is a structure in a dynamic steady-state as opposed to a rigid stable composite.

1.4.3. Cell cycle regulation of centromeres

The centromere structure is not static, and undergoes considerable rearrangements throughout the cell cycle. Being reliant on CENP-A nucleosomes for its specification on chromatin implies that the "centromere state" becomes diluted between the two chromatids after replication. As such, consistent replenishment of centromere components, in each cell cycle, is crucial their maintenance.

Centromere DNA in most animal cells seems to replicate during mid-late S phase, but unlike canonical histones, new CENP-A is not incorporated into centromeric loci at this time. CENP-A assembly takes place just after mitotic exit in early G1, with the drop in Cdk activity at anaphase being a catalyst for its initiation. Otherwise, CENP-A is distributed between sister chromatids during S phase, and remains in this "diluted" amount until the next assembly stage.

CENP-A assembly, in vertebrates and S. pombe, starts with targeting of the Mis18 complex to centromeres. Mis18 recruitment has been found to be promoted via interaction with CENP-C. Mis18 does not seem to interact directly with CENP-A, but is essential for CENP-A localization. It is believed that Mis18 "primes" centromeres for subsequent CENP-A assembly, but the process remains unclear. It is believed that it affects the local chromatin state, by somehow influencing acetylation levels at the centromere.

Once "priming" has taken place, CENP-A/H4 tetramers are brought and assembled into centromeres by association with HJURP. HJURP is a conserved Scm3 domain-containing protein, which can assemble CENP-A into nucleosomes. Some reports showed that HJURP is directly recruited to centromeres by Mis18, via its Mis18β subunit.
The CENP-A assembly process starts in late telophase and continues throughout most of G1, finishing with a maturation process, involving MgcRacGAP, which seems to lock new CENP-A in place. Curiously, the players in this process do not seem to be conserved in several lineages, such as *D. melanogaster*, where CENP-A is assembled via the protein CAL-1.

### 1.4.4. Centromeres and disease

Perturbation of centromere/kinetochore function is one way through which healthy cells can fall down the path of tumourigenesis. One of the hallmarks of many solid tumours is high levels of CIN (chromosomal instability), which is an elevated frequency of whole-chromosome mis-segregation. At least three centromere-related mechanisms can lead to this outcome: faulty sister chromatid cohesion, deregulation of the mitotic checkpoint and overly-stable kinetochore-microtubule attachments. The resulting lagging chromosomes can lead to aneuploidy or even chromosomal damage during cytokinesis and micronucleus formation, causing abnormal gene expression/proteotoxic stress and genome rearrangements, respectively, both of which are strong drivers for subsequent acquisition of cancerous properties.

As such, understanding the mechanisms of chromosome segregation is crucial to our understanding of human disease: how it can occur, how it evolves and possibly, how it can be prevented or cured.
1.5. Centrochromatin

1.5.1. The epigenetic nature of centromeres

For the majority of eukaryotes, centromere location on a chromosome does not seem to be strictly defined by any particular targeting motif in DNA. Although association with some sort of characteristic DNA sequence is common in most eukaryotic centromeres (α-satellite repeats, in the case of humans), they can apparently also form in unrelated sequences. The latter was first observed in human clinical isolates that showed striking karyotypic differences in centromere position, compared to the overall population. These so-called neocentromeres are often formed in regions of DNA with little to no sequence similarity to centromeric DNA, so they are not the consequence of chromosomal translocations of whole centromere loci. They can arise in the wake of genome rearrangements that have led to the loss of the original centromeric region: indeed, most clinically identified neocentromeres in humans are of this sort. Alternatively they can occur spontaneously, and the original centromeric position can still be found in its regular location on the chromosome, but has lost centromeric activity. This is also observed as a consequence of translocations that lead to chromosome fusion. Stable fused chromosomes generally bear only one active centromere (the other having become inactivated).

Neocentromeres provided evidence that centromeres are epigenetic. Neocentromere locations are stable and heritable, and centromere relocation events have been suggested to have occurred throughout the evolution of several species, after becoming fixed in a population. Neocentromeres bear all marks of active centromeres, such as CENP-A, recruitment of kinetochore proteins and formation of a primary constriction. Interestingly, neocentromeres do not co-localize with CENP-B signals, as they are not located on α-satellite repeats (which contain CENP-B boxes), however CENP-B remains bound to inactive centromere loci.
Current models of centromere specification point to CENP-A being both necessary and sufficient for triggering identity and memory of this locus. Experiments conducted in *D. melanogaster* have shown that incorporation of CENP-A into non-centromeric loci, via its overexpression, can induce ectopic kinetochore formation \(^{322}\). Further confirming this, recruitment of tagged CENP-A to a specific chromosomal locus also assembled a functional kinetochore and recruited additional CENP-A, an arrangement that persisted even in the absence of constitutive CENP-A recruitment \(^ {199}\). *In vitro*-reconstituted CENP-A chromatin, when introduced into *X. laevis* egg extracts, also recruits some aspects of functional kinetochores \(^ {293}\). Although in human cells CENP-A overexpression also causes its mislocalization across non-centromeric loci, it has so far proved unable to effectively nucleate functional ectopic kinetochores \(^ {198,254}\). However, artificial tethering of HJURP or CENP-C (a constitutive centromere protein) to a non-centromeric locus, in human and chicken chromosomes, was successful in nucleating functional kinetochores, which persisted in the absence of tethering \(^ {205,246}\).

The accumulation of evidence, from multiple organisms, all points to CENP-A being an epigenetic marker for centromere specification that, once nucleated beyond a certain threshold (be that in quantity or post-translational modification), initiates a feedback loop which establishes further CENP-A recruitment, thus maintaining centromere identity \(^ {323-325}\).

**1.5.2. Conservation of centromere organization in eukaryotes**

Although the genomic location of centromeres may ultimately be promoted in an epigenetic fashion by CENP-A in most eukaryotes, it does often bear distinct sequence traits that could potentially influence centrochromatin behaviour (fig. 1.4). Extended regions of repetitive DNA and a higher AT base content are the DNA-related factors that seem to be most conserved in their association with centromeres, instead of sequence itself, throughout most eukaryotes. Yet, while centromeric sequences are very poorly conserved between
species, they are usually quite similar between all the chromosomes of a given species\textsuperscript{326–330}, a striking observation given that each centromere should presumably be maintained independently from each other in \textit{cis}, via epigenetic means, and therefore evolve independently.

With the exception of species with holocentric chromosomes, most eukaryotes have a single centromeric region which defines the primary constriction of mitotic chromosomes. Within these more "conventional centromeres" we can distinguish two types: "point" centromeres which are defined by a very small centromeric region (1-2 nucleosomes long), and "regional" centromeres which encompass larger, Kb- or Mb-long, genomic regions.
Fig. 1.4: Centromere structure organization

Schematic interpretation of sequence and chromatin domains in centromeres of model eukaryotes. Diagrams show domain organization and size, with description of each feature. Note that while some sequences are centromere-associated in some organisms (yellow boxes/arrows) not all are CENP-A associated (green boxes/arrows). Mapped chromatin states are shown below, as well as detected transcripts.
1.5.2.1. Yeast point centromeres

The centromere of *S. cerevisiae* is the best-studied model for point centromeres (fig. 1.4), which also occur in a few other fungal species, and are believed to have evolved from epigenetically-specified regional centromeres. They rely on a subset of point-centromere-specific proteins to promote sequence-dependent binding to a ~120bp DNA recognition sequence which comprises the minimal centromere nucleation platform. This effectively makes these centromeres non-epigenetic. Despite that, however, local chromatin dynamics share some similarities with regional centromeres: they reside within hypoacetylated chromatin and require low-level transcription to be maintained. Yet, strong transcriptional activity can dismantle these centromeres, something which is also observed in human cells. As such, a sufficient level of transcription (but not excessive) seems to play a role in centromere function, regardless of whether a centromere is epigenetically defined or not (see below).

1.5.2.2. Yeast regional centromeres

The regional centromeres of *S. pombe*, those whose organization is perhaps the best studied, have a very regular structure, clearly distinguishing a core centromere domain from an outer pericentromeric domain, composed of arrays of different DNA repeats (fig. 1.4). The core is where CENP-A accumulates, which also shows a low level of transcription. The outer pericentromeric region, which surrounds the core, is composed of repetitive DNA, is heterochromatic and enriched in cohesin. Between the core and the pericentromere repeats stand inverted repeats which contain tRNA genes, forming an effective transcription-dependent barrier against spreading of pericentromeric heterochromatin onto the core. This domain arrangement, of a non(or less)-heterochromatic core surrounded by a heterochromatic pericentromere, where at least one of those components is composed of
tandemly repetitive DNA, is conserved in other yeasts such as *C. albicans* and *N. crassa*. However, in these organisms those domains are not necessarily determined by their sequence (repeat type) like in *S. pombe*.

Functional dissection of the importance of these regions, through *de novo* centromere formation studies, revealed surprising centromere chromatin determinants. Experimental neocentromere generation showed a clear trend of neocentromeres forming on gene-poor regions and associated with repetitive and/or heterochromatic regions, in *C. albicans* and *S. pombe*. The association of centromeric function and heterochromatin in *S. pombe* has long been established, and in fact, mutants for heterochromatin components fail to nucleate neocentromeres. *De novo* centromere generation on ectopic minichromosomes in *S. pombe* revealed that both the central core and a portion of the outer repeats are critical for its establishment, and a robust outer repeat region seemed to be required for meiotic stability. More surprisingly, sequence determinants can be bypassed completely through ectopic nucleation of heterochromatin on minichromosomes containing no centromere sequences. This approach was able to generate centromeres in loci adjacent to the nucleation site, which are maintained even in subsequent withdrawal of heterochromatin nucleation.

1.5.2.3. Plant centromeres

Centromeres of studied plant models assemble on Mb-long tandem arrays of AT-rich centromere satellite repeats, interspersed with retrotransposons (fig. 1.4). These satellites share a common sequence ancestry, but show chromosome-specific differences, and are found exclusively on centromeric regions. Yet, barley and maize neocentromeres can form outside of centromere repeats or indeed any repeat regions, indicating that they are also not an absolute requirement in these organisms.
1.5.2.4. Animal centromeres

The animal organisms whose centromeres have been characterized present large tandemly repeated arrays of satellite DNA in their centromeric regions, but many show no sequence distinction between a core and an outer pericentromere (fig. 1.4).

In *D. melanogaster*, centromeres form on heterochromatic AT-rich repeats, which are interspersed with retrotransposons. These can be highly heterogeneous between chromosomes, and the repeats and transposable elements they harbour are not always exclusively centromeric. Neocentromere "hotspots" in *D. melanogaster* seem to be primarily influenced by heterochromatin boundary locations.

In *X. laevis* and mammals, centromeric sequences consist of a single type of AT-rich satellite repeat, containing a CENP-B box motif, organized in tandem higher-order-repeated arrays and present in the majority of chromosomes. Primate, especially human, satellites have been extensively characterized, and conform to chromosome-specific but related satellite families, which organize themselves in repeat arrays of 0.3-5Mb, containing a higher-order repetitive structure of slightly different monomers. The pericentromere can be distinguished somewhat from the centromere core, as it shows increased sequence degeneration (including nonfunctional CENP-B boxes) and lacks higher-order organization of the array. Mouse satellite organization is distinct from this: it is divided between minor satellites, which harbour CENP-A and contain CENP-B boxes, and major satellites (of different sequence composition) which are highly heterochromatic and flank the centromeric minor satellite array. This arrangement provides, like in *S. pombe*, an apparently sequence-determined pericentromere.
1.5.2.5. Genomic determinants in human neocentromeres

Several non-centromeric satellite regions in the human genome have the potential to nucleate de novo centromeres \(^{384}\). However, most characterized human neocentromeres, although arising in gene-poor regions, share no sequence similarity between them, and have no association with satellites or higher AT-content, with only some of them being located in proximity to repetitive elements \(^{318,385-389}\). In at least one human neocentromere, it has been shown that, although full kinetochore function had been established, the distribution of centromeric proteins throughout the primary constriction was altered, resulting in unfocused CPC targeting to the inner centromere and consequently an inefficient error-correction ability for that kinetochore \(^{390}\). This neocentromere and two others were also shown to suffer premature loss of cohesion during prolonged mitotic arrest \(^{390,391}\). Such findings, albeit few so far, suggest that full and efficient centromere function may not only be dependent on CENP-A 'seeding' and kinetochore nucleation, but could possibly involve other genomic aspects (dependent on sequence and/or chromatin state). These may somehow play a role in providing a suitable 'cradle' for centromere maintenance and function. Evolutionary evidence from several organisms \(^{392-396}\) hints at cycles of centromere relocation, each relocation followed by subsequent acquisition and expansion of repetitive sequences in their surroundings (up to Mb-length tandem arrays). This encourages the idea that centromeres do have genomic preferences and may even promote these advantageous genomic rearrangements. The mechanisms that could allow such events to occur are, however, still conjecture \(^{394,397}\).

In light of the evidence, it is paramount to understand the conditions that make an efficient centromere, both in their naturally evolved settings (repetitive arrays or otherwise)
or in recently established regions (such as neocentromeres). And while it is likely that species-specific differences, of how chromatin and the genome as a whole is regulated, play a role in determining some particulars of centromere behaviour (such as point centromeres in yeasts \textsuperscript{331} or holo-centromeres \textsuperscript{398-400}), several of them are likely (and some have been proven) to be sufficiently conserved for parallels to be drawn across species.

### 1.5.3. Implication of chromatin state in centromere function

The study of centromere organization first pioneered in \textit{S. pombe} suggested that specific organization of chromatin states at and around centromeres could bear significance to their function \textsuperscript{342}. The characteristic organization of a heterochromatic pericentromere surrounding a non-heterochromatic CENP-A-containing core region is somewhat of a conserved feature. Observations in chicken and \textit{D. melanogaster} also suggested that there is spatial compartmentalization between the core and pericentromere, and the pathways that assemble each seemed to be genetically independent \textsuperscript{401,402}.

#### 1.5.3.1. Nucleosome distribution

Study of the chromatin organization of the centromere core showed that stretches of CENP-A nucleosomes were interspersed with those containing canonical H3 \textsuperscript{195}. This has led to the formulation of several structural models, in which alternating CENP-A and H3 nucleosomes in the centromere core, together with CCAN proteins and the adjacent pericentromere, may contribute to a higher-order 3D structure of the primary constriction \textsuperscript{195,247,403}. 
1.5.3.2. Centromere core vs pericentromere

Centromeric H3 nucleosomes in human and *D. melanogaster* cells were shown to contain chromatin marks characteristic of euchromatic regions, such as H3K4me2 and H3K36me2, but with otherwise low levels of histone acetylation \(^{404,405}\). This active chromatin state has since been shown to be required to maintain CENP-A assembly and to allow *de novo* centromere nucleation in ectopic centromeric DNA templates \(^{294,405,406}\). This semi-euchromatic signature contrasts with the adjacent pericentromere region, which is highly heterochromatic. Indeed, chromatin manipulations of centromeres in *S. pombe* and human cells have shown that heterochromatin inactivates centromeres \(^{294,345,346,407,408}\). Thus, the centromere CENP-A containing core seems to clearly have chromatin properties distinct from its surrounding regions.

1.5.3.3. Centromere-associated chromatin factors

Reinforcing the idea that euchromatin marks are important to centrochromatin is the finding of several factors required for CENP-A assembly that are themselves associated with transcription. Chromatin remodellers CHD1 and FACT, for instance, are required for assembly of new CENP-A in chicken cells \(^{409}\), albeit not in *D. melanogaster* \(^{410}\). The FACT-associated RSF complex was shown to interact with centromeres in G1 and is also required for maintenance of centromere function in human cells \(^{411}\). The nucleosome-assembling NPM1 phosphoprotein also directly interacts with pre-nucleosomal CENP-A and is required for CENP-W stability in human cells, and in *D. melanogaster* may be involved in centromere clustering and anchoring to nucleoli, which stabilize pericentric heterochromatin \(^{214,412}\). As such, although differences can be found between the CENP-A assembly process of flies and other animals, several similarities still seem to be present.
More recently, a new set of human and chicken centromere-associated proteins was discovered\textsuperscript{413}, which were found to be involved in transcriptional activation: Eaf6/CENP-28\textsuperscript{414}, PHF2/CENP-35\textsuperscript{415,416} and MSL1v1/CENP-36\textsuperscript{417,418}. Perhaps the most poignant piece of evidence supporting those mentioned above was the detection of the active, elongating state of RNA Pol II, at the centromeres of human mitotic chromosomes\textsuperscript{419}. In conclusion, the evidence suggests that particular chromatin remodelling processes take place at centromeres and, at least in human and chicken, these seem to be related to transcription.

1.5.3.4. Centromeric transcripts

Confirming the finding that centrochromatin seemed to bear marks of active transcription and recruit transcription proteins, centromeric transcripts have been shown to be present in cells of several eukaryotes, such as \textit{S. pombe}, \textit{S. cerevisiae}, rice or human cells\textsuperscript{341,419–423}. While transcription within the adjacent pericentromere may contribute to overall centromeric fidelity by allowing heterochromatin maintenance through the RNAi machinery\textsuperscript{402,424,425}, transcripts from within the centromere core have been suggested to perform distinct functions. Centromeric transcripts have been suggested to interact with CENP-C and Aurora B, and contribute to their functions\textsuperscript{426,427}. At this point, it is still unknown if the transcripts themselves have function in CENP-A assembly or maintenance. Due to the fact that centromeres can form in regions devoid of canonical centromere DNA, it seems unlikely that the sequence of the transcript is essential. However, transcripts \textit{in situ} could still have an impact if the centromere shows dependency on a particular co-transcriptional regulation process, regardless of their sequence.
1.5.3.5. Chromatin state and epigenetic memory of centromere function

It has been shown that a change in chromatin state plays a role in activation of centromeric function in otherwise inactive centromere sequences. Centromere inactivation can occur to prevent dicentric chromosomes (likely the result of a recombination event causing chromosome fusion) from breaking apart during mitosis by uncoordinated spindle attachments between the two kinetochores. It is unknown whether the process is purely stochastic or regulated. It involves the complete loss of centromere memory proteins (CENP-A and CCAN proteins) and loss of sister chromatid constriction in mitosis. It has been shown that inactive centromeres become more highly heterochromatinized, and that antagonizing such chromatin silencing on those regions can actually reactivate them.

Taken together, the evidence suggests the following: that centromere function, and possibly the process of CENP-A assembly itself, is somehow promoted by either open chromatin, active transcription, the accumulation of transcripts in cis or by some consequence of the transcriptional process (such as histone remodelling). Or perhaps even a combination of the above.

1.5.4. Artificial chromosomes and de novo nucleation of centromeres

Some of the most significant and clear insights into the determinants of centromere identity have come from the development of artificial chromosomes. This reductionist approach consists of attempting to generate a chromosomal locus de novo, such as a centromere or telomere, through its nucleation on input DNA sequences associated with each locus. This approach proved instrumental in understanding the genetic determinants of yeast centromeres, telomeres and replication origins, and allowed generation of designer YACs (yeast artificial chromosomes. Further application of these principles in yeast and
human cells has led to some of the most significant advances in the current understanding of centromere chromatin genomic determinants.

HACs (human artificial chromosomes) are generated through introduction of ectopic \(\alpha\)-satellite DNA arrays into human cells. The first attempts utilized co-transfections of arrays of \(\alpha\)-satellite repeats together with telomeric sequences and random genomic DNA. Further refinement, taking a cue from YACs, was to clone the above in a single vector before introduction of the linear construct into human cells, thus allowing further control of the composition of the final chromosome. Later, it was found that telomere sequences were dispensable if the transfected construct was circular in nature. This opened the way for simplifying the array generation procedure, as it could now be manipulated inside bacteria or yeast cells, and augmented through rolling circle amplification. Before, stretches of cloned \(\alpha\)-satellite had to be used, limiting the potential to alter the sequence of its repeats in order to probe for specific key centromeric determinants. Synthetic repeats can now be designed and amplified to generate a tandem array long enough to promote de novo centromere formation.

The most recent advances in HAC generation have allowed the creation of HACs to which specific proteins can be targeted (fig. 1.5). Such HACs contain \(tetO\) sequences within their repeats, which allow for TetR-fusion proteins to selectively bind to the HAC and recruit enzymatic functions in \textit{cis}. This technique has allowed for specific modulation of HAC chromatin to great success. In this way, by manipulating only a single chromosome within a cell, it has become possible to probe centromere function without affecting cell viability and also visualizing long-term effects on centromere function, as the HAC is a non-essential chromosome.
The exact events that occur during HAC generation are still very much unknown. The resulting HACs are often amplified in size \(^{438,439}\), possibly due to homologous
recombination between the repeats of the input naked DNA. An additional by-product of the HAC generation procedure is integration of the input array into another chromosome, which occurs frequently during the HAC formation process. This generally results in inactivation of potential centromere function through heterochromatin accumulation. Isolated clones with such integrated arrays have since been used as negative controls for centromere function (since they share the same DNA sequence with HACs but carry no centromeric activity). Furthermore, testing conditions which could re-activate centromeric activity in integrated arrays has advanced significantly our understanding of the chromatin determinants that allow epigenetic memory of this locus.

1.6. Aims of this work

In the remainder of this document, I will detail my efforts to further our understanding of the chromatin determinants of centromere function. The systems I used to achieve this relied heavily on synthetic biology tools: mapping centromere and neocentromere chromatin on genomically engineered cells, and conditional targeting of chromatin modifiers to HACs. The studies were conducted in vertebrate cells: in chicken DT40 cells and human HeLa cells. The latter system (HAC-targeting in human cells) has been extensively used by the Masumoto, Larionov and Earnshaw labs, and has been expansively characterized.

Within the overall goal of understanding the chromatin determinants of centromere function, I have focused on the interplay between repressive chromatin marks and centrochromatin, as this association is a conserved phenomenon throughout eukaryotes. However there is little evidence of whether or not the repressive pericentromere chromatin directly contributes to centromere function, as opposed to marks of active transcription (which have previously been shown to play a role in centrochromatin maintenance).

My efforts are broken down into three separate projects, as detailed below.
1.6.1. Chromatin mapping of non-repetitive chicken centromeres

The chromatin landscape of vertebrate centromeres is poorly characterized due to its repetitive nature: the study of fine detail of chromatin features, particularly at the centromere core, is usually relegated to microscopy approaches. However, a few species of vertebrates, such as chicken, have some centromeres that are not composed of tandem repeats.

In a collaboration with the Fukagawa lab, I have mapped the coverage of several chromatin marks of interest on these non-repetitive centromeres of chicken. Additionally, by engineering one of these centromeres for conditional excision, several neocentromere clones were subsequently isolated, and the chromatin state pre- and post-neocentromere formation was characterized.

This was the first time native vertebrate centromeres (non-neocentromeres) could be mapped for chromatin marks at high-resolution, and the evolution of a locus into a neocentromere could be studied in vertebrates.

1.6.2. Centromere activity can coexist with polycomb chromatin

Previous manipulations of HAC chromatin have led to the finding that directly removing marks associated with active transcription resulted in loss of CENP-A assembly. Conversely, heterochromatin targeting led to swift inactivation of locus as well. This suggested that an actively transcribing state is important for centromere function. However, it is still unclear how this state interacts with the surrounding heterochromatin and how it remains transcriptionally active in such conditions.

In an effort to further understand the response of centrochromatin to silencing, I nucleated polycomb chromatin on a HAC through the H3K27 methyltransferase EZH2. Polycomb achieves effective transcriptional repression in slightly different ways than
heterochromatin, and is more sensitive to the underlying chromatin state. By analyzing how the change into polycomb state affected centrochromatin, and how it compared to inactivation by heterochromatin, I attempted to tease out which components of the centrochromatin signature are essential.

1.6.3. Heterochromatin stabilizes the core centromere and prevents DNA damage

Centromeres are usually associated with heterochromatin in most eukaryotes. It has been shown in some organisms that the presence of heterochromatin contributes to centromere function, by providing increased cohesion at the primary constriction and by preventing deleterious recombination between repeats, that could otherwise lead to loss of a centromere by excision. However, it is unclear whether the presence of heterochromatin contributes directly to centrochromatin and CCAN stability, or why their association is conserved in evolution. Evidence from *S. pombe* and *D. melanogaster* seems to indicate that is the case, in that heterochromatin can promote ectopic centromere nucleation, but evidence from human cells shows the opposite effect.

In order to clarify whether heterochromatin and centrochromatin function can be dissociated, I have selectively erased HAC heterochromatin, using the H3K9 demethylase JMJD2D. In this way it was possible to remove heterochromatin from a single chromosome, without interfering with endogenous centromeres or de-regulating overall cellular heterochromatin, as opposed to previous studies that relied on RNAi and gene deletions. I subsequently proceeded to analyze how centrochromatin and the CCAN responded to the change, and how it affected mitotic function of the HAC.
2. Material & Methods

2.1. Materials

2.1.1. Buffers, solutions & reagents

All solutions were prepared in double-distilled water. Chemicals and reagents were purchased from Sigma-Aldrich or Merck, excepting SDS solution (Severn Biotech Ltd) and Tween-20 (BioRad).

Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x DNA loading Buffer</td>
<td>40% Sucrose; 2mg/ml Orange G</td>
</tr>
<tr>
<td>LB medium</td>
<td>1% tryptone; 0.5% yeast extract; 10mM NaCl; pH7.4</td>
</tr>
<tr>
<td>PBS</td>
<td>65mM Na2PO4; 8.8mM KH2PO4; 137mM NaCl; 2.7mM KCl; pH7.4</td>
</tr>
<tr>
<td>Protease Inhibitor Cocktail (CLAP)</td>
<td>1µg/ml each chymostatin, leupeptin, antipain, pepstatin</td>
</tr>
<tr>
<td>TAE</td>
<td>40mM Tris-acetate; 1mM EDTA; pH8.0</td>
</tr>
<tr>
<td>TBS</td>
<td>25mM Tris; 137mM NaCl; 3mM KCl; pH7.4</td>
</tr>
<tr>
<td>TE</td>
<td>10mM Tris; 1mM EDTA; pH8.0</td>
</tr>
<tr>
<td>KCM buffer</td>
<td>10mM Tris-HCl pH8.0; 120mM KCl; 20mM NaCl; 0.5mM EDTA; 0.1% Triton X-100</td>
</tr>
<tr>
<td>ChIP Dilution buffer 1</td>
<td>50mM Tris-HCl pH8.0; 2mM EDTA; 0.2% SDS; 134mM NaCl; 0.88% Triton X-100; 0.088% Sodium-Deoxycholate</td>
</tr>
<tr>
<td>ChIP Dilution buffer 2</td>
<td>50mM Tris-HCl pH8.0; 167mM NaCl; 1.1% Triton X-100; 0.11% Sodium-Deoxycholate</td>
</tr>
<tr>
<td>ChIP Elution buffer 1</td>
<td>10mM Tris-HCl pH8.0; 1mM EDTA; 1% SDS</td>
</tr>
<tr>
<td>ChIP Lysis buffer</td>
<td>10mM Tris-HCl pH8.0; 10mM NaCl; 0.5% NP-40</td>
</tr>
<tr>
<td>ChIP Post-Elution Buffer</td>
<td>10mM Tris-HCl pH8.0; 9mM EDTA; 600mM NaCl</td>
</tr>
<tr>
<td>ChIP RIPA-150 buffer</td>
<td>50mM Tris pH8.0, 1mM EDTA pH8.0, SDS 0.1%, Triton X-100 1%, Na-Deoxycholate 0.1%, and 150mM NaCl</td>
</tr>
<tr>
<td>ChIP RIPA-500 buffer</td>
<td>50mM Tris pH8.0, 1mM EDTA pH8.0, SDS 0.1%, Triton X-100 1%, Na-Deoxycholate 0.1%, and 500mM NaCl</td>
</tr>
</tbody>
</table>
2.1.2. Bacterial and eukaryotic cell culture

**Eukaryotic cell culture**

Human cells were grown in DMEM (+L-Glutamine, +pyruvate) (Gibco), supplemented with 10% (v/v) FBS (Invitrogen) and 100 U/ml penicillin G and 100 µg/ml streptomycin sulphate (Invitrogen), otherwise referred as cDMEM. Cells were grown at 37°C in humidified atmosphere containing 5% CO2. Cell harvesting was performed by washing twice in pre-warmed D-PBS (Invitrogen) and using TrypLE Express (Invitrogen) to release adherent cells, before passage into fresh culture vessels, or downstream applications.

All cell lines containing an α-satellite^{TetO} array (1C7 HAC cells of 1F10 integration cells) based on chr. 17 repeats and a consensus α-satellite sequence were maintained in the presence of 4 µg/ml blasticidin S (Invitrogen). Those with arrays based on chr. 21 repeats (HeLa 2-4) were maintained in the presence of 400 µg/ml G418 (Gibco).

Chicken DT40 cells were grown in RPMI 1640 medium supplemented with 10% (v/v) FBS (Invitrogen), 1% (v/v) chicken serum, 100 U/ml penicillin and 100 µg/ml streptomycin, otherwise referred as cRPMI. Cells were grown at 39°C in humidified atmosphere containing 5% CO2.

**Table 2**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Name of cell line</th>
<th>Details</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>HT1080 AB2.2.18.21</td>
<td>HT1080-derived line containing an α-satellite^{TetO} HAC, based on repeats of chr.17 and a consensus α-satellite sequence. HAC Bsr gene grants resistance to Blasticidin S.</td>
<td>Nakano et al., 2008</td>
</tr>
<tr>
<td>Human</td>
<td>HeLa 1C7</td>
<td>Generated by cellular fusion of HeLa and HT1080 AB2.2.18.21.</td>
<td>Cardinale et al., 2009</td>
</tr>
<tr>
<td>Human</td>
<td>1C7-LSD1</td>
<td>Derived from HeLa 1C7, stably expressing TetR-EYFP-LSD1, from plasmid construct TYIP-LSD1.</td>
<td>Bergmann et al., 2010</td>
</tr>
<tr>
<td>Human</td>
<td>1C7-EZH2</td>
<td>Derived from HeLa 1C7, stably expressing TetR-EYFP-EZH2, from plasmid construct TYIP-EZH2.</td>
<td>This work.</td>
</tr>
<tr>
<td>Human</td>
<td>HeLa 1F10</td>
<td>Generated by cellular fusion of HeLa and HT1080 AB2.5.30, containing an integrated, non-centromeric α-satellite^{TetO} array based on repeats of chr.17 and a consensus α-satellite sequence. The array's Bsr gene grants resistance to Blasticidin S.</td>
<td>Generated in house by Stefano Cardinale. HT1080 AB2.5.30 was a kind gift from Hiroshi Masumoto.</td>
</tr>
<tr>
<td>Human</td>
<td>HeLa OHAC 2.4</td>
<td>HeLa-derived line containing an α-satellite^{TetO} HAC, based on repeats of chr.21 and a consensus α-satellite sequence. HAC Kan^R gene grants resistance to G418.</td>
<td>Tachiwana et al., 2013, a kind gift from Hiroshi Masumoto.</td>
</tr>
<tr>
<td>Human</td>
<td>2-4-EYFP-only</td>
<td>Derived from HeLa OHAC 2-4, stably expressing TetR-EYFP, from plasmid construct pJETY3-TetR-EYFP.</td>
<td>Ohzeki et al, 2012 [294], a kind gift from Hiroshi Masumoto</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>Human</td>
<td>2-4-JMJD2D</td>
<td>Derived from HeLa OHAC 2-4, stably expressing TetR-EYFP-JMJD2D, from plasmid construct pJETY3-TetR-EYFP-JMJD2D.</td>
<td>This work.</td>
</tr>
<tr>
<td>Human</td>
<td>2-4-JMJD2D D195A</td>
<td>Derived from HeLa OHAC 2-4, stably expressing TetR-EYFP-JMJD2D D195A, from plasmid construct pJETY3-TetR-EYFP-JMJD2D D195A.</td>
<td>This work.</td>
</tr>
<tr>
<td>Chicken</td>
<td>DT40 #Z3</td>
<td>Generated by insertion of loxP sites and selection markers, via homologous recombination, in the vicinity of the centromere of chromosome Z.</td>
<td>Shang et al, 2010 [445], Shang et al, 2013 [283].</td>
</tr>
<tr>
<td>Chicken</td>
<td>DT40 #BM23</td>
<td>Derived from DT40 #Z3, via excision of centromere of chromosome Z followed by selection for neocentromere formation.</td>
<td>Shang et al, 2010 [445], Shang et al, 2013 [283], described in this work.</td>
</tr>
<tr>
<td>Chicken</td>
<td>DT40 #0514</td>
<td>Derived from DT40 #Z3, via excision of centromere of chromosome Z followed by selection for neocentromere formation.</td>
<td>Shang et al, 2013 [283], described in this work.</td>
</tr>
<tr>
<td>Chicken</td>
<td>DT40 #0514-CENP-A-FLAG</td>
<td>Derived from DT40 #0514, stably expressing CENP-A-FLAG.</td>
<td>Shang et al, 2013 [283], described in this work.</td>
</tr>
</tbody>
</table>

### Bacterial Escherichia coli strains

Chemically competent *E. coli* TOP10 cells (F– mcrA Δ(mrr-hsdRMS-mcrBC) \(\phi 80lacZΔM15 Δ lacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1\) (grown in-house) or *E. coli* DAM* (ara-14 leuB6 hhuA31 lacY1 tssX78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10) TetS endA1 rpsL136 (StrR) dam13::Tn9 (CamR) xylA-5 mtl-l thi-l mcrB1 hsdR2) purchased from NEB were routinely grown at 37°C on LB-Agar plates or in liquid LB shaking at 180rpm.
2.1.3. Oligonucleotide primers

All oligonucleotides were synthesized and purchased from Sigma-Aldrich. Oligonucleotide stocks were diluted in double-distilled water to 100µM working solutions. All oligonucleotides were stored at -20°C.

Table 3

<table>
<thead>
<tr>
<th>Cloning</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EZH2_5'BamHI</td>
<td>aaaggatccATGGGCCAGACTGGGAAGAAATC</td>
</tr>
<tr>
<td>EZH2_3'BamHI</td>
<td>aaaggatccTCAAGGGGATTTCCATTCTTTTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequencing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TYIP_3'Sprime</td>
<td>GGTGAATATCAAATCCTCTTCG</td>
</tr>
<tr>
<td>TYIP_SEQ1</td>
<td>CCGACAATTTGCAATGGAAGAAATCCTG</td>
</tr>
<tr>
<td>TYIP_SEQ2</td>
<td>CTCTGGCTTTAACAGAAGAACC</td>
</tr>
<tr>
<td>TYIP_SEQ3</td>
<td>ATGTGATAGCGCTGGACACCC</td>
</tr>
<tr>
<td>TYIP_SEQ4</td>
<td>CTGGAGTACAACTACACAGGCC</td>
</tr>
<tr>
<td>pIRES_5'Sprime</td>
<td>GGGAGGCTCATATATAAAGCAGGC</td>
</tr>
<tr>
<td>pJETY3-JMJD2D-191&amp;295mut fw</td>
<td>TTTGAGGGGAAAAATATGGGAAGAACC</td>
</tr>
<tr>
<td>pJETY3_seq IRES rv</td>
<td>TGGTCCAGCAGGCGCCACCTCCG</td>
</tr>
<tr>
<td>pJETY3-JMJD2D #1 fw</td>
<td>TGGTCCAGGAAATCCATGTAACATAATG</td>
</tr>
<tr>
<td>pJETY3 reverse</td>
<td>AAGAGAACAGGGCGAGGTTCCTCG</td>
</tr>
<tr>
<td>GFP fw</td>
<td>CATGGCTCTGATGAGATCGTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site-directed mutagenesis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EZH2_H689A fw</td>
<td>TTTGCAAATGcTTCGTTAAATCCAAACTGCTATGC</td>
</tr>
<tr>
<td>EZH2_H689A rv</td>
<td>GATTTACCGAgcATTTGCAAAACGATTTTGGTAC</td>
</tr>
<tr>
<td>JMJD2D N295A 5'</td>
<td>CAACTGCGGAGGCGCATCGTTGCCACTCCCG CGATGG</td>
</tr>
<tr>
<td>JMJD2D N295A 3'</td>
<td>CCACGCGGAGTGCGAAAATCAGATGGCCTCTCG CAGTTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ChIP or RT-qPCR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>αTetO fw</td>
<td>CCACGCGGTCATCAGTGATAAGAA</td>
</tr>
<tr>
<td>αTetO rv</td>
<td>TCGACATCTGGTTTTAGTCTGTCG</td>
</tr>
<tr>
<td>TetO_rv (for RT-qPCR)</td>
<td>GTAAAATCGTGCTGACCAAGAG</td>
</tr>
<tr>
<td>TetO-21 fw</td>
<td>GTTAGAATCTGCAAGTGGATATTGCAC</td>
</tr>
<tr>
<td>TetO-21 rv</td>
<td>CTGATAGGGAGAGCTTCTGCTGTAAG</td>
</tr>
<tr>
<td>Bsr fw</td>
<td>CAGGAGAACATATCCGGCAGTAC</td>
</tr>
<tr>
<td>Bsr rv</td>
<td>TCCATTCGGAACACTGAATCA</td>
</tr>
<tr>
<td>α21-l fw</td>
<td>GTCTAATTTTTTATGGAATATTCCG</td>
</tr>
<tr>
<td>α21-l rv</td>
<td>AGGGAATGCTTCCATATAAAAATC</td>
</tr>
<tr>
<td>Sat2 fw</td>
<td>TCGATAGGAAATCGGAATGGA</td>
</tr>
<tr>
<td>Sat2 rv</td>
<td>GCATGCGATGGCCGTAAG</td>
</tr>
<tr>
<td>PABPC1_5’UTR fw</td>
<td>CGCCGGTTAGTGCTGAGAATG</td>
</tr>
<tr>
<td>PABPC1_5’UTR rv</td>
<td>GGTGGTCTCGCTGCTCTCACC</td>
</tr>
<tr>
<td>PABPC1-10kb fw</td>
<td>CAGAGATCCTGGGCGCAGAG</td>
</tr>
<tr>
<td>PABPC1-10kb rv</td>
<td>CTTCTCCAACAACCCAGAAAT</td>
</tr>
<tr>
<td>MYT1+40kb fw</td>
<td>ACGAGGATAGTTGAGGTGAGGGA</td>
</tr>
<tr>
<td>MYT1+40kb rv</td>
<td>GGCGATGAAATCGCTCCTG</td>
</tr>
<tr>
<td>BAC fw</td>
<td>GATTATCAAGTTTATTACTCTGAATGGCTATC</td>
</tr>
<tr>
<td><strong>BAC_rv</strong></td>
<td>AGCGCAAGAAGAAATATCCACCG</td>
</tr>
<tr>
<td><strong>Zcen AdjL fw</strong></td>
<td>GCAGGCGAGGGTGCTCATGGAGGGC</td>
</tr>
<tr>
<td><strong>Zcen AdjL rv</strong></td>
<td>GCCCCACTTTTGTTTGTGGTGCTGACCC</td>
</tr>
<tr>
<td><strong>Zcen EdgeL fw</strong></td>
<td>GAGGAGGTTTTGAGGCTCCCCCATCCTGG</td>
</tr>
<tr>
<td><strong>Zcen EdgeL rv</strong></td>
<td>AACCCCCCTTTTACTAGGGAAGGGTGCC</td>
</tr>
<tr>
<td><strong>Zcen Core fw</strong></td>
<td>GGCTGCCGGAAGTGCTCTGAAGCTTGC</td>
</tr>
<tr>
<td><strong>Zcen Core rv</strong></td>
<td>AACTCTTGCCCTGACGTCGTGCC</td>
</tr>
<tr>
<td><strong>Zcen Gap fw</strong></td>
<td>GCTTTCTGGGCGAGTCACCCCTTTGAC</td>
</tr>
<tr>
<td><strong>Zcen Gap rv</strong></td>
<td>CCCCCCTCCATTTTCTTCCAGGCCG</td>
</tr>
<tr>
<td><strong>Zcen EdgeR fw</strong></td>
<td>TGGCCTGTTCTGTTTCTAATGGGGCAGG</td>
</tr>
<tr>
<td><strong>Zcen EdgeR rv</strong></td>
<td>GTGTCATCCACCAGACACTTTTGCC</td>
</tr>
<tr>
<td><strong>Zcen AdjR rv</strong></td>
<td>CGAGAGAGTGTCAGAGCTGTTG</td>
</tr>
<tr>
<td><strong>5cen AdjL fw</strong></td>
<td>TGGCCCACTTGTTCTATACTGGGCTCC</td>
</tr>
<tr>
<td><strong>5cen AdjL rv</strong></td>
<td>TGGGCCCCTCCAGCTGCCAGGGC</td>
</tr>
<tr>
<td><strong>5cen EdgeL fw</strong></td>
<td>GGATGCAGCCGCTGAGCTGTGCC</td>
</tr>
<tr>
<td><strong>5cen EdgeL rv</strong></td>
<td>GCAAGCTGGGCTGGGCTGGGG</td>
</tr>
<tr>
<td><strong>5cen Gap fw</strong></td>
<td>ACCCCCCAAACCCAGAGATAAAGAAG</td>
</tr>
<tr>
<td><strong>5cen Gap rv</strong></td>
<td>GTCAAGGTCTCCAGCTCTGGGTG</td>
</tr>
<tr>
<td><strong>5cen Core fw</strong></td>
<td>GCCGAGGCAGCTGGGGCTTTCCAG</td>
</tr>
<tr>
<td><strong>5cen Core rv</strong></td>
<td>GTCAGGGCTCTGGGCTGGGGC</td>
</tr>
<tr>
<td><strong>5cen EdgeR fw</strong></td>
<td>CTCCCTGCCATCCACTGCTCTCTGCC</td>
</tr>
<tr>
<td><strong>5cen EdgeR rv</strong></td>
<td>CCAACCTTTGTCCAGCTTCCCTGGG</td>
</tr>
<tr>
<td><strong>5cen AdjR fw</strong></td>
<td>ACCCCCCACACCATCCCCCTGCCAG</td>
</tr>
<tr>
<td><strong>5cen AdjR rv</strong></td>
<td>CTGGCACCATTTCCAGCTGCC</td>
</tr>
<tr>
<td><strong>BM23 AdjL fw</strong></td>
<td>GCCCTGATCTCATGGTCTCCAGGGG</td>
</tr>
<tr>
<td><strong>BM23 AdjL rv</strong></td>
<td>ACCAGGATACAAACACCAGGATCTCC</td>
</tr>
<tr>
<td><strong>BM23 EdgeL fw</strong></td>
<td>GCCCTGATCTCAGTCCTGGTCCAGGG</td>
</tr>
<tr>
<td><strong>BM23 EdgeL rv</strong></td>
<td>ACCAGGATACAAACACCAGGATCTCC</td>
</tr>
<tr>
<td><strong>BM23 Core fw</strong></td>
<td>AGTCTGGGAGGAGGAGCTTGG</td>
</tr>
<tr>
<td><strong>BM23 Core rv</strong></td>
<td>ACCTCCCTTGGTGCTGGTGCTGG</td>
</tr>
<tr>
<td><strong>BM23 Gap fw</strong></td>
<td>TGCAACCGGAATGCAACCAGGAG</td>
</tr>
<tr>
<td><strong>BM23 Gap rv</strong></td>
<td>GCCATGCTCTGGCACTTAAATGGCAG</td>
</tr>
<tr>
<td><strong>BM23 EdgeR fw</strong></td>
<td>TGAACAAGCCTAGGGTCTCTAGCC</td>
</tr>
<tr>
<td><strong>BM23 EdgeR rv</strong></td>
<td>AGCATTGGCATCCACGCTGTGAGC</td>
</tr>
<tr>
<td><strong>BM23 AdjR fw</strong></td>
<td>CCCCCACGCAAGCTGCTGTACCACAC</td>
</tr>
<tr>
<td><strong>BM23 AdjR rv</strong></td>
<td>GCCGAGCCTCCAAAAGTGGGGCATGG</td>
</tr>
<tr>
<td><strong>0514 AdjL fw</strong></td>
<td>ACCAGACGGGGCTCACAAGGGG</td>
</tr>
<tr>
<td><strong>0514 AdjL rv</strong></td>
<td>GAGGGGTGCTCCAGCAGGGAGAGG</td>
</tr>
<tr>
<td><strong>0514 EdgeL fw</strong></td>
<td>TGCTGAAGGAGGCTGAGTTGGC</td>
</tr>
<tr>
<td><strong>0514 EdgeL rv</strong></td>
<td>GGACCAGGTCTGCCAGGAGCC</td>
</tr>
<tr>
<td><strong>0514 Core fw</strong></td>
<td>GGATGCAAGCACTGGAAGACTGCC</td>
</tr>
<tr>
<td><strong>0514 Core rv</strong></td>
<td>TGCGCTGACCTGAAGGAGGAGG</td>
</tr>
<tr>
<td><strong>0514 Gap fw</strong></td>
<td>CTCTCGCTCCCGGTCCG</td>
</tr>
</tbody>
</table>
2.1.4. Primary antibodies

<table>
<thead>
<tr>
<th>Target epitope (host)</th>
<th>Origin</th>
<th>Working (application)</th>
<th>dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENP-A, A1 (mouse)</td>
<td>Hiroshi Masumoto</td>
<td>1:500 (IF), 1:100 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>Gg CENP-A</td>
<td>Tatsuo Fukagawa</td>
<td>1:500 (IF)</td>
<td></td>
</tr>
<tr>
<td>CENP-C, R554 (rabbit)</td>
<td>in house</td>
<td>1:500 (IF)</td>
<td></td>
</tr>
<tr>
<td>H3, total, ab10799 (mouse)</td>
<td>Abcam</td>
<td>5µg (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K27me3, 1E7 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:50 (ChiP), 1:200 (IF)</td>
<td></td>
</tr>
<tr>
<td>H3K27me3 07-449 (rabbit)</td>
<td>Upstate</td>
<td>1:200 (IF)</td>
<td></td>
</tr>
<tr>
<td>H3K36me2, 2C3 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:50 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K4me3, CMA004 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:50 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K4me2, CMA003 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:50 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K4me2, 07-030 (rabbit)</td>
<td>Upstate</td>
<td>1:200 (IF)</td>
<td></td>
</tr>
<tr>
<td>H4K20me3 27F10 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:200 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K9me3, 07-523 (rabbit)</td>
<td>Upstate</td>
<td>1:1000 (IF)</td>
<td></td>
</tr>
<tr>
<td>H3K9me3, 2F3 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:50 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>Hec1 Ab3613 (rabbit)</td>
<td>Abcam</td>
<td>1:1000 (IF)</td>
<td></td>
</tr>
<tr>
<td>normal mouse IgG</td>
<td>Calbiochem</td>
<td>10µg (ChiP)</td>
<td></td>
</tr>
<tr>
<td>RING1A ASA3 (rabbit)</td>
<td>Paul Freemont</td>
<td>1:250 (IF)</td>
<td></td>
</tr>
<tr>
<td>CENP-T r42F10 (rat)</td>
<td>Kinya Yoda</td>
<td>1:10 (IF)</td>
<td></td>
</tr>
<tr>
<td>α-tubulin B512 (mouse)</td>
<td>Sigma-Aldrich</td>
<td>1:5000 (IF)</td>
<td></td>
</tr>
<tr>
<td>H2A.Z 07-594 (rabbit)</td>
<td>Upstate</td>
<td>1:100 (IF)</td>
<td></td>
</tr>
<tr>
<td>H3K79me2 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:200 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>H3K16ac 1B2 (mouse)</td>
<td>Hiroshi Kimura</td>
<td>1:200 (ChiP)</td>
<td></td>
</tr>
<tr>
<td>HP1α MAB3584 (mouse)</td>
<td>Chemicon-Millipore</td>
<td>1:2000 (IF)</td>
<td></td>
</tr>
<tr>
<td>53BP1 NB100-304</td>
<td>Novus Biologicals</td>
<td>1:500 (IF)</td>
<td></td>
</tr>
</tbody>
</table>
### 2.1.5. Secondary antibodies

All secondary antibodies for immunofluorescence analysis were purchased from Jackson ImmunoResearch Laboratories. Secondary antibodies against mouse, rabbit, rat and sheep IgG were conjugated to either FITC, Alexa488, TexasRed, Alexa594, Cy5 or Alexa647. All antibodies were used at a 1:200 dilution, except for Alexa594 (1:1000).

### 2.1.6. Commercial kits

**Table 5**

<table>
<thead>
<tr>
<th>Description (catalogue number)</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIAFilter Plasmid Mini kit (12243)</td>
<td>Qiagen</td>
</tr>
<tr>
<td>QIAprep Spin miniprep kit (27106)</td>
<td>Qiagen</td>
</tr>
<tr>
<td>QIAquick Gel Extraction kit (28704)</td>
<td>Qiagen</td>
</tr>
<tr>
<td>QIAquick PCR purification kit (28104)</td>
<td>Qiagen</td>
</tr>
<tr>
<td>QuikChange Site-directed Mutagenesis kit (200518)</td>
<td>Stratagene</td>
</tr>
<tr>
<td>Transcriptor High Fidelity cDNA synthesis kit (05091284001)</td>
<td>Roche</td>
</tr>
</tbody>
</table>

### 2.2. Molecular biology techniques

#### 2.2.1. Preparation of bacterial plasmid DNA

**Transformation of chemically competent E. coli cells**

Individual aliquots of 50µL of chemically prepared competent *E. coli* were thawed on ice. An appropriate amount of plasmid DNA or ligation product was added, and incubated for 30min on ice. Cells were the submitted to a 60sec. heat shock at 42°C in a water bath, and subsequently allowed to recover on ice for 3min. 1mL of LB medium (pre-warmed to 37°C) was added to each sample, which were then incubated at 37°C with gentle agitation for 45min.

Each transformed sample was then plated out on LB-agar plates (pre-warmed to 37°C) containing appropriate antibiotics for selection. Resistance to kanamycin sulphate or ampicillin (Sigma-Aldrich) was selected for in the presence of 50 µg/ml or 100 µg/ml of antibiotic, respectively. Incubation overnight at 37°C allowed for selection of resistants due to plasmid incorporation.

**Isolation of plasmid DNA from E. coli**

*E. coli* bacterial cultures, derived from a single colony picked from LB-agar, were grown overnight at 37°C, in LB medium supplemented with appropriate antibiotic. Miniprep or midiprep isolation of plasmid DNA was performed with appropriate extraction kit supplied by Qiagen, according to the manufacturer's instructions. Purified DNA was re-suspended in TE buffer and...
quantified by spectrophotometric measurement of optical density at 260nm and 280 nm, using a Beckman DU530 UV/Vis spectrophotometer.

2.2.2. Molecular cloning of plasmid constructs

Polymerase chain reaction

DNA sequences of interest were amplified using proof-reading PCR, using the Expand High Fidelity PCR kit (Roche) with sequence-specific oligonucleotide primers. As template, plasmid DNA or a suitable amount of random hexamer-primed total cellular cDNA was used. PCR reactions in a final volume of 50 µl were prepared with double-distilled water and contained 2.6U of the supplied enzyme mix, 200 µM of each dNTP, 200 nM of each primer and 1.5 mM MgCl2 in a 1x dilution of the reaction buffer provided in the kit.

Reactions were run on a Biometra T3000 thermocycler. Standard PCR programme consisted of: a two minute initial denaturation / enzyme heat-activation step at 94°C, followed by cycles of denaturation of dsDNA for 10 sec at 98°C, annealing at a primer-specific temperature for 30 sec, and subsequent elongation at 72°C for one minute per kb of the target sequence. This denaturation/annealing/elongation cycle was repeated 20-25 times, then followed by a final extension at 72°C for 10 minutes.

Specificity and yield of the reaction was subsequently analyzed by agarose gel electrophoresis. Validation of the sequence of the PCR product, after subcloning into a plasmid, was verified by sequencing, as described below.

Agarose gel electrophoresis and purification of DNA fragments

Agarose gels between 0.8 and 2.5 % agarose (Sigma-Aldrich) were prepared in TAE buffer and were supplemented with 0.3 µg/ml ethidium bromide. DNA was loaded onto gel wells in a final 1x dilution of loading buffer. Gel electrophoresis was performed immersed in TAE, at a constant voltage of 100-120 V for a suitable time period required for DNA fragment migration. Agarose gel blocks containing the desired DNA fragments were excised under visualization with low-intensity UV light. Recovery and clean-up of DNA from gel blocks was performed using the QIAquick Gel Extraction kit, following the manufacturer’s instructions.

Restriction, ligation and cloning of DNA

To generate compatible complementary DNA ends, insert PCR products were generated with oligonucleotide primers containing 5’ restriction enzyme target sites present in the desired locus of the target plasmid. Inserts and plasmids were digested with the required restriction endonucleases, in 50 µl reactions prepared in double distilled water, containing 5-10 U of the relevant endonuclease and a final 1x dilution of the recommended restriction enzyme buffer. Where recommended, BSA was added to a final concentration of 100 µg/ml.

When required, the 5’ phosphate group of restriction enzyme-digested plasmid DNA was removed using 5 U of alkaline Calf Intestinal Phosphatase (CIP) (NEB). Reactions were incubated for a minimum of 30min at 37°C.

Digestion reactions were incubated at the required temperature for two hours, prior to agarose gel electrophoresis analysis and purification.

Ligation of purified DNA fragments (inserts and plasmids) was performed using T4 DNA ligase, in a reaction volume of 20 µl containing 1 µl of the T4 DNA ligase and a final 1x dilution of the respective Ligation Buffer. A molar ratio of vector-to-insert DNA between 1:3 and 1:6 relative to 1 µg vector DNA was used. Reactions were incubated at room temperature for 3h prior to transformation of 10 µl reaction volume into E. coli cells.

All restrictions enzymes, ligase, CIP, BSA and respective buffers were purchased from NEB.

Site-directed mutagenesis

Performed with the QuikChange Site-directed Mutagenesis kit (Stratagene) as per the manufacturer's instructions.
2.2.3. Sequencing of plasmids

Sequencing reactions were performed based on the dideoxynucleotide method, using the BigDye v3.1 Cycle Sequencing Kit purchased from Applied Biosystems. 100-250 ng super-coiled plasmid DNA, 1μM of the sequencing primer and 4 μl BigDye mix were used in a final reaction volume of 10 μl. The following sequencing programme was used: an initial denaturation step for two minutes at 96°C, followed by 25 cycles of 30 sec denaturation at 96°C, 15 sec annealing at 50°C and four minutes extension at 60°C. Sequencing reactions were further processed and loaded for analysis on an ABI 3730 DNA Analyzer (Applied Biosystems) by The GenePool Sequencing Facility (The University of Edinburgh). Sequence analysis of associated ABI sequence files was performed using Sequencher software (Gene Codes Corporation).

2.2.4. Reverse transcriptase reaction

Reverse Transcriptase reactions were typically performed on total RNA extracted with TRIzol reagent (Invitrogen), and using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche), according to the manufacturer’s instructions. All reaction steps were carried out on a Biometra T3000 thermocycler. For real-time RT-PCR analysis of centromeric transcripts, synthesis of cDNA from 2 μg RNA was primed using 60 μM random hexamer primers, and reverse transcription was performed for 30 minutes at 50°C. cDNA products were further diluted 1:2 and stored at -20°C until qPCR analysis. See section "qPCR for analysis of transcript samples" in section 2.2.5 (pg. 54) for details.

2.2.5. Quantitative Polymerase chain reaction analysis

Quantitative PCR was performed using a SYBR Green master mix (Sigma) containing the hot start enzyme, dNTPs and SYBR Green dye, in a final reaction volume of 20 μl. All primers were used at a final concentration of 400 nM. Reactions were run on a LightCycler 480 system (Roche), in 96 well plates (Roche). Cycle parameters were used as follows.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Ramp Rate</th>
<th>Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>95°C Hot Start</td>
<td>5 min</td>
<td>4.4°C/sec</td>
<td>none</td>
</tr>
<tr>
<td>95°C Amplification (x40 cycles)</td>
<td>15 sec</td>
<td>4.4°C/sec</td>
<td>none</td>
</tr>
<tr>
<td>61°C</td>
<td>15 sec</td>
<td>2.2°C/sec</td>
<td>none</td>
</tr>
<tr>
<td>72°C</td>
<td>45 sec</td>
<td>4.4°C/sec</td>
<td>single</td>
</tr>
<tr>
<td>Melt Curve</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95°C</td>
<td>30 sec</td>
<td>4.4°C/sec</td>
<td>none</td>
</tr>
<tr>
<td>55°C</td>
<td>30 sec</td>
<td>2.2°C/sec</td>
<td>none</td>
</tr>
<tr>
<td>99°C</td>
<td></td>
<td>0.3°C/sec</td>
<td>continuous (2°C)</td>
</tr>
</tbody>
</table>

For every plate and primer pair used, a serial dilution of the relevant template DNA was included to determine a log-scale standard curve, and account for differential reaction efficiencies. Specificity of reactions was validated by product melting curve analysis. Reaction crossing points (LightCycler 480) were determined using the corresponding software, and values were exported to, further processed and analyzed in Excel (Microsoft). Crossing points (LightCycler 480) were determined using the 2nd derivative maximum algorithm.

**qPCR for analysis of ChIP samples**

2 μl of Input or ChIP’ed DNA was used in each reaction. To account for differential primer efficiency and quality of sonicated chromatin, standard curves were prepared from the corresponding
Input material, and attributed arbitrary values. Detection of DNA quantity in samples was calculated relative to this standard curve. Primer pairs used for ChIP are listed in section "2.1.3. Materials: Oligonucleotide primers".

For each antibody and locus analyzed by ChIP, the amount of DNA detected in the sample was presented as percentage of recovery of immunoprecipitated antigen, relative to amount detected in the Input (total antigen in the original extract).

For comparison analyses of 1C7-derived cell lines from different experimental conditions, quantifications for each individual locus were further normalized to the value of the α21-I satellite locus, following subtraction of background signal (IgG), for each separate antibody used.

No ChIP signals were normalized to the levels quantified for either H3 or H4, for each respective locus. This decision was made to preserve information (as histone quantifications in chicken cells are shown, separately). The biological readout and consequence of chromatin marks, in any given locus, is dependent on their actual enrichment on chromatin, independently of the availability of the histone substrate for each particular mark. For example, the total enrichment of H3K9me3 in a given locus is what is relevant to its biological function (heterochromatin formation, via H3K9me3-interacting proteins), regardless of how much H3 is available at that locus to be modified to H3K9me3.

**qPCR for analysis of transcript samples**

Real-time PCR analysis of cDNA was performed using a SYBR Green Mastermix (Sigma) on a LightCycler480 system (Roche). A cDNA amount equivalent to ~70 ng for low-expressed loci (such as α-satellite loci) or ~0.7 ng (highly-expressed loci such as Bsr or β-actin) of sample cDNA was used per reaction. Primer pairs used for transcript analysis are listed in section "2.1.3. Materials: Oligonucleotide primers". For each primer pair and every qPCR run, a standard curve was created from genomic DNA derived from the corresponding cell line, thereby calculating the transcript copy numbers in the sample relative to the copy number in the genome. Background values (no reverse transcriptase) were subtracted, and all values were normalized to β-actin expression and arbitrarily multiplied by 1x10^6 for ease of visualization.

### 2.2.6. Expression plasmid constructs

<table>
<thead>
<tr>
<th><strong>Table 7</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plasmid</strong></td>
</tr>
<tr>
<td>TetR-EYFP</td>
</tr>
<tr>
<td>tYIP</td>
</tr>
<tr>
<td>tYIP-tol2</td>
</tr>
<tr>
<td>tYIP-KAP1</td>
</tr>
<tr>
<td>TetR-EYFP-HP1α</td>
</tr>
<tr>
<td>tYIP-LSD1</td>
</tr>
</tbody>
</table>
2.3. Eukaryotic cell culture techniques

2.3.1 Transfection of plasmid DNA

**Transfections for transient expression of plasmid construct**

Transient transfections were performed using either Fugene HD (Roche) or Fugene 6 (Roche). Cells were seeded in an appropriate dilution, to achieve 50-60% confluency on the day of transfection. The transfection complex was prepared essentially according to the manufacturer’s instructions, in 100 µl of OptiMEM (Invitrogen). Ratio between transfection reagent and plasmid DNA was 3µl:1µg. Up- or downscaling of the transfection complex mixture was done proportionally to the surface area of the culture vessel used. For samples intended for microscopy analysis, cells were seeded in 12-well plates, on top of sterile glass coverslips.

Were desired, puromycin (for TetR-EYFP- or tYIP-based constructs) or hygromycin (for pJETY3-based constructs) was added one day after transfection, for 24 hours, to enrich for transfected cells. Transfected cells used for RT-qPCR or ChIP analysis were typically selected for in 2 µg/ml or greater than 3µg/ml puromycin, respectively.

**Transfections to generate cell lines stably expressing plasmid construct**

HeLa 1C7 cells stably expressing TetR-EYFP-EZH2 were generated by transfection with TYIP-EZH2-To12, using Fugene HD (Roche) as described previously. Clonal cell lines were isolated by limiting dilution and grown in cDMEM, in the presence of 4µg/mL blasticidin S (Invitrogen), 1µg/mL doxycycline (Sigma) and 1-2µg/mL puromycin (Sigma).

HeLa 2-4 cells stably expressing TetR-EYFP-JMJD2D were generated by transfection with TetR-EYFP-JMJD2D, using Fugene HD (Roche) as described previously. Clonal cell lines were

<table>
<thead>
<tr>
<th>Construct</th>
<th>Details</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCENP-A-SNAP-IP</td>
<td>CENP-A-SNAP-3xHA fusion construct cloned into pIRESpuro2 (Clontech), digested with BamHI and NotI.</td>
<td>Puromycin, A kind gift from Lars Jansen</td>
</tr>
<tr>
<td>tYIP-EZH2</td>
<td>PCR product of EZH2 (digested with BamHI) cloned into tYIP (digested with BamHI).</td>
<td>Puromycin, This work</td>
</tr>
<tr>
<td>tYIP-EZH2 H689A</td>
<td>Site-directed mutagenesis on tYIP-EZH2, to introduce H689A mutation.</td>
<td>Puromycin, This work</td>
</tr>
<tr>
<td>tYIP-EZH2-Tol2</td>
<td>PCR product of EZH2 (digested with BamHI) cloned into tYIP-Tol2 (digested with BamHI).</td>
<td>Puromycin, This work</td>
</tr>
<tr>
<td>pJETY3</td>
<td>TetR-EYFP and IRES-Hyg cloned into pJTI Fast DEST (digested with PstI and EcoRI).</td>
<td>Hygromycin, A kind gift of Hiroshi Masumoto, Ohzeki et al., 2012</td>
</tr>
<tr>
<td>TetR-EYFP-JMJD2D</td>
<td>PCR product of JMJD2D cloned into pJETY3.</td>
<td>Hygromycin, A kind gift of Hiroshi Masumoto and Takahiro Nagase</td>
</tr>
<tr>
<td>TetR-EYFP-JMJD2D D195A</td>
<td>Site-directed mutagenesis on TetR-EYFP-JMJD2D, to introduce H689A mutation.</td>
<td>Hygromycin, This work</td>
</tr>
<tr>
<td>TetR-EYFP-BMI1</td>
<td>PCR product of BMI1 cloned into pJETY3.</td>
<td>Hygromycin, A kind gift of Hiroshi Masumoto</td>
</tr>
</tbody>
</table>
isolated by limiting dilution and grown in cDMEM containing 10% FBS (Gibco), in the presence of 4\(\mu\)g/ml blasticidin S (Invitrogen), 1\(\mu\)g/ml doxycycline (Sigma) and 400\(\mu\)g/ml hygromycin (Sigma).

Nuclear localization and targeting to the HAC after doxycycline wash out was confirmed by fluorescence microscopy. Doxycycline wash out experiments where conducted as previously described.

2.3.2. Doxycycline washout experiments

To prevent binding of tetR fusion constructs to the alphoid\(^{tetO}\) array, all stably expressing cell lines were grown in the presence of 1 \(\mu\)g/ml doxycycline. To initiate targeting of tetR fusion constructs, a washout protocol was established.

Cells were seeded in cell culture dishes, and wash-out was initiated the subsequent day. Culture medium was removed and the cell layer was rinsed twice with excess D-PBS pre-warmed to 37°C. Cells were then incubated in DMEM without doxycycline for 30 minutes at 37°C. These steps were repeated once followed by a final two washes in D-PBS. Cells were then incubated in cDMEM as previously, but omitting doxycycline from the medium, and allowing cells to grow for the desired time period.

2.3.3. Monastrol arrest-and-release assay

Adapted from Kapoor et al 2006\(^{446}\). Cells were treated for 2h with 100\(\mu\)M Monastrol, with or without the presence of 2\(\mu\)M ZM447432 as a positive control. Cells were washed and released into fresh medium (containing or not ZM447432) for 45 min before fixation. This enriches the population of cells in metaphase.

2.3.4. CENP-A-SNAP experiments

**quench-chase-pulse and pulse-chase**

For quench-chase-pulse experiments, cells were transiently co-transfected with a TetR-EYFP fusion construct and a plasmid expressing CENP-A~SNAP. After 24h, free SNAP-tag was quenched for 30 minutes in medium containing 10\(\mu\)M non-fluorescent bromothylpteridine (BTP, SNAP-Cell Block, NEB). BTP was washed from the cells by three washes with D-PBS followed by another 3 washes after 30min, and chased for 7h to allow synthesis of new CENP-A and progression into G1. Newly-synthesized SNAP-tagged CENP-A was then fluorescently labelled in medium containing 3\(\mu\)M TMR-Star (NEB) for 15 minutes, washed three times with D-PBS, and fixed for immunofluorescence after 30min.

For pulse-chase experiments, the cells were transfected as above, and after 24h, initially labelled with 3\(\mu\)M TMR-Star for 30min. Cells were then washed three times with D-PBS and again three times after 30min. The cells were then chased for a suitable amount of time.

Quantification of HAC-associated TMR-Star signal was performed in the same manner as for centromeric proteins, which is detailed below.

2.3.5. Preparation of chromosome spreads and stretched chromatin fibers

Confluent cultures were incubated for 2-3 hours in 0.1 \(\mu\)g/ml colcemid (KaryoMax, Invitrogen) or 300 nM TN-16 (Wako, Osaka, Japan) at 37°C. Human cells thus arrested in mitosis were collected by shake-off, centrifuged at 800 rpm, and re-suspended in hypotonic buffer (75mM KCl) for 10 minutes at 37°C. Cells were then cytopun (Shandon Cytospin 4) onto ethanol-washed glass slides at 1800rpm for 5min, and subsequently processed for immunofluorescence.

In the case of non-adherent chicken DT40 cells, the culture was first enriched with cells arrested in mitosis using colcemid, and a fraction of the total culture was centrifuged and re-suspended in hypotonic buffer, as described above.

Stretched chromatin fibers of mitotic chromosomes were obtained from metaphase spreads subjected to unfixed immunofluorescence procedure (see below).
2.3.6. Immunofluorescence analysis

**Indirect immunofluorescence staining**

Cells expressing fluorescently labelled fusion constructs and/or to be processed for indirect immunofluorescence staining were grown on sterile no. 1.5 glass cover slips. Cover slips were washed twice in pre-warmed D-PBS, and cells fixed in 2.5 - 4% PFA/PBS for 5-10 min at room temperature. Cells were permeabilized for five minutes in PBS-T (0.2 % Triton X-100 in PBS) and either counterstained in 1 µg/ml Hoescht 333342 (Sigma) in PBS and mounted on Superfrost glass slides in VectaShield (Vector Labs), or further labelled with antibodies as follows.

Fixed and permeabilized cells were pre-blocked in 3 % BSA/PBS-T for 15 minutes at 37 °C prior to incubation with appropriately diluted primary antibody in 1 % BSA/PBS-T for one hour at 37 °C. Cells were washed in PBS-T and subsequently incubated for 30 minutes at 37 °C with fluorophore-conjugated secondary antibody (Jackson Labs) diluted 1:200 in 1 % BSA/PBS-T. After excess antibody was washed off in PBS, cells were counterstained with Hoescht 333342 and mounted onto glass slides as above.

**Staining of unfixed metaphase spreads**

Preparation and staining of unfixed mitotic chromosomes was essentially performed as described in Keohane et al, 1996. In brief, chromosome spreads were prepared and cytopsnn onto glass slides, as described previously, and subsequently incubated in KCM buffer (10mM Tris pH 8.0; 120mM KCl; 20mM NaCl; 0.5mM EDTA; 0.1% Triton X-100) for 10 min. Samples were then labelled with primary and secondary antibodies (diluted in 1% BSA/KCM buffer), fixed in 4% PFA/KCM and counter stained with Hoescht 333342. Cells were subsequently mounted in VectaShield (Vector Labs) and covered with glass slides.

**Microscopy analysis**

Images were acquired on a DeltaVision Core system (Applied Precision) using an inverted Olympus IX-71 stand, with an Olympus UPlanSApo x 100 oil immersion objective (NA 1.4) and a 250W Xenon light source. Camera (Photometrics Cool Snap HQ), shutter and stage where controlled through SoftWorx (Applied Precision). Z-series were collected with a spacing of 0.2 mm, and image stacks were subsequently deconvolved in SoftWorx.

For time-lapse live-cell microscopy, cells were grown in coverglass chambers (Lab-tek II, Thermo Fisher Scientific), and culture medium was replaced with warm Leibovitz medium (Gibco) for 1h before start of imaging. Afterwards, cells were imaged for an appropriate amount of time, and stage temperature was maintained at 37°C.

All microscopy images after acquisition and optional deconvolution were projected and analyzed in ImageJ. All quantifications were performed using the software's tool suite.

**Fluorescence signal quantification**

All fluorescence signal quantification was performed on non-deconvolved microscopy images acquired at a 1x1 binning, using identical exposure conditions for each experimental subset. Cells displaying more than one HAC were only quantified for one of those, determined randomly. Furthermore, for CENP signal quantification, only those cells that showed a clear overlap of CENP staining and EYFP (HAC) signal were considered.

To quantify centromeric proteins in interphase cells, an ImageJ macro (HAC & CRaQ), adapted from that of Bodor et al, 2012, was used to assess HAC centromere protein levels relative to endogenous centromeres. HAC-associated signals from maximum intensity projections were measured and the nuclear background signal subtracted. The same measurements were applied to endogenous centromeres detected by the macro, and the HAC associated signal was normalized against the median measurement of those centromeres.

To quantify HAC-associated chromatin mark signals, or centromere protein signals in mitotic cells, an ImageJ macro (HAC Area quantifier) was used. Briefly, for each cell, a maximum intensity projection of 5 Z-stacks, centered on the HAC, was used for quantification. A region of interest was thresholded to the boundaries of the TetR-EYFP signal, defining the HAC area. HAC-associated
signals in other channels were then quantified, and the mean signal of 3 regions adjacent to the HAC was subtracted, as background signal.

2.4. Biochemical techniques

2.4.1. Extraction of total cellular RNA

Total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer’s instructions.

2.4.2. Extraction of genomic DNA

Sub-confluent cell cultures were harvested, cells were washed once in D-PBS and then re-suspended in TE buffer containing 0.5 % SDS. RNase A was added to a final concentration of 20 µg/ml and lysates were incubated for one hour at 37°C. Subsequently, proteinase K was added to a final concentration of 100 µg/ml, and lysates incubated overnight at 50°C. Genomic DNA was isolated by standard Phenol/Chloroform extraction, ethanol-precipitated and re-suspended in TE buffer. Concentration and quality of the DNA was assessed by spectrophotometric determination of the OD at 260nm and 280nm using a Beckman DU530 UV/Vis spectrophotometer.

2.4.3. Chromatin immunoprecipitation

A minimum of 5x10⁶ human cells or 50x10⁶ chicken DT40 cells were isolated for each separate ChIP experiment. Exponentially growing cells were collected and re-suspended in D-PBS to a final concentration of 1x10⁶/ml or 10x10⁶/ml (for human or chicken cells respectively) and crosslinked in a final 1% formaldehyde (Fischer Scientific) for 5 minutes at room temperature, followed by quenching in 0.5M glycine for an additional 5 minutes.

Cells where washed in TBS and lysed in lysis buffer (10mM Tris pH8.0; 10mM NaCl; 0.5% NP-40) containing protease inhibitors (1µg/ml CLAP; 0.5µg/ml aprotinin; 1mM PMSF) for 10 minutes on ice. Crosslinked chromatin was briefly washed again in lysis buffer before being snap-frozen and stored.

Chromatin pellets derived from human cells were re-suspended in 300µl Dilution Buffer 1 containing protease inhibitors, and sheared by sonication in a Bioruptor sonicator (Diagenode) for 16 cycles of 30sec on / 30sec off at a high setting at 4°C. Afterwards, insoluble debris was spun down for 10min at 10000g. Supernatants were then mixed with 500µl of Dilution Buffer 2 and 500µl RIPA buffer containing 150mM NaCl (RIPA-150) and protease inhibitor, and used as ChIP Input material.

To digest chromatin from chicken DT40 cells, sample pellets were re-suspended in Dilution Buffer 1 (without EDTA added) supplemented with 1mM CaCl. Material was digested with 200 U/mL of Micrococcal nuclease (Worthington Biochem. Corp.) for 30min at 21°C. The reaction was stopped with addition of EDTA to a final concentration of 2mM. After digestion, chromatin was sheared by sonication for 9 cycles, with the same settings as above. Finally, Dilution Buffer 2, RIPA-150 and protease inhibitors were added, as above, to constitute the ChIP Input.

Anti-mouse IgG Dynabeads M-280 (Invitrogen) were pre-blocked with BSA and subsequently coupled with the relevant antibodies for 4-6 hours in RIPA-150/0.5% BSA at 4°C, washed twice in RIPA-150/0.5% BSA, and 500µl of the sheared chromatin was incubated with the beads over night at 4°C. Beads were then washed twice with RIPA-150 containing protease inhibitors, followed by two washes in RIPA-500 and a final wash in TE pH8.0.

Initially, for ChIP experiments described in section “3.2. Results: Centromere activity can coexist with polycomb chromatin”, the following elution protocol was used. Antibody/chromatin complexes where eluted at 65°C in TE/1% SDS. An equal volume of Post-Elution Buffer (10mM Tris pH8.0; 9mM EDTA; 600mM NaCl) was added and crosslinks reversed at 65°C over night. Samples where treated with 100µg/ml RNase A and 5µl of proteinase K buffer followed by phenol/chloroform extraction, and DNA was finally re-suspended in TE.
For subsequent ChIP experiments, described in sections "3.1. Results: Chromatin mapping of non-repetitive chicken centromeres" and "3.3. Results: Heterochromatin stabilizes the core centromere and prevents DNA damage", improvements in DNA purification were made. Elution was performed in 100µl of double-distilled water containing 10% Chelex beads (BioRad), previously inactivated for 10min at 93°C. Samples were decrosslinked at 93-100°C for 12min, and treated with RNase A and Proteinase K as described above. Proteinase K was subsequently inactivated at 93°C for 10min, samples were quickly centrifuged, and 60 µl of bead-free supernatant, containing ChIP'ed DNA, was recovered and stored at -20°C until qPCR analysis.

See section "qPCR for analysis of ChIP samples" in section 2.2.5 (pg. 53) for details.
3. Results

3.1. Chromatin mapping of non-repetitive chicken centromeres

3.1.1. Background

Centromere genomic organization remains poorly characterized in many eukaryotes, despite knowledge of the sequence of their underlying repeats, thus presenting a problem for the full mapping of their domain composition. This is due to the repetitive nature of these regions: sequence assemblage of long tandem repetitive arrays is unfeasible due to complete overlap of most sequences. This fact has made in-depth mapping of natural centromeres, through whole-genome sequence mapping methods, nigh-impossible. This has limited the characterization of centromeric chromatin marks and transcripts to FISH and multi-copy ChIP-qPCR techniques only, and indirectly by biochemical co-purifications.

There is, thankfully, available evidence from model organisms that possess centromeres amenable to mapping, either by lack of repetitive sequences or by being small enough to be cloned, such as those of yeasts. These have yielded significant insights into how centromere chromatin might be regulated in-situ, and the distribution patterns of adjacent chromatin signatures (as discussed in Introduction). However, centromeres of vertebrates remain poorly characterized at high resolution.

Previous work from Shang et al has led to the characterization of centromeric repeats of Gallus gallus, and the finding that three of its natural centromeres are not located on tandem repeat arrays. Those centromeres have therefore been fully sequenced. Together with members of the genus Equus (which also possess some repeat-less
centromeres)\textsuperscript{395,450}, chickens are among the few vertebrates with natural centromeres which are fully sequenced.

Like humans\textsuperscript{394,451}, there is similarity between the centromere repeat units of each chicken chromosome: they consist of a characteristic repeat sequence interspersed with a retrotransposon, in arrays spanning roughly 400 Kb\textsuperscript{445}. However, thanks to precise ChIP-seq mapping onto the chicken reference genome\textsuperscript{452}, the centromere cores of chromosomes Z, 5 and 27, as measured by CENP-A coverage, were found to span only about 35Kb\textsuperscript{445}. For the first time, established centromeres (non-neocentromeres) of a vertebrate were fully sequenced and the pattern of their CENP-A distribution known.

Shang \textit{et al} proceeded to develop an ingenious model system where, through genomic engineering of the region surrounding the chr. Z centromere, they introduced flanking loxP sites, allowing for conditional excision of the centromeric region\textsuperscript{445}. Expression of Cre recombinase in these cells triggers excision of the chr. Z centromere, and leads to mis-segregation of the chromosome by its inability to attach to the mitotic spindle. This eventually causes cell death\textsuperscript{445}.

Capitalizing on this conditional chr. Z centromere, Shang \textit{et al} designed a pipeline to experimentally generate neocentromeres in DT40 cells, much like similar studies in \textit{C. albicans}\textsuperscript{347,453} and \textit{S. pombe}\textsuperscript{348}, by selection of survivors who can nucleate a new centromere on the Z chromosome after excision.

The work that follows reports on the characteristics of chicken natural centromeres and neocentromeres, their influence on local genome and chromatin processes, and possible determinants of neocentromere location.
3.1.2. Results

The following results are the product of a collaborative effort with the Fukagawa and Earnshaw labs (the latter being my contribution), and respective credit is attributed throughout the text. Part of these results have been published in "Chromosome engineering allows the efficient isolation of vertebrate neocentromeres" (included as section 6.4 of this document), and will be referenced as appropriate. The remaining unpublished results are the product of my own PhD research only.

**Conditional generation and mapping of neocentromeres in chicken chromosomes**

The chicken genome has 3 chromosomes whose centromeres are not located on a repetitive tandem array: chr. 5, 27 and Z. Given that the Z sex chromosome is present in a single copy in chicken DT40 cells, making targeted genomic engineering and interpretation of the results much easier, Shang et al proceeded to build a conditional system to eliminate this centromere from the chromosome.

They introduced, by homologous recombination, two flanking loxP sites 46Kb away from each edge of the CENP-A domain. This allows for Cre-induced conditional excision of the entire CENP-A domain, by homologous recombination between the two sites. To select desired clones after excision of the intervening region, a double positive-negative selection strategy was employed. A Herpes Simplex Virus thymidine kinase (HSV-tk) gene was introduced inside the excision region, and a zeocin resistance (BleoR) gene outside the region (which only becomes expressed after excision). The resulting DT40 cell line was dubbed #Z3 (fig. 2.1A). After expression of Cre recombinase, the CENP-A domain of the chr. Z centromere was excised as an extrachromosomal circle; subsequent addition of FIAU (5-iodo-2'-fluoroorauracil) to the culture medium was used to kill cells that retain expression of HSV-tk, and Zeocin used to select for cells that express BleoR. This resulted in a surviving population that had lost the original centromere region, had successfully recombined the
centromere flanking regions of the Z chromosome together, and must have secured successful mitotic transmission of the Z chromosome (by neocentromere formation or chromosome fusion) in order to sustain life (fig. 2.1A).

This approach led to the isolation and identification of 136 clones, 92.64% of which had formed neocentromeres somewhere else on the Z chromosome (fig. 2.1B), while the remaining clones survived due to fusion of the Z to another chromosome. Several neocentromeres were tested for their ability to recruit kinetochore proteins and correctly segregate during mitosis. The majority behaved similarly to the original Z centromere, including correct localization of Aurora B and retention of cohesion under long-term mitotic arrest. This is in contrast to some human neocentromeres, which have demonstrated some instability regarding these characteristics 390,391.
Fig 2.1 – Conditional generation and mapping of neocentromeres in chicken DT40 cells (by Shang et al 2013)

A. Schematic of engineered chr. Z in DT40 #Z3 cells, and pipeline of centromere excision. Genomic engineering, by homologous recombination, was used to introduce flanking loxP sites on each side of the centromere of chromosome Z. Adapted from Shang et al 2013.

#Z3 cells were subjected to excision by loxP recombination, via conditional expression of Mer-Cre-mer recombinase by addition of hydroxytamoxifen. Selection of successful recombinants was achieved by two parallel methods. First, introduction of a bleomycin (zeocin) resistance cassette, which only becomes expressed after excision, via reconstitution of a β-actin promoter upstream of the BleoR gene. Second, by introduction of a Herpes Simplex Virus thymidine kinase (HSV-tk) cassette inside the excised region, which when expressed provides negative selection when 5-iodo-2'-fluoroarauracil (FIAU) is present in the growth medium. After selection, survivors had successfully excised centromeres and had lost the extra-chromosomal circular excision product. To sustain life, survivors must have either generated neocentromeres in another locus of the Z chromosome, or fused the Z chromosome to another. A total of 136 isolates were analyzed for presence of either outcome.

B. Immunofluorescence-FISH microscopy analysis of isolated clones from conditional centromere removal. Adapted from Shang et al 2013.

Isolates were characterized for either chromosomal fusion or neocentromere formation, and the latter were categorized (from Type I-V or Fusion) according to the position of the neocentromere in the Z chromosome. Active centromere detection was achieved with an antibody against CENP-T, and chromosomal orientation was confirmed with a FISH probe against the macro-satellite in the q arm of the Z chromosome.

C. High-resolution mapping of CENP-A distribution on original chr. Z centromere and neocentromeres. Adapted from Shang et al 2013.

ChIP-seq. analysis of CENP-A (or FLAG-tagged CENP-A) was performed in each of the cell lines, by mapping immunoprecipitated DNA against the chicken genome database. Graph shows distribution of CENP-A across the (neo)centromere region, inset shows distribution across the whole chromosome (in red). "Gaps" in CENP-A distribution are indicated by orange arrows. Additional local genomic features (genes, non-coding RNAs, in green and blue) are also mapped relative to CENP-A signal. GC content of the analyzed region and distribution of repetitive elements is shown in the two bottom lines of each graph.
**A**

**Zoen-cKO (Z#3)**

![Diagram of Zoen-cKO (Z#3)]

- ChrZ centromere (42.62 Mb)
- LoxP
- 46 kb
- 35 kb
- 46 kb
- 127 kb
- + Cre
- 127 kb
- + FAU
- + Zeocin

**B**

- **Neocentromere categories:**
  - Type I: acrocentric (p arm)
  - Type II: submetaacentric (p arm)
  - Type III: acrocentric
  - Type IV: submetaacentric (q arm)

Fluorescent centromeric function secured through fusion with another chromosome.

**C**

**Z#3**

- CENP-A IP
- LOC430532
- ChrZ
- another chr

**#BM23**

- CENP-A IP
- LOC10085784B

**#0514 CA-FLAG**

- CENP-A IP
- MAMDC2
- SMC6

**Repeat Elements (RE):** DNA transposons, LTR elements, LINEs, Low complexity, Simple repeats
With this resource, Shang et al proceeded to characterize several of these centromeres in-depth, by performing ChIP-seq to analyze its genomic distribution of CENP-A. They showed that neocentromeres occupy regions of roughly about 45Kb (283, fig. 2.1C). Looking at this data, I frequently observed deep 'gaps' in the CENP-A signal, that otherwise did not conform with the overall distribution, and seemed to indicate CENP-A-poor regions within the centromere (fig. 2.1C). These gaps could be found both at the original Z centromere and at several neocentromeres. Given the fact that ChIP-seq signals represent the average coverage of CENP-A across the total cell population, the recurrent presence of deep gaps in the population suggested that certain underlying genomic features might hamper CENP-A deposition. Shang et al also noticed that neocentromeres formed mostly on regions of higher-than-average AT-content, seldom formed on top of genes, and do not show an absolute requirement for any sort of DNA transposons within them, in contrast to what had been suggested for some human neocentromeres 455. Additionally, the chicken neocentromeres showed no bias for any specific repetitive element or indeed any sequence similarity between them (fig. 2.1C).

From these results they concluded that neocentromere formation was independent of any detectable underlying sequence. Curiously, 76% of isolated neocentromeres were are metacentric, indicating that neocentromere formation was inclined to take place near the location of the original centromere (fig 2.1B, 2C). This had also been observed in C. albicans neocentromere formation studies 347,453. Thanks to ChIP-seq analysis, Shang et al were able to detect very low levels of CENP-A accumulation, higher than the genomic average, in a 2Mb region surrounding the active original Z centromere and the characterized neocentromeres. Strikingly, this enrichment was not present if the locus did not contain an active centromere. This result was independently verified for chromosome 5 as well, in which one of its copies in DT40 cells was engineered in the same way as for chromosome Z. Shang et al hypothesized that these flanking low-level accumulations of CENP-A might provide nucleation sites for neocentromeres.
Neocentromere formation negatively impacts gene transcription

A transcription-associated state of chromatin, mainly characterized by the H3K4me2 and H3K36me2 marks, has been shown to be present at natural centromeres of some metazoans 335,456, and to be required for centromere function in a human artificial chromosome 405,407,408. Additionally, transcription-associated chromatin marks have been shown to co-localize partially with neocentromeres in human cells 391,457,458. However, chicken centromere mapping, as assayed by super-resolution microscopy on chromatin fibers, had revealed no evidence for H3K4me2 co-localization with CENP-A 247.

To verify such claims in this system, and to investigate putative chromatin determinants for neocentromere formation, I analyzed the Z centromere and two neocentromere loci (cell lines #BM23 and #0514) by ChIP. To identify possible shifts in chromatin signature due to neocentromere formation, I compared the chromatin state of each locus, before (in parental #Z3 cells) and after (in isolated clones) neocentromere formation. Additionally, the #0514 cell line is the only isolate which Shang et al found to have formed a neocentromere on an active gene (fig. 2.1C, 2B, 283), thus constituting a prime opportunity to observe possible effects of centrochromatin on gene activity.

Using the mapped ChIP-seq. distribution of CENP-A, primer sets were designed against specific loci, as follows: flanking regions, edges of CENP-A boundaries, the core of CENP-A signal distribution and one distinct CENP-A 'gap' mentioned previously. After performing ChIP against the above loci, I observed that enrichment for H3K4me2 did not correlate with centromere function, as it was not significantly enriched in either the original Z centromere or the #BM23 neocentromere locus, before or after formation of its neocentromere (fig. 2.2A). The #0514 neocentromere locus, on the other hand, had a significant enrichment of H3K4me2 corresponding to the CENP-A "gap" locus before neocentromere formation. This corresponds to the TSS of the MAMDC2 gene, and both the distribution of H3K4me2 and low CENP-A density fit with the reported dynamics of a
promoter locus. Upon neocentromere formation (in #0514 cells), the H3K4me2 signal was significantly reduced, suggesting neocentromere formation may have a negative effect on transcription activity (fig. 2.2A). In my experiments, distribution of H3K36me2 also did not show significant correlation with CENP-A distribution, among the neocentromeres analyzed, although in #0514 cells some redistribution of the signal occurred upstream of the MAMDC2 gene, upon neocentromere formation (fig. 2.2A).

Quantification of cellular transcripts by Shang et al showed that in #0514 cells, MAMDC2 was expressed 10-fold less than wild-type DT40, suggesting that indeed neocentromere formation can impair transcription (fig. 2.2C), just as observed in C. albicans and S. pombe.

The presence of the MAMDC2 TSS seemed to disrupt CENP-A distribution. While most observed "gaps" in the CENP-A signal were relatively narrow (1Kb or less), the gap in the #0514 neocentromere, which corresponded to the TSS of the MAMDC2 gene, was quite pronounced (fig. 2.1C, 2.2A). CENP-A enrichment downstream of the TSS, on the body of the gene, appeared unaffected, indicating the effect is specific for the TSS-proximal region.

Because the Z chromosome in #0514 cells segregates efficiently, these results suggest that presence of an underlying TSS, while having a local negative effect on CENP-A distribution, has little impact on overall centromere function. On the other hand, centromere formation can negatively impact gene transcription.
Fig 2.2 – Neocentromere formation negatively impacts local active transcription and chromatin state, but otherwise can form on transcribed regions

A. ChIP analysis of active transcription-related chromatin marks on chr. Z centromere and neocentromeres, before and after neocentromere formation. Antibodies against H3K4me2 and H3K36me2 were used for immunoprecipitation, in #Z3 cells (before neocentromere formation, in grey) or in #BM23 and #0514 cells (after neocentromere formation, in pink). The original centromere of chr. Z (in #Z3 cells) was also analyzed. Primer locations, used for qPCR of ChIP products, are shown below each graph, mapped relative to the CENP-A ChIP-seq. distribution. Primers against the gene body of active SMC2 gene were used as positive controls for an actively transcribed locus. Mean of three independent experiments, error bars show SEM. Note that the #0514 neocentromere co-localizes with the promotor of the active gene MAMDC2 (see text).

B. Gene expression analysis of genes co-localizing with the characterized neocentromeres. By Shang et al, 2013. Transcript level, for both chicken embryos and DT40 cells, was analyzed by RT-qPCR. Of all those genes, only MAMDC2 (which co-localizes with the neocentromere of #0514 cells) was found to be expressed in DT40 cells.

C. Quantitative expression analysis by RT-qPCR of MAMDC2, in wild-type DT40, or #0514 cells (stably expressing CENP-A-FLAG or otherwise). By Shang et al 2013. MAMDC2 transcripts levels decrease by 10-fold or more in #0514 cells compared to wild-type DT40. Mean of three independent experiments, errors bars denote SD.

D. Mapping of neocentromere locations (in pink) on chr. 5, after excision of its centromere, and relative position of annotated genes. By Shang et al 2013. Most neocentromeres form in positions adjacent to the original centromere, but otherwise can also form on top of actively transcribed genes (‘Gene ON’).
Actively transcribing regions are not refractory to neocentromere formation

In previous studies of experimental neocentromere generation, the genomic distribution of neocentromeres was primarily found on sites proximal to the original centromere, and on intergenic regions with no sequence similarity between them.\(^{347,453}\) Distribution of the neocentromeres across the Z chromosome, in the several lines isolated by Shang et al., showed a clear bias for formation on intergenic regions as well, with only a few neocentromeres formed on top of mapped genes.\(^{283}\) Of those, only the MAMDC2 gene was found to be expressed in DT40 cells (fig. 2.2B).\(^{283}\) This initially suggested that transcribed regions could somehow prevent centromere nucleation, with the neocentromere in the #0514 clone being an exception. Some of the centromere regions of O. sativa do indeed contain genes, yet CENP-A is only associated with the intergenic regions between those genes, in interspersed stretches,\(^{463}\) suggesting that they might be incompatible with each other. But Shang et al. put forward an alternative hypothesis: if neocentromere nucleation indeed negatively affects transcription, then due to the fact that the Z chromosome is present in only a single copy in DT40 cells, neocentromere formation on any essential gene would impair its expression and could lead to cell death. As such, surviving clones would only have neocentromeres that formed in non-expressed regions of the Z chromosome, where they do not compromise essential gene expression.

To test this, Shang et al. engineered chr. 5 (which is present in two copies in DT40 cells) for centromere excision, in the same way as chr. Z. After excision of the original centromere, they analyzed the genomic distribution of neocentromeres on chromosome 5. Neocentromeres thus generated were found to map both on top of active as well as inactive genes (fig. 2.2D). This strongly suggested that neocentromere formation is not impaired by local active transcription. Shang et al. have not yet confirmed whether the presence of underlying genes on these neocentromeres caused any impact on their mitotic performance. But nonetheless, these results suggested that it is not the local chromatin environment that
determines neocentromere location, but rather the effect of centromeres on underlying transcription that selects out incompatible locations.

**Strategy for in-depth characterization of chicken centrochromatin**

Given that chicken (neo)centromeres had shown themselves amenable to in-depth characterization due to their ability to be fully mapped, especially at the chromatin and transcript level, I decided to use this system to further characterize the nature of centrochromatin, in a vertebrate system. The pipeline to achieve this goal consisted of systematic mapping of several histone modifications, as well as transcripts, across the non-repetitive centromeres of chromosome Z and 5, both in asynchronous exponentially growing cultures (as a surrogate for the interphase state) and in cells arrested in prometaphase (as a surrogate for the mitotic state). Detection of chromatin marks was performed by using the same ChIP mapping approach as before (fig. 2.2A). Additionally, the TSS and a downstream intron region of the SMC2 gene were used as active gene control regions. Identification of novel centromere-associated chromatin marks (and distributions thereof) would then be followed up by high-resolution mapping by ChIP-seq. What follows are my efforts towards the completion of this goal.

"Gaps" in CENP-A distribution show reduced H3 levels, but not H4

The observation of deep 'gaps' in CENP-A distribution, and the fact that most mapped chicken centromeres possess one or more of these gaps\(^{283}\) led me to hypothesize that they might present evidence for centromere-specific effects occurring at the chromatin level. One possibility is that they are nucleosome-free regions (devoid of either H3 or CENP-A nucleosomes), like those found at active gene TSSs\(^{105}\), which could potentially be evidence of centromeric transcription. On the other hand, CENP-A "gaps" could be due to incompatibility of some local genomic feature with CENP-A assembly. These could be
nucleosomes which for some reason cannot be replaced by CENP-A, or conversely the locus induces constitutive CENP-A turnover.

To analyze nucleosome density across chicken centromeres, I analyzed the distribution of both histone H3 and H4 across the centromeres of chr. Z and 5 in #Z3 DT40 cells (fig. 2.3). H3 mapping on both centromeres revealed decreased H3 levels along the whole mapped CENP-A domain, in line with its replacement by CENP-A in centromeric nucleosomes. Surprisingly, the "gap" region of chr. Z centromere showed greatly reduced H3 levels, comparable to those at the SMC2 TSS. This was not significant for the "gap" in the centromere of chr. 5. This observation initially suggested that "gap" loci may have reduced nucleosome occupancy. However, analysis of H4 shows a more uniform distribution across the whole region, without any significant reduction at the gap sites (fig. 2.3).

To determine whether the centromere nucleosome distribution changed when the chromosome is in a mitotic state, I arrested #Z3 DT40 cells for 13h in prometaphase with nocodazole, before collection for ChIP. Under these conditions, although H4 levels don't change significantly, there is an overall reduction in H3 levels which stretches to the centromere-adjacent regions (of both centromere Z and 5) (fig. 2.3). H3 levels at "gap" loci of chr. 5 were still not significantly different from the "core" locus (fig. 2.3). It is possible that the H3 levels at these loci have no correlation with the observed lack of CENP-A mapped by ChIP-seq. On the other hand, my choice of primers for this locus may not have been as close to the observed "gap" by CENP-A as those for Chr. Z "gap", thus explaining the differences between the two "gap" loci at these centromeres.

In summary, I have shown that across the centromeres of chr. 5 and Z, H3 levels may have a cell-cycle-dependent variation that is not reflected in H4. Additionally, the severely reduced signal for both CENP-A and H3 in the "gap" loci of the chr. Z centromeres suggests that the locus may have particular characteristics previously unknown at vertebrate centromeres. Further analysis will be required to understand histone levels and distribution in these regions, whether they be artefacts or evidence of real biological processes.
Fig 2.3 – ChIP analysis of histone distribution on centromeres of chr. Z and 5 shows decreased H3 density at 'gap' regions

ChIP mapping of centromeres of chr. Z and 5, using the same primer design strategy as before (fig. 2.2A). Primer locations (red arrows) relative to CENP-A ChIP-seq mapping (in dark grey) are shown at the top. Two regions in the active SMC2 gene were used as controls for transcribed chromatin: the TSS and an intron region in the gene body. Mouse normal IgG was used to assess pulldown specificity and enrichment. Antibodies against total H3 and H4 were used to assess nucleosome and specific histone density across the analyzed regions. DT40 #Z3 samples from both asynchronous exponentially growing cultures (grey) and cultures arrested for 13h in microtubule depolymerizing agent TN-16 (red) were tested. Mean of three independent experiments (except for H4 measurements in mitotically arrested cells, where n=1), error bars denote SEM. Blue arrows indicate 'gap' loci.

Student's t test was used to compare individual loci measurements of H3 levels. Chr Z asynchronous "core" vs "gap" *** p=0.0005, Mitotic arrest "core" vs "gap" n.s. p=0.4542. Chr 5 asynchronous "core" vs "gap" n.s. p=0.9180, Mitotic arrest "core" vs "gap" n.s. p=0.3045.

Two-way ANOVA was used to compare variation of H3 or H4 levels across all the centromere loci, for each chromosome, between asynchronous or mitotically arrested cells. Chr. Z H3 levels Asynchronous vs Mitotic Arrest: ** p=0.0036, H4 levels Asynchronous vs Mitotic Arrest: n.s. p=0.1453. Chr. 5 H3 levels Asynchronous vs Mitotic Arrest: *** p=0.0002, H4 levels Asynchronous vs Mitotic Arrest: n.s. p=0.3791.
<table>
<thead>
<tr>
<th>SMC2 active gene control</th>
<th>( \text{CENP-A} ) mapping by ChIP-seq</th>
<th>% INPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asynchronous</td>
<td>Mitotic arrest</td>
<td></td>
</tr>
</tbody>
</table>

**H3**

- **Variation across centromere:**
  - \( p = 0.0436 \)
  - \( **p = 0.0002 \)

**H4**

- **Variation across centromere:**
  - \( n.s. \ p = 0.1453 \)
  - \( n.s. \ p = 0.3791 \)
Chicken centromeres Z and 5 are not enriched for canonical marks of active chromatin, but do show signs of open chromatin

To probe for chromatin marks related to transcriptional activity on the centromeres of chr. Z and 5, I immunoprecipitated chromatin marks H3K4me2 and H3K36me2, using the same ChIP approach as described before, in both asynchronous and mitotically arrested cells.

I observed no specific centromeric enrichment for H3K4me2 or H3K36me2 on either the Z and 5 centromeres, in contrast with levels at the control active SMC2 gene (fig. 2.4). H3K4me2 and H3K36me2 levels on these centromeres also remain low in mitotically arrested cells (fig. 2.4), consistent with previous observations in stretched chromatin fibers and contrasting with observations in human cells. This may suggest that, at least for chicken non-repetitive centromeres, a chromatin signature similar to active genes is absent, unlike in human and D. melanogaster centromatin.

It is known that transcription can still occur without broad H3K4me2 or H3K36me2 distribution across the expressed region, which is only prevalent in higher expressing genes, but active TSSs always seem to be enriched with H3K4me3 in human cells (see "6.1 Chromatin state signatures"). Additionally, if the observed H3 depletion at "gap" loci indeed indicates lower nucleosome density, such a feature could be evidence of active TSSs, which are mostly depleted from nucleosomes. In order to test whether these centromeres contain a TSS proximal to the primer locations chosen, I performed ChIP for H3K4me3. I observed no enrichment for this mark at any of the centromere loci tested, for either chr. Z and 5, both in asynchronous or mitotically arrested cells, when compared to the high levels found at the TSS of SMC2 (fig. 2.4). Even in the "gap" regions, no significant H3K4me3 levels are observed. Therefore, no evidence of any centromeric TSS can be confirmed at this level of ChIP resolution.

To further probe for any association of chicken centromeres with an active transcription chromatin state, I decided to analyze additional markers associated with active transcription, by ChIP. H4K16ac mostly correlates with open chromatin, and
H3K79me2 \(^{465-467}\) is a more ubiquitous chromatin mark than H3K4me3/2 for transcribed regions, including lower-expressing ones \(^{92,93}\) (see "6.1 Chromatin state signatures").

Surprisingly, I found an enrichment of H4K16ac around both the centromeres of chr. Z and 5, comparable to that found at the SMC2 gene (fig. 2.4), although the centromere core itself shows lower enrichment. Surprisingly, the "gap" locus of chr. Z shows significantly less enrichment than surrounding loci (fig. 2.4), in spite the fact that H4 levels throughout all these loci being quite similar (fig. 2.3). This modification is greatly reduced upon entry into mitosis, an observation that had been reported previously in mouse cells, where this mark is globally reduced on chromosomes during mitosis \(^{468}\). Surprisingly, my results show that the SMC2 gene retained significant enrichment during mitosis, both at the promoter and gene body (fig. 2.4). Possibly not all loci (especially active genes) lose this mark during mitosis.

H3K79me2 shows a clear enrichment around centromeres in mitotically arrested cells by ChIP, but not in asynchronous cells (fig. 2.4). The distribution of this mark has been shown to mirror that of transcriptional activity and its enrichment on the chromosome increases throughout the cell cycle \(^{465,466,469}\). The mitotic levels of H3K79me2 across the centromere core were not very pronounced (fig. 2.4), which may indicate that transcriptional activity is more prevalent in centromere-adjacent regions than at centromeres themselves. Interestingly, I could not observe high signals for H3K79me2 on either locus of the SMC2 gene (which is actively transcribed), compared to that of centromeres. This might mean that mitotic centromeres have very high levels of H3K79me2 (compared to other actively transcribed regions). Additional control regions, such as an intergenic non-transcribed region or a silenced region, will be required to validate these findings.

My results suggest that the vicinity of centromeres of chromosomes Z and 5 of chicken may be in an open chromatin state during interphase, or even undergo active transcription. Further testing of the temporal distribution of these marks, both on the centromere and genome-wide, will be required for correct assessment of these observations.
Fig 2.4 – ChIP analysis of chr. Z and 5 for active transcription chromatin marks shows deficiency of most canonical chromatin marks, but presents some evidence of a transcriptional signature

ChIP mapping for active transcription marks on centromeres of chr. Z and 5, using the same analysis strategy as described in fig. 2.3. Pulldown specificity and enrichment controls, using mouse normal IgG, is shown in fig. 2.3. Antibodies against marks associated with actively transcribing chromatin were used, as follows: H3K4me2 tracks 5’ region of active genes; H3K36me2 tracks throughout gene bodies; H3K4me3 is highly enriched at TSSs and weakly at enhancers; H4K16ac is enriched in open chromatin and H3K79me2 tracks with transcription level. DT40 #Z3 samples were collected from either asynchronous exponentially growing cultures (grey) or cultures arrested for 13h in microtubule depolymerizing agent TN-16 (red). Mean of three independent experiments (except for measurements of the marks H3K4me3, H4K16ac and H3K79me2, in mitotically arrested cells and asynchronous quantifications of chr. 5 centromere, where n=1), error bars denote SEM. Statistical tests for differences between asynchronous and mitotically arrested cells were not performed, due to n=1 for the mitotically arrested condition. Mann-Whitney U test was used to compare H4K16ac values, in asynchronous cells, between SMC2 intron and chr. Z Adj L (n.s. p=0.4), SMC2 intron and chr. Z Edge L (n.s. p=1.0). Student’s t test was used to compare H4K16ac values, in asynchronous cells, between chr. Z Edge L and chr. Z Gap (* p=0.0466), chr. Z Gap and chr. Z Edge R (* p=0.0236), chr. Z Core and Chr. Z Gap (n.s. p=0.1115).
Non-repetitive chicken centromeres are heterochromatin-poor, but instead are enriched in polycomb chromatin

Given that the centromeres of chromosome Z, 5 and 27 are non-repetitive, Shang et al decided to confirm if they were also of a heterochromatic nature. ChIP-seq. for H3K9me2 revealed there is actually little enrichment for this mark in any of those 3 centromeres (fig. 2.5A). Their subsequent analysis of the neocentromeres of clones #BM23 and #0514 yielded the same observation (fig. 2.5A), that these centromeres were active and stable in the absence of the heterochromatin mark H3K9me3. This apparent absence of centromere-associated heterochromatin could be visualized by immunofluorescence microscopy, showing that, while most other chicken chromosomes contained an accumulation of H3K9me3 in the vicinity of their centromeres, enrichment of this mark at the centromere of chromosome Z could not be detected (fig. 2.5B).

Interestingly, these observations suggested that association of heterochromatin with centromeres may only be due to the fact that they are located within repetitive regions, and additionally that centromere maintenance in chicken cells does not necessarily depend on heterochromatin. Together with evidence from previous studies showing that centromeres in several eukaryotes seem sensitive to excessive transcription, I hypothesized that maybe an alternative type of repressive chromatin might be surrounding these centromeres in chicken. Previous studies suggesting that human neocentromeres can form in heterochromatin-poor regions did not address whether such locations were nonetheless under transcriptional repression, possibly via other chromatin states.

I decided to test the centromeres of chromosomes Z and 5 for the presence of the polycomb mark H3K27me3. Immunofluorescence staining revealed that the (peri)centromeric region of the Z chromosome is indeed enriched for H3K27me3, while showing a clear absence of H3K9me3 (fig. 2.5C). Further microscopy examination confirmed presence of H3K27me3 adjacent or co-localizing with CENP-A at the Z chromosome (fig. 2.5D). ChIP analysis of these two chromatin marks showed that there may
be a low-level accumulation of H3K9me3 at centromeres Z and 5, and H3K27me3 is significantly enriched on the regions surrounding the CENP-A core in both centromeres, compared to the control SMC2 active gene (fig. 2.5E). The levels of H3K9me3 observed by ChIP contrast with both immunofluorescence (fig. 2.5B,C) and ChIP-seq (fig. 2.5A), but the reason might be due to the absence of a bona fide heterochromatin control locus (which would be much more enriched with H3K9me3), but I was unable to locate such a locus to date. ChIP quantification also showed a reduction in the levels of both of these marks, particularly H3K27me3, in mitotically arrested cells. This contrasts with my immunofluorescence observations of the centromere of chr. Z in mitosis. One likely explanation is that this reduction may be a consequence of mitotic phosphorylation of the H3S10 and H3S28 residues, adjacent to H3K9 and H3K27 respectively, carried out by Aurora B at mitosis onset. These phosphorylations are known to occlude the epitopes for both of these antibodies (Hiroshi Kimura personal communication), however they still seem able to detect significant H3K9me3 and H3K27me3 by immunofluorescence throughout mitotic chromosomes. Alternatively, there may be mitosis-specific dynamics in the levels of these marks that I was not able to detect in my microscopy analysis.

In conclusion, I was able to show that the non-repetitive pericentromeres of chr. Z and 5, while being poorly enriched for heterochromatin, are in fact associated with polycomb marks. This not only supports that the heterochromatin state per se may not be essential for pericentromere function in chicken cells, as these are fully functional native centromeres, but suggests that a repressed chromatin state (be it heterochromatin or otherwise) is what contributes centromere function.
Fig 2.5 – The centromeres of chromosomes Z and 5 are poor for heterochromatin H3K9me3 mark, but enriched in polycomb H3K27me3 mark

A. ChIP-seq mapping of centromeres of chr. Z, 5 and 27, and neocentromeres on chr. Z of clones #BM23 and #0514, using antibodies against CENP-A and heterochromatin mark H3K9me3. By Shang et al 2013. H3K9me3 levels in and around selected centromeres are quite low (axis relative to bona fide heterochromatin-rich regions, not shown).

B. Immunofluorescence microscopy analysis of DT40 karyotype shows lack of heterochromatin on chr. Z. By Shang et al 2013. Antibodies were used against heterochromatin mark H3K9me3 (red), and CENP-T (green) to detect centromeres. Most chromosomes have an enrichment of H3K9me3 at the primary constriction, but that is not the case for chr. Z. A LacO array present in the Z chromosome was used, targeted by GFP-LacI (green), for ease of detection of this chromosome. DNA was stained with DAPI (blue).

C. The chromosome Z centromere is enriched in H3K27me3 instead of H3K9me3. Immunofluorescence microscopy analysis of DT40 #Z3 karyotype, by KCM chromosome spread, using antibodies against H3K27me3 (green) and H3K9me3 (red). DNA was stained with Hoescht 333342 (blue). Inset: chr. Z with arrow pointing to primary constriction. Scale bar = 5 µm.

D. Immunofluorescent analysis of Z chromosome shows enrichment of polycomb at the primary constriction. DT40 #Z3 chromosome spreads as in C were stained with antibodies against H3K27me3 (green) and CENP-A (red) for centromere localization. DNA was stained with Hoescht 333342 (blue). Arrows indicate centromere location.

E. ChIP mapping for heterochromatin and polycomb marks on centromeres of chr. Z and 5, using the same analysis strategy as described in fig. 2.3. Pulldown specificity and enrichment controls, using mouse normal IgG, is shown in fig. 2.3. Antibodies against heterochromatin mark H3K9me3 and polycomb marks H3K27me3 were used. DT40 #Z3 samples from both asynchronous exponentially growing cultures (grey) and cultures arrested for 13h in microtubule depolymerizing agent TN-16 (red) were tested. Mean of three independent experiments (except for H3K27me3 measurements in mitotically arrested cells, as well as asynchronous cells for chr. 5, where n=1), error bars denote SEM. Two-way ANOVA was used to compare variation of H3K9me3 or H3K27me3 levels across all the centromere loci, for each chromosome, between asynchronous or mitotically arrested cells. Chr. Z H3K9me3 levels Asynchronous vs Mitotic Arrest: ** p=0.0011, H3K27me3 levels Asynchronous vs Mitotic Arrest: n.s. p=0.0573. Chr. 5 H3K9me3 levels Asynchronous vs Mitotic Arrest: ** p=0.0054, H3K27me3 levels Asynchronous vs Mitotic Arrest: ** p=0.0070.
3.1.3. Discussion

The discovery of sequenced, non-repetitive centromeres in an vertebrate organism tractable for genomic engineering has allowed for both the conditional generation of neocentromeres, and the in-depth chromatin and transcriptional mapping of those same centromeres and neocentromeres. This valuable resource has allowed not only for the discovery that neocentromeres have no obvious sequence requirements for where they can form on the chicken genome, but also that their presence impairs transcriptional activity. Additionally, we were able to show that presence of CENP-A is compatible with open transcribing chromatin, but not on TSS regions. These conclusions have been reported in Shang et al, 2012.

Histone "gaps": a nucleosome-depleted region in centromere cores?

In most of the centromeres mapped in this study, at least one deep gap in the CENP-A distribution could be observed in each. Analysis of histone density, on the chicken centromeres of chromosome Z and 5, revealed that H3 levels across these "gap" loci were apparently quite reduced. While this initially suggested that nucleosome density in this region was low, the distribution of H4 across the same region showed no observable reduction (although H4K16ac levels were significantly reduced in this region in comparison with other centromeric loci). The observed reduction in H3 (despite no change in H4) could be evidence for a transcriptional process, such as the presence of a TSS (the SMC2 TSS also showed less enrichment when compared to a region downstream in the same gene), but current evidence is not sufficient to support this. On the other hand, loss of H3 epitope accessibility could also account for my observations. This could be a result of a distinct orientation and positioning of nucleosomes, or on the other hand the binding of specific proteins, either kinetochore proteins or any other protein recruited by some process triggered by the underlying DNA sequence (such as a mobile genetic element).
At this time, the sequences corresponding to the "gaps" in CENP-A distribution have not yet been studied. While there was no significant similarity among repeat elements present in the sample of neocentromeres isolated, such elements are nonetheless present in many of them, albeit with no conservation between them. It is possible that promoters of unspecific mobile elements, present in the DNA within the centromere core, could be providing some level of transcriptional activity, as observed in some human neocentromeres.

An elusive transcriptional "footprint" in chicken centromeres?

By virtue of being able to study vertebrate centrochromatin at higher resolution than previously thought possible, one hypothesis I sought to test was whether I could find evidence for transcription-related chromatin marks in these centromeres, as observed in human cells. Like previous studies of chicken centrochromatin, I was unable to find any evidence of the canonical transcription marks H3K4me2 and H3K36me2, either on neocentromeres or on non-repetitive centromeres of chromosomes Z and 5. Furthermore, ChIP mapping for H3K4me3, which could potentially allow me to pinpoint the existence of TSS-like loci, also gave no positive results for the qPCR loci selected. This included the "gap" loci, the best candidates for TSS-like entities at centromeres. However, because conventional ChIP allows only stringent analysis of chosen primer locations, such low coverage could allow TSS-associated peaks of H3K4me3, occupying only a very short region on the chromosome, to be missed. Thus, while there is little evidence to indicate there is transcription initiation within centromere cores, my mapping has not excluded it completely.

Interestingly, despite ChIP for H3K4me3/2 and H3K36me2 having provided no evidence any sort of such a transcription-associated signature at chicken centromeres, mapping of H4K16ac and H3K79me2 showed distinct enrichment patterns around the centromere core in chromosomes Z and 5. H3K79me2 co-localizes with transcriptional
activity and increases over the cell cycle up to mitosis \(^{465,466,469}\), and H4K16ac maps to open chromatin, but is reported to be absent from chromatin in mitosis \(^{71,100,101,464,468}\). Mapping of these histone marks, in asynchronous or mitotically-arrested DT40 cells, matched their expected cell cycle distribution. H4K16ac and H3K79me2 are functionally associated with each other \(^{476,477}\), and their distribution across the centromeres studied is similar, but at first glance contradictory: H4K16ac is absent in mitosis, whereas H3K79me2 only shows enrichment in mitosis. One interpretation could be that, while H4K16ac (correlating with open chromatin) is present at centromeres throughout interphase, H3K79me2 accumulation rises slowly during the cell cycle (reaching highest levels at mitosis), thus being mostly non-enriched in the majority of my asynchronous population. Subsequently, upon mitotic entry, H4K16ac is lost, while the more stable H3K79me2 methylation modification, after achieving maximum enrichment levels, remains on centromeres.

Along with its association with transcription, H3K79me2 has been demonstrated to confer anti-silencing properties, by antagonizing the actions of histone deacetylases \(^{476,478–481}\). H3K79 methylation in \(S.\cerevisiae\) has also been shown to be involved in DNA repair pathways \(^{482,483}\), and is required in DNA repair via sister chromatid recombination, in a process that recruits and requires cohesin \(^{484}\).

In light of these correlations, and given that I observe little evidence for canonical transcription marks, it is not impossible that I have found evidence of open chromatin that nonetheless does not necessarily implicate active transcription (such as DNA repair). Centromere-associated proteins, such as CENP-S/X \(^{250,251,485,486}\), HJURP \(^{487}\), and even CENP-A itself \(^{488}\), have been reported to be associated with the DNA repair machinery, and open chromatin is a requirement for initiating DNA repair \(^{489–491}\). As such, discriminating the exact relevance and meaning of the histone marks analyzed here, especially whether they necessarily implicate transcription or not, is paramount. Furthermore, it is also possible that different processes that promote open chromatin (or confer anti-silencing), regardless of how they achieve it, are able to promote centromere function. In view of this, it might not be
unexpected that such processes could vary from centromere to centromere, depending on their chromatin environment. This might help to explain the seeming disparity of findings when it comes to uncovering a consistent centrochromatin signature. Evidence for such, so far, is lacking, and transcription is to date the best candidate for such a role.

**Centromeres show persistent association with a repressive chromatin state, but which type it is may be flexible**

Chicken centromeres located outside of tandem repeats are poor in heterochromatin marks, but the fact that they may be enriched with other types of repressive chromatin, such as polycomb, could be evidence for a common link between centromeres and silenced chromatin. On the other hand, it may be an artefact of the propensity of neocentromere formation in intergenic regions, in order to avoid selective pressure from disrupting gene expression by neocentromere formation. Intergenic regions may be naturally enriched with repressive chromatin marks, given the absence of transcription. In either case, the recurring observation of centromeres in stretches of repressive chromatin is remarkable, especially when centromeres have been suggested to require active transcription to retain their function.

Heterochromatin and polycomb have been found to be mostly mutually exclusive, across large regions, on metazoan chromosomes \(^91,94,492–494\), although some interactions between the two proteins networks have been demonstrated \(^495–497\). Polycomb has been reported to "take over" previously heterochromatic regions in Suv39h1/h2 mouse knock-outs, which are defective for heterochromatin formation \(^126,493,498,499\). The same study showed that kinetochores can function with "polycomb pericentromeres", albeit with poor segregation fidelity. The fact that I was able to find enrichment for polycomb marks surrounding chicken centromeres, after several reports of heterochromatin-poor centromeres, may be evidence that similar features of both chromatin states are contributing to centromere function. Interestingly, in *D. melanogaster*, polycomb has been reported to associate with
cohesin subunits and requires active cohesin to perform its silencing functions\textsuperscript{500}. Polycomb could then be contributing to pericentromeric cohesin enrichment, in the same way as heterochromatin has been suggested to.

Neocentromere formation in heterochromatin-poor regions has been reported before in \textit{S. pombe} and human cells\textsuperscript{\textdegree 391,471,473}, and comparative genomics with other avians has suggested that the centromeres of chromosomes Z and 5 of chicken are evolutionarily recent\textsuperscript{472,501}. This would fit with proposed models for centromere evolution\textsuperscript{\textdegree 394,397}, in that 'young' centromeres, or neocentromeres, have not yet acquired (through as yet unknown mechanisms) repetitive elements in their vicinity. In this model, the presence of heterochromatin would only occur when centromeres acquire repetitive sequences. Possibly, until such events take place, centromeres might associate with alternative repressive chromatin states, although the reason for why they would do so is still unclear.

\textit{Reconciling centromere-associated silenced chromatin with neocentromeres located on genes}

The persistent association of a repressed chromatin state with centrochromatin can be put into question by the work of Shang \textit{et al}, described previously, where they showed that neocentromere formation could occur on active genes, which could imply that a repressed state was altogether dispensable. To reconcile this, I put forward the following hypothesis: those genes may nonetheless harbour repressive modifications. It has been shown that many genes can bear repressive chromatin modifications throughout the gene body, downstream of the TSS\textsuperscript{502}. Furthermore, some genes can even be dependent on a repressive chromatin environment for their expression, as shown for chr. 4 of \textit{D. melanogaster}\textsuperscript{503–506}. It is also possible that those genes may present mono-allelic expression (in which one of the gene alleles is constitutively silenced\textsuperscript{507}), and that neocentromere formation occurred on the silenced allele. Interestingly, chromatin control of mono-allelic expression has been suggested to fall under the purview of polycomb chromatin, and its
interplay with the H3K36me3 mark \(^{508-511}\). It is unknown at this point what is the chromatin state of the neocentromere-associated genes on chr. 5. Therefore, it might be possible that, if these gene alleles are associated with some form of repressive chromatin, a link between transcriptional repression and centromeres might still exist.

### 3.1.4. Conclusion and outlook

My observations in chicken non-repetitive centromeres have enhanced our understanding of the chromatin environments surrounding naturally evolved vertebrate centromeres. Additionally, they provided further evidence for variability on possible chromatin mark determinants at centromeres, particularly in respect to those related to active transcription.

While I have shown evidence for polycomb chromatin associated with the two natural centromeres of chromosome 5 and Z, it remains to be tested whether neocentromeres also present accumulation of repressive marks, be they H3K9me3, H3K27me3 or other. Additionally, testing the relevance (if any) of polycomb chromatin on these centromeres is required to confirm whether it contributes in any way to centromere stability, or if on the other hand the correlation of neocentromeres with intergenic regions is derived simply due to their interference with active essential gene function.
3.2. Centromere activity can coexist with polycomb chromatin

3.2.1. Background

Centromeres seem to coexist with repressive chromatin environments while maintaining a low level of transcriptional activity, through mechanisms unknown. However, heterochromatinization of human centromeres, via tethering of Suv3-9h1, HP1 or KAP-1, can lead to their inactivation through loss of CCAN proteins and CENP-A assembly. De novo centromere formation is also opposed by heterochromatin: by reversing the repression state (by knockdown of the heterochromatin pathway or promoting acetylation at satellite repeats), one can effectively promote de novo centromere formation in human cells. This suggests that centrochromatin may be incompatible with silencing.

Given that an actively transcribing chromatin state has been suggested as essential for centromere maintenance (in human cells and S. cerevisiae), one might assume that repression across centrochromatin must not be allowed to happen, if centromere function is to be preserved. Thus, it is intriguing that centromeres seem to be able to sustain themselves next to heterochromatin, which is known to spread to adjacent loci, yet they seem unable to tolerate assembly of heterochromatin directly across their whole region. Sullivan and Karpen first observed, in both Drosophila and human cells, that H3K9me3 coverage was absent from the CENP-A domain of the centromere, and some chromatin marks associated with active transcription co-localize with CENP-A domains in both these organisms and also in mouse cells.

These observations together hint at the existence of two separate functional domains on human α-satellite arrays: centrochromatin (CENP-A rich and transcribed) and the pericentromere (heterochromatic and repressed). Furthermore, they imply that the
mechanisms that allow centromeres to cope with neighbouring heterochromatin do not seem sufficient to withstand direct heterochromatin nucleation. Absence of heterochromatin enrichment over the centromere core was first shown in *S. pombe*\textsuperscript{342,521}, and spreading of heterochromatin to the CENP-A region correlates with centromere function becoming impaired. In this organism, centromere/pericentromere boundaries are maintained by tRNA genes flanking both sides of the centromere core, apparently preventing heterochromatin spreading\textsuperscript{342,345,346}. In most regional centromeres, however, there seem to be no known sequence determinants within centromere repeats that one could identify as constitutive heterochromatin barriers. Human CENP-A regions only occupy a fraction of the homogenous α-satellite array, the remainder being heterochromatin-rich. Heterochromatin covers also the adjacent degenerate satellite repeats of the pericentromere\textsuperscript{519}. As such, any boundary or anti-repression activity within centrochromatin, that can explain its resistance against transcriptional repression, is likely to be promoted by centrochromatin itself or by CCAN-specific proteins.

So far, manipulations of centrochromatin using the HAC system have demonstrated how centromeres respond to shifts into repressed states (by nucleating heterochromatin, or removal of active transcription marks) or open chromatin states (by increased histone acetylation, or highly amplified levels of transcription)\textsuperscript{294,339,405,407,408}. In an effort to further understand the centromere's relationship with repressive chromatin, I decided to test its interaction with polycomb silencing\textsuperscript{150}. Polycomb proteins repress transcription through slightly different mechanisms than heterochromatin. These include ubiquitination of H2AK119 (which blocks RNA polymerase II transit) and topological chromatin compaction\textsuperscript{157–161}. Through this induced repressive mechanism, I sought to explore whether centromere transcription could be affected by changing chromatin state without forming classical heterochromatin, possibly by affecting RNA polymerase directly. Polycomb has been shown to be able to coexist with active chromatin marks\textsuperscript{159,522–524}. Additionally, polycomb
chromatin does not completely abolish active centromere functions. Knockout mice for Suv39h1/h2, which lack heterochromatin, show an encroaching of polycomb repression over previously heterochromatic regions, yet centromere identity (if not their full function) is still preserved. Cells of these mice presented chromosome mis-segregation defects which led to frequent tumour formation and low fertility, yet the animals were still viable \(^99\). Together with my finding that chicken centromeres can be enriched in polycomb marks (in section "3.1. Results: Mapping of non-repetitive chicken centromeres"), this raises interesting questions about how centrochromatin interacts with different types of adjacent repressive chromatin states, and what mechanisms it uses to coexist with them. Particularly given that active centromere transcription may play a crucial role in its stability, which contrasts with the otherwise apparently repressive environment to transcription.

To promote formation of polycomb chromatin on the HAC centromere, I performed targeting of the H3K27 methyltransferase EZH2, part of the PRC2 repressive complex, in the hopes of finding new interactions between active centrochromatin and a new type of transcription repression other than heterochromatin. Depending on the response of the HAC centromere, this experiment should help me understand if centrochromatin is susceptible to this type of repression as well. The results of this experiment might help me predict what other regions of the genome might be suitable for neocentromere formation. Furthermore, any factors implicated in centrochromatin's relationship with repressive chromatin may be relevant for the maintenance of any boundary function, if indeed such boundaries exist.

### 3.2.2. Results

**Study of centromere-associated repressive histone modifications**

I first sought to understand the native chromatin state of pericentromeres: how the chromatin marks of the two repression states (heterochromatin and polycomb) are arranged
at human (peri)centromeres, relative to each other, and how they distribute on and/or around the CENP-A domain. To achieve this, I performed immunofluorescence staining for chromatin marks H3K9me3 and H3K27me3 on human mitotic chromosomes, and stretched mitotic chromatin fibers.

I observed that H3K9me3 is enriched at most pericentromeric regions on mitotic chromosomes, but close observation of stretched chromatin fibers showed this mark indeed has lower enrichment across the CENP-A domain, or is even altogether absent there (fig. 3.1A), as reported previously in human and D. melanogaster cells. Distribution of H3K9me3 and H3K27me3 on chromosomes, which is mostly non-overlapping, showed that at least some pericentromeres are highly enriched in H3K27me3, instead of H3K9me3, whereas others present no enrichment for either mark (fig. 3.1B). Some stretched chromatin fibers show H3K27me3 signals adjacent or overlapping with CENP-A stretches, just as observed for H3K9me3, less dense or absent across the CENP-A domain (fig. 3.1C).

To quantitatively analyse the levels of these marks, I performed ChIP in the HeLa 1C7 cell line, which contains a HAC whose synthetic α-satellite TetO repeats are a composite based on those of chromosome 17 and a consensus α-satellite monomer containing TetO sites. Quantification of H3K9me3 and H3K27me3 shows that the HAC centromeric repeats are enriched with H3K9me3, similarly to the centromere of chr. 21 (taken as an endogenous centromere control). Yet both of these loci have a slightly lower H3K9me3 level than adjacent pericentromeric regions, such as the BAC region of the HAC (non-α-satellite) and the pericentromeric repeat Sat2 (fig. 3.1D). The constitutively active Bsr genes (located on the HAC) and PABPC1 (endogenous) show low levels of H3K9me3, as expected for non-repressed regions. H3K27me3 is detected at low levels on centromere arrays (of both the HAC and chr. 21), as reported previously, whereas its levels are negligible on active
genes Bsr and PABPC1 (fig. 3.1D). Importantly, the levels of H3K27me3 on centromere regions do not compare to those found on the *bona fide* polycomb-repressed gene MYT1, which are much higher. These results indicate that there could either be small regions of polycomb chromatin within the α-satellite arrays of the HAC and chr. 21, or a low-level H3K27me3 accumulation across those regions. In either case, the polycomb chromatin state is not enriched at these centromeres.

In summary, my mapping of chromosomes shows that the type of repressive chromatin at pericentromeres can be quite variable. Centromeres can be found adjacent to either heterochromatin or polycomb chromatin, or even neither of those. The two marks seem to not overlap with each other, and their distribution tends to be less dense or even absent at centromere cores.
**Fig 3.1 – Distribution of repressive chromatin marks H3K9me3 and H3K27me3 at (peri)centromeres**

A. H3K9me3 is less enriched across CENP-A domains. Mitotic chromosomes of HT1080 cells, stained with antibodies against H3K9me3 (red) and CENP-A (green), and DNA stained with hoescht 333342 (blue). Top panels show example of a chromosome spread, lower panels show examples of stretched chromatin fibers. Scale bar = 2µm.

B. H3K27me3 is enriched at some pericentromeres. Mitotic chromosomes of HeLa 1C7 cells, were stained with antibodies against H3K9me3 (red) and H3K27me3 (green), and DNA stained with hoescht 333342 (blue). Insets show individual chromosomes, arrows indicate the primary constriction (centromere). Scale bar = 5µm.

C. H3K27me3 can localize adjacent to, but is less enriched at, CENP-A domains. Mitotic stretched chromatin fibers of HeLa 1C7 cells, stained with antibodies against CENP-A (red) and H3K27me3 (green), and DNA stained with hoescht 333342 (blue). Scale bar = 2µm.

D. Chromatin analysis by ChIP of HeLa 1C7 cells shows that HAC and chr.21 (peri)centromeres are primarily heterochromatic. Mouse antibodies against H3K9me3 and H3K27me3 were used. Mouse unspecific IgG was used as negative pulldown control. Pulldown DNA was quantified by qPCR. Primer sets used are shown at the top. Mean of three independent experiments, error bars denote SD.
Tethering of EZH2 on the HAC establishes polycomb markers and reduces centromere proteins

To understand if centromeres can withstand direct polycomb repression, I used the H3K27 methyltransferase EZH2, a subunit of the polycomb PRC2 complex, to nucleate H3K27me3 on the HAC. As a control for methyltransferase activity versus any contribution by its interacting partners, I generated a catalytically inactive mutant of the enzyme, EZH2 H689A. Both were cloned into a HAC-targeting vector, generating constructs TetR-EYFP-EZH2 and TetR-EYFP-EZH2-H689A, which were transiently transfected into Hela 1C7 HAC cells.

At 3 days after transfection, I observed an enrichment for H3K27me3 at the HAC 3 days after transfection, as well as an accumulation of RING1A (fig. 3.2A), one of the subunits of the PRC1 polycomb complex, which is recruited by the H3K27me3 mark. Thus, TetR-EYFP-EZH2 tethering recapitulates the chromatin events that lead to PRC1 recruitment. Consistent with this, the HAC-associated active transcription mark, H3K4me2, was visibly decreased (fig. 3.2B, C), supporting a conversion of the chromatin state at the HAC from active to repressed.

The EZH2 H689A catalytically inactive mutant, however, also induced enrichment of H3K27me3 and loss of H3K4me2 in most cells. However, only a few cells showed RING1B recruitment (fig. 3.2A, B, C). It seems that, although not as efficiently as the WT protein, EZH2 H689A still nucleates the polycomb chromatin state. It was reported that EZH2 H689A was defective for H3K27 methyltransferase activity. It is possible that transfected TetR-EYFP-EZH2-H689A recruits endogenous WT EZH2 to the HAC, through the EED and Pcl2 subunits of PRC2. It therefore does not seem to be a good control for EZH2 methyltransferase activity.

Analysis of HAC centromere proteins shows a visible reduction of CENP-T upon EZH2 tethering after 3 days (fig. 3.2B, D), suggesting that this change in chromatin state affects the HAC centromere. Again EZH2 H689A also causes the same effect (fig. 3.2B, D),
demonstrating that it indeed cannot be used as a negative control for H3K27me3-specific effects on the HAC. Characterization of centromere proteins CENP-A and CENP-C showed that EZH2 has a milder effect on these components, but together with the reduction in CENP-T supports that the HAC centromere is negatively affected by EZH2 (fig. 3.2E). This indicates that centrochromatin is sensitive to the polycomb repressive state.

Together, these results suggest that centrochromatin may be as sensitive to EZH2 and polycomb repression as it is to heterochromatin-related factors, such as HP1\textsuperscript{407} or Suv39h1\textsuperscript{204}. 
Fig 3.2 – EZH2 tethering to the HAC recapitulates the polycomb silencing pathway and reduces levels of centromere proteins

A. EZH2 tethering to the HAC nucleates H3K27me3 and recruits RING1A. HeLa 1C7 cells were transiently transfected for 3 days with the TetR-EYFP constructs shown (green), and processed for immunofluorescence using antibodies against H3K27me3 (red) and RING1A (blue), and DNA stained with hoescht 333342 (gray). Arrows and insets depict the HAC. Scale bar = 5µm.

B. EZH2 tethering to the HAC slightly affects CCAN proteins, but severely reduces H3K4me2. As in A, but stained with antibodies against CENP-T (red) and H3K4me2 (blue).

C. EZH2 tethering effectively reduces active chromatin mark H3K4me2 at the HAC. Qualitative assessment of H3K4me2 HAC enrichment, as shown in B. Single experiment, no. of cells analyzed: EYFP only n= 18, EZH2 n= 22, EZH2 H689A n= 17.

D. EZH2 tethering reduces CENP-T at the HAC. Quantification of CENP-T levels, as shown in B. Single experiment, cells analyzed n≥17 for each condition. Mann-Whitney U statistical test showed significant differences between results: EYFP vs EZH2 *p=0.0129, EYFP vs EZH2 H689A *p=0.0424.

E. EZH2 tethering can affect both CCAN and CENP-A levels. Quantification of CENP-C and CENP-A levels at the HAC, after 3 days of transient transfection with TetR-EYFP or TetR-EYFP-EZH2. Total of three independent experiments, each experiment CENP-A n≥19 cells, CENP-C n≥19 cells, blue line denotes median. Mann-Whitney U statistical test showed significant differences between results: CENP-A **p=0.0019, CENP-C *p=0.0139.
**Long-term tethering of EZH2 does not inactivate the HAC centromere, and has no effect on HAC mitotic fidelity**

To further analyze and quantify the effects of polycomb nucleation on the HAC centromere, I generated a stable HeLa 1C7 cell line expressing TetR-EYFP-EZH2 (subsequently referred as 1C7-EZH2). These cells were grown in doxycycline-supplemented medium to prevent untimely tethering of EZH2 to the HAC. I then analyzed the levels of HAC centromere proteins, the fidelity of its mitotic segregation, and the effect of polycomb chromatin on CENP-A assembly, over a long-term time course of EZH2 tethering.

Surprisingly, by tethering EZH2 over a 12 day time-course, I observed a significant reduction of centromere proteins (CENP-C and CENP-T), but their levels did not decrease further beyond 5 days of tethering (fig. 3.3A, B, C). Following an initial significant drop in CENP-A levels, the protein showed no further change after 5 days of EZH2 tethering (fig. 3.3A, B). Furthermore, this analysis revealed no significant increase in HAC alignment or segregation defects, over the tethering period (fig. 3.3D). Thus the centromeric activity of the HAC was not compromised.

Given that EZH2 tethering reduces H3K4me2 and CENP-A (slightly) on the HAC, and that the tethering of the H3K4me2 demethylase LSD1 to the HAC was shown to be sufficient to impair CENP-A assembly\(^\text{405}\), I hypothesized that polycomb repression could also be causing CENP-A assembly to be reduced. To this hypothesis, I analysed the first event of CENP-A assembly under tethering conditions, by transiently co-transfecting 1C7 cells with a plasmid expressing CENP-A-SNAP and a second plasmid expressing one of the following TetR-EYFP fusion constructs: EYFP only, EZH2 or LSD1. After transfection, cells were arrested with thymidine in S-phase for 18h, then released and labelled with the non-fluorescent BTP SNAP substrate, which quenches all existing free SNAP-tags in the cell. Cells were then allowed to progress into G1 and undergo assembly of newly synthesized CENP-A-SNAP, which bears free SNAP-tags. Newly-synthesized centromeric CENP-A-SNAP was then specifically labelled with the fluorescent substrate TMR-Star\(^\text{286}\).
Measurement of TMR-Star levels at the HAC centromere revealed that while LSD1 significantly reduces CENP-A assembly (as shown previously in Bergmann et al 2010⁴⁰⁵), EZH2 only shows a lesser, non-significant reduction (fig. 3.3E). This supports that the chromatin changes introduced by EZH2 on the HAC are not sufficient to impair maintenance of centromere memory, despite causing a slight overall reduction of total CENP-A levels.

Thus, the HAC centromere can persist despite the presence of polycomb chromatin and suffers few negative effects on its mitotic behaviour or ability to be reassembled every cell cycle.
Fig 3.3 – Long-term EZH2 tethering has little effect on centromere proteins and mitotic fidelity

A. Long-term EZH2 tethering in 1C7-EZH2 cells reduces but does not deplete the HAC centromere. 1C7-EZH2 cells were grown in the absence of doxycycline for a 12-day time course, and at each time point samples were processed for immunofluorescence using antibodies against CENP-A, CENP-C and CENP-T. Total of two independent experiments for CENP-A and CENP-C quantifications, no. of cells analyzed: CENP-A n≥28 cells and CENP-C n≥24 cells each timepoint. Single experiment for CENP-T quantifications, no. of cell analyzed n≥20 cells each timepoint. Mann-Whitney U statistical test was used to evaluate significance between differences observed.

B. Microscopy analysis of CENP-A staining, at 3 and 8 days of EZH2 tethering to the HAC, as described in A. Antibodies against CENP-A (red) were used and DNA was stained with hoescht 333342 (blue).

C. The HAC CCAN persists during long-term EZH2 tethering, even during mitosis. Example images of CENP-C staining, at 3 and 12 days after EZH2 tethering, as described in A. Antibodies against CENP-C (red) were used and DNA was stained with hoescht 333342 (blue). Arrows and insets denote the HAC chromatids and associated CENP-C signal.

D. Tethering of EZH2 to the HAC has little effect on HAC mitotic fidelity. Quantification of HAC mitotic defects in 1C7-EZH2 cells, at each time point described in A. A HAC scored as positive for a mitotic defect if one of the following was true: it did not congress in metaphase, it lagged during anaphase, or chromatids mis-segregated during anaphase/telophase. Transient transfection of TetR-EYFP, for 3 days, into HeLa 1C7 cells, was used as a negative control. Mean of three independent experiments, n≥40 cells each timepoint, error bars denote SD. Fisher’s exact test was used to evaluate significance between differences observed: EYFP 3d vs EZH2 **p=0,0018, EZH2 1h vs 6d **p=0,0011, EYFP 3d vs EZH2 12d n.s. p=0,4167.

E. EZH2 tethering has little effect on CENP-A assembly. 1C7 cells were co-transfected with plasmids expressing CENP-A-SNAP and a TetR-EYFP construct, as shown, for 5h, and then arrested in S phase with thymidine. After 17h, cells were released and a 9h quench-chase-pulse for CENP-A-SNAP was performed, to specifically label newly synthesized CENP-A-SNAP with TMR-Star. TetR-EYFP was used as a negative control, and LSD1 as a positive control. Single experiment, n≥16 cells each condition. Mann-Whitney U statistical test showed no significant differences between results of EYFP vs EZH2, but did for EYFP vs LSD1: *p=0,0131.
**EZH2 tethering to the HAC generates a polycomb chromatin signature, but does not affect HAC centromere transcription**

As the HAC centromere seemed to be able to cope with polycomb marks, these experiments raise the question of how exactly EZH2 has changed HAC chromatin, and whether it inhibits centromere transcription. I therefore analyzed the chromatin state of EZH2-tethered HACs, and the level of transcripts emanating from their centromere.

ChIP quantification on 1C7-EZH2 cells after 5 days of EZH2 tethering revealed no change in CENP-A levels at the HAC α-satellite arrays. I observed a large enrichment of H3K27me3 at HAC α-TetO repeats, supporting effective establishment of a polycomb chromatin state at the HAC centromere. The difference in H3K27me3 status of HAC chromatin, before and after tethering, was also easily observable in chromosome spreads of 1C7-EZH2 cells (fig. 3.4B), and co-localized with the centromeric regions of each HAC chromatid. Additionally, ChIP analysis confirmed that active chromatin marks H3K4me2 and H3K36me2 were severely reduced across the HAC centromere repeat array (fig. 3.4A). H3K36me2, a mark demonstrated to be enriched across centromeric regions and associated with transcription elongation, has been reported to antagonize polycomb repression. Yet, my experiments demonstrate that presence of H3K36me2 on the chromatin substrate is not sufficient to resist nucleation of polycomb chromatin throughout the HAC (fig. 3.4A).

This experiment thus yielded the surprising result that EZH2 tethering to the HAC can efficiently dismiss most active transcription marks from the HAC centromere array, without loss of CENP-A or mitotic fidelity.

Given that active chromatin marks on the HAC are severely reduced, and that these marks have been shown to be required for transcriptional activity, I hypothesized that HAC centromere transcription might be repressed after polycomb establishment. However, quantification of HAC centromeric transcription, by RT-qPCR, before and after 3 days of
EZH2 tethering, showed no significant change (fig. 3.4C). Thus, in a surprising way, centromeric transcription was not impaired by nucleation of polycomb chromatin, despite a severe reduction in chromatin marks associated with active transcription. Tethering of LSD1 to the HAC was shown to achieve reduction of HAC centromere transcription within just 1 day. EZH2 tethering, on the other hand, showed no effect on transcription even after 3 days, when H3K4me2 levels already seem severely reduced (fig. 3.2B, C).

It might be argued that polycomb repression is unable to efficiently repress transcription in this region, possibly due to the underlying DNA sequence or the nature of the α-satellite repeats. Additionally, the efficiency of EZH2-dependent polycomb repression has been shown to be significantly affected by the local chromatin state. To demonstrate that tethering of TetR-EYFP-EZH2 to an α-TetO array can indeed cause efficient transcriptional repression, I transiently transfected TetR-EYFP-EZH2 into both HeLa 1C7 and HeLa 1F10 cells. 1F10 cells possess a similar α-satellite TetO array as the HAC, but it is present as an array integrated into an endogenous chromosome arm, and lacks centromere activity. 3 days after transfection, I extracted RNA from cells and analyzed transcripts by RT-qPCR. Whereas 1C7 cells showed a surprising increase in α-TetO transcription from the HAC following EZH2 tethering, 1F10 cells showed a significant decrease in α-TetO transcripts from its non-centromeric array (fig. 3.4D).

These results demonstrate that EZH2 tethering has the potential to induce transcriptional repression on an α-satelliteTetO array, and that resistance of the HAC to polycomb repression may be promoted by the presence of a centromere.
Fig 3.4 – EZH2 tethering can severely reduce active transcriptional marks on the HAC, but has not effect on HAC centromere transcription

A. EZH2 tethering severely reduces HAC H3K4me2 and H3K36me2, but has no effect on CENP-A. Chromatin analysis by ChIP of 1C7-EZH2 cells. Cells were washed free of doxycycline and grown for 5 days before harvesting and processing for ChIP. Cells grown in doxycycline were processed in parallel. Mouse antibodies against CENP-A, H3K27me3, H3K4me2, and H3K36me2 were used; mouse unspecific IgG was used as negative pulldown control. Pulldown DNA was quantified by qPCR. Primer sets used are shown below graphs. Mean of three independent experiments, error bars denote SEM. Student's t test was used to evaluate significance of differences observed.

B. Example of HAC H3K27me3 levels, before and after EZH2 tethering. Chromosome spreads of 1C7-EZH2 cells, grown in presence or absence of doxycycline, were stained with antibodies against CENP-C (red) and H3K27me3 (green), and DNA was stained with hoescht 33342 (blue). Arrows denote HAC chromatids.

C. EZH2 tethering does not affect levels of centromeric HAC transcripts. Quantification of transcripts from 1C7-EZH2 cells, grown in presence or absence of doxycycline. Cells were grown as in A, harvested for RNA extraction, cDNA was retrotranscribed and transcripts were quantified by qPCR. Expression level is normalized to genomic copy number (for repeats) and normalized to β-actin. Mean of three independent experiments, error bars denote SEM. Student's t test was used to evaluate significance of differences observed.

D. The HAC centromere protects α-satellite transcription from polycomb repression. HeLa 1C7 and HeLa 1F10 cells were transiently transfected with the depicted TetR-EYFP constructs for 3 days, and subsequently harvested for transcript quantification as in C. Untransfected cells were included as a negative control for TetR-EYFP binding. Mean of three independent experiments, error bars denote SD. Student's t test was used to evaluate significance of differences observed.
**HAC centromere resistance to polycomb is independent of PRC1 release during mitosis**

The fact that the HAC centromere can be enriched with polycomb repressive chromatin, but suffer no ill effects on its transcriptional output, suggests that mechanisms might be in place to allow centromeres to transcribe and bear some level of active transcription marks even within this chromatin context. One hypothesis is that the repressive mechanism might be cell-cycle regulated, and possible "windows of opportunity" for centromere transcription might exist. The repressive polycomb PRC1 complex has been shown to mostly delocalize from chromosomes during mitosis\(^{537-539}\). This is caused by its inability to bind H3K27me3 due to the mitotic phosphorylation of adjacent H3S28\(^{540,541}\), catalyzed by Aurora B. Aurora B phosphorylates several chromatin targets upon mitotic entry, including the H3S10 and H3S28 residues on nucleosomes\(^{542-544}\). Given that centromeres have been reported to transcribe during mitosis\(^{419}\), and the initial steps of CENP-A assembly start during late anaphase\(^{286}\), it is possible that this particular "window of opportunity" exists during mitosis, when PRC1 is prevented from binding, allowing centromeres to transcribe and sustain themselves while under polycomb repression.

To explore this hypothesis, I analyzed PRC1 localization at different mitotic stages, by immunofluorescence microscopy. Confirming previous observations\(^ {537-539}\), PRC1 enrichment could no longer be detected on the HAC upon entry into mitosis, and only started returning during telophase, reaching normal levels in early G1 (fig. 3.5A). HAC H3K27me3 enrichment also seems to be reduced during mitosis (fig. 3.5A). However, the antibody used in this staining does not bind the H3K27me3 epitope when H3S28ph is present (Hiroshi Kimura, personal communication). In any case, it seems that the HAC centromere does enjoy a period of time seemingly free of PRC1. This could potentially allow for transcription and/or other essential processes which allow centromere maintenance.
To test the hypothesis that relocation of the PRC1 repressor during mitosis allows the HAC centromere to subsist in the presence of polycomb repression, I treated 1C7-EZH2 cells with Aurora B inhibitor ZM447439. Addition of ZM447439 is expected to lead to loss of Aurora B-dependent phosphorylations, such as H3S10ph and H3S28ph, and I confirmed this by immunofluorescence (fig. 3.5B). Analysis of PRC1 distribution, by immunofluorescence staining of RING1A, revealed that Aurora B inhibition is sufficient to retain PRC1 at the HAC during mitosis (fig. 3.5B).

Since PRC1 can be retained at the HAC during mitosis, I proceeded to test whether in these conditions the HAC centromere was impaired. Given that centromere transcription seems to occur during mitosis, and CENP-A replenishment is initiated during mitotic exit, I examined the level of CENP-A assembly at the HAC, by transfecting CENP-A-SNAP, as described before, following retention of PRC1 during the preceding mitosis by Aurora B inhibition. Quantification of newly-assembled CENP-A at the HAC showed no significant difference with or without Aurora B inhibition (fig. 3.5C). Thus, PRC1 retention on HAC polycomb chromatin is not sufficient to impair CENP-A assembly. This behaviour was the same, whether measured for the first events of CENP-A assembly after EZH2 tethering (1 day) or after 4 days of polycomb establishment, when total CENP-A levels have already been slightly reduced (fig. 3.3A).
Fig 3.5 – Retention of PRC1 at HAC chromatin during mitosis is insufficient to affect CENP-A assembly

A. PRC1 relocates away from chromatin during mitosis. Microscopy images of mitotic 1C7-EZH2 cells, after 4 days of EZH2 tethering (green), were analysed for PRC1 signal at the HAC, as detected by its RING1A subunit, both at interphase and at various stages of mitosis. Antibodies against H3K27me3 (red) and RING1A (blue) were used for detection, and DNA was stained with hoescht 333342 (gray). Note that H3K27me3 signal is also shown as being reduced during mitosis, and it is known that my anti-H3K27me3 antibody suffers from lack of epitope accessibility during mitosis caused by occlusion by the adjacent H3S28p modification (Hiroshi Kimura, personal communication).

B. Inhibition of Aurora B activity leads to retention of PRC1 at the EZH2-tethered HAC. Microscopy images of mitotic 1C7-EZH2 cells, after 1 day of EZH2 tethering (green), in the presence or absence of Aurora B inhibitor ZM447439. Leftmost column: confirmation of the inhibition of Aurora B phosphorylation activity, antibodies against α-tubulin (red) and H3S10 phosphorylation (gray) were used, and DNA was stained with hoescht 333342 (blue). Rightmost columns: PRC1 is retained at the HAC during mitosis upon inhibition of Aurora B, antibodies against RING1A (red) were used to detect the PRC1 complex, and DNA was stained with hoescht 333342 (blue).

C. Schematic of the experimental conditions for D). Asynchronous cell populations were used, but only those who had visible CENP-A~SNAP signals at their centromeres, corresponding to those cells that had assembled CENP-A~SNAP during the experimental duration, were analyzed.

D. Effect of PRC1 retention at mitotic HACs tethered with EZH2, on CENP-A assembly. 1C7-EZH2 cells were grown in the absence of doxycycline, for a total of 1 or 4 days before fixation. 1 day before fixation a plasmid expressing CENP-A-SNAP was transfected into cells, and 9h before fixation a quench-chase-pulse assay was performed to specifically label newly-synthesized CENP-A-SNAP with TMR-Star. During the quench-chase-pulse period, cells were grown in parallel in medium supplemented either with DMSO or ZM447439. Total of two independent experiments, n≥23 cells for each condition. Mann-Whitney U statistical test was used to evaluate significance between differences observed. A reduction in CENP-A assembly could be noted between controls of day 1 and day 4, a loss of total assembly which is likely due to reduction of the centromere recruitment platform, in line with results in fig. 3.3A.
**A**

- EZH2
- H3K27me3
- RING1A
- + DNA

- EZH2
- H3K27me3
- RING1A
- + DNA

**B**

- EZH2 α-tubulin
- H3S10ph DNA
- + DMSO

- EZH2 α-tubulin
- H3S10ph DNA
- + ZM447439

**C**

Quantification of CENP-A loading (with retention of PRC1 during mitosis)

HeLa 1C7 ~ TetR-EYFP-EZH2

Stable cell line

Assynchronous cell population

Transflect plasmid expressing CENP-A–SNAP

Quench free CENP-A–SNAP with non-fluorescent BTP

Label newly-synthesized CENP-A–SNAP with fluorescent TMR-Star

Wash and process for microscopy analysis

Quantify CENP-A–SNAP HAC signal in early G1 cells

Doxycline washed out for a total of either 1 day or 4 days

**D**

Loading of newly-synthesized CENP-A–SNAP at HAC centromere

![Graph showing loading of CENP-A–SNAP at HAC centromere with comparison between 1 day and 4 days]
Together, these results show that release of PRC1 from HAC chromatin during mitosis is not the reason why the HAC centromere is resistant to polycomb repression. Given that LSD1 tethering, which reduces centromere transcription within 24h, can affect the first cycle of CENP-A assembly after tethering (fig. 3.3E)\textsuperscript{405}, one might gather that PRC1 mitotic retention probably has little to no effect on HAC centromere transcription as well, but this has not been tested.

*The HAC centromere shows little sensitivity to repressive “bookmarks”, but is affected by direct tethering of secondary repressors*

Having excluded that release of PRC1 during mitosis is not the reason why centromeres persist within polycomb repressive chromatin, and given that PRC1 is responsible for the majority of effective polycomb repression in mouse cells \textsuperscript{161}, I hypothesized that maybe centrochromatin itself was preventing establishment of effective PRC1 function at the CENP-A domain (fig. 3.6A). That is, although EZH2 activity can target PRC1 to the HAC α-satellite arrays, there may exist some aspect of the centrochromatin environment which prevents PRC1 from establishing a repressive state.

To explore this hypothesis, I examined the consequences of bypassing the H3K27me3 recruitment step, by direct tethering of PRC1, on centromere stability and function. I transiently transfected HeLa 1C7 cells for 3 days, with the following TetR-EYFP fusion constructs: EZH2 and BMI1 (a subunit of PRC1\textsuperscript{150}, fig. 3.6B), with TetR-EYFP as a negative control and KAP-1 as a positive control. KAP-1 is part of a strong silencing scaffold complex \textsuperscript{546,547} whose swift negative effects on the HAC CCAN and CENP-A have been previously studied \textsuperscript{408}. By quantifying the levels of CENP-A, which is the most persistent of centromere markers after inactivation by a chromatin repressor \textsuperscript{408}, I was able to compare the different effects these chromatin silencer complexes have on the HAC centromere.
Direct tethering of PRC1, through its BMI1 subunit (fig. 3.6D, E), caused an increase in HAC mis-segregation and achieved reduction of CENP-A, after 3 days, to levels comparable to that observed with KAP-1. On the other hand, as previously observed in 1C7-EZH2 cells, a much smaller reduction on HAC CENP-A and mitotic fidelity was detected upon EZH2 tethering (fig. 3.6D, E).

As such, I can conclude that direct tethering of PRC1, as opposed to its recruitment by EZH2-mediated H3K27me3, can significantly perturb the HAC centromere. In this way, I have showed that presence of the PRC1 complex on the HAC can indeed have negative effects on its centromere, indicating that it is not immune to perturbations induced by polycomb factors. But somehow these do not occur when PRC1 is brought to the HAC via H3K27me3, even considering that PRC1 is extensively recruited across the whole HAC array in those conditions (as observed by microscopy, fig. 3.2A, fig. 3.6B).
Fig 3.6 – HAC centromere resists repression by nucleation of repressive "bookmarks", but not by direct tethering of secondary repressor complexes

A. Diagram of the two-step pathway of both the polycomb and heterochromatin transcription repression systems.
B. TetR-EYFP-BMI1 can recruit PRC1 complex to the HAC. 1C7 cells were transiently transfected with the TetR-EYFP constructs shown for 3 days, fixed and processed for immunofluorescence. Antibodies against RING1A were used to detect enrichment of PRC1.
C. TetR-EYFP-Suv39h1-SET can nucleate H3K9me3 at the HAC. 1C7 cells were transfected, as in A, with the TetR-EYFP constructs shown. Antibodies against CENP-A (red) and H3K9me3 (blue) were used for confirmation of each construct's effect on the HAC. KAP-1 was used as a positive control for centromere inactivation.
D. Nucleation of H3K27me3 or H3K9me3 "bookmarks" is insufficient to significantly impair HAC CENP-A levels, but direct tethering of PRC1 or HP1 causes severe loss of CENP-A. HeLa 1C7 cells were transiently transfected with TetR-EYFP constructs for 3 days before fixation and processing for immunofluorescence. Antibodies against CENP-A were used for HAC signal quantification. Mean of three independent experiments, n≥18 cells per condition in each experiment. Mann-Whitney U test was used to evaluate significance between differences observed.
E. As in D, but quantification of HAC mis-segregation after transient transfection with each respective TetR-EYFP construct. HAC mis-segregation was scored as an uneven number of HAC EYFP signals in transfected cells in early G1, assayed by the presence of a midbody. Mean of three independent experiments, n≥25 cells per condition in each experiment, error bars denote SD. Fisher's exact test was used to evaluate significance of difference between differences observed.
F. Centromeres show H3K4me3 signals within CENP-A domain. Stretched chromatin fibers from mitotic HT1080 cells were stained with antibodies against H3K4me3 (red) and CENP-A (green), and DNA was stained with hoescht 333342.
G. Centromeres show H2A.Z signals within or across CENP-A domain. As in D, but stained with antibodies against H2A.Z (red) and CENP-A (green).
Due to similarities in the two-step recruitment mechanism of the polycomb system, I decided to test if the same behaviour was observed for the heterochromatin system, which is more normally found surrounding centromeres. For this purpose, I tethered the Suv39h1 SET domain (which catalyzes H3K9me3) and HP1 (which enacts heterochromatin repression) to the HAC, as TetR-EYFP constructs. Previous reports had shown that tethering of either full-length Suv39h1 or HP1 (roughly the heterochromatin equivalents to polycomb EZH2 and PRC1, respectively) could achieve HAC centromere inactivation.\textsuperscript{294,407} However, Suv39h1 has been shown to be a scaffold which directly recruits many repressive factors, including HP1\textsuperscript{131,132,548}, which could exacerbate the repressive phenotype in a H3K9me3-independent manner. To ascertain only the effects of H3K9me3 on the HAC centromere, and possibly better mimic the effect of EZH2, I used a truncation of Suv39h1 composed of only its H3K9 methyltransferase SET domain (fig. 3.6C).

Interestingly, the direct tethering of HP1 significantly reduces HAC CENP-A and induces an increase in HAC mis-segregation to similar levels as BMI1, although neither as penetrant as the KAP-1 tethering positive control (fig. 3.6D, E). This suggests that indeed both of these downstream repressors can severely perturb CENP-A assembly and cause centromere malfunction. More surprisingly, tethering of Suv39h1-SET has the same effect on CENP-A levels as EZH2, indicating that nucleation of H3K9me3 on the HAC centromere (fig. 3.6C) per se does not lead to its inactivation, much like H3K27me3. However, I cannot yet exclude that long-term tethering of Suv39h1 could have more severe consequences than EZH2 tethering.

Together, these preliminary results show new clues regarding centromere sensitivity to repressive chromatin. Nucleation of initial repressive "bookmarks", such as H3K9me3 and H3K27me3, seems to contribute little to centromere inactivation, but direct tethering of downstream repressors, such as PRC1 and HP1, bypasses that apparent resistance.
Evidence for transcription start site chromatin modifications at centromeres

The fact that somehow H3K9me3 and H3K27me3 nucleation did not seem sufficient to negatively affect the HAC centromere, together with the observation that H3K9me3 is less enriched across the centromere core (fig. 3.1A), led me to hypothesise that the locus may be refractory to the establishment or maintenance of these marks. It is known that actively transcribing regions, especially promoter loci, recruit chromatin remodelling complexes which remove repressive marks such as H3K9me3 and H3K27me3. Furthermore, nascent transcripts were actively detected from mitotic centromeres. To investigate if there was any evidence for promoter-associated chromatin marks on human centromeres, I examined stretched chromatin fibers, since the discrete point signals of such marks would likely be too diluted in a ChIP-qPCR analysis on repetitive regions.

Immunofluorescence staining on chromatin fibers for H3K4me3 and H2A.Z, which are almost exclusively found at TSSs or enhancer sites (see Appendix table 6.1), allowed me to find centromeres with low-level accumulations of these marks, interspersed with CENP-A (fig. 3.6F,G). The accumulation was mostly punctate, similar to that observed in TSSs, and the distribution of the marks was more readily observed in the more stretched, thinner fibers. I was not able to determine if the HAC also possesses enrichment for these marks, as its fibers cannot be distinguished from those of other chromosomes, due to release of TetR binding in the staining protocol used. A quantification of these marks across different CENP-A-enriched fibers has not yet been performed. This will be required for accurate assessment of whether this enrichment is true for all centromeres. However, these initial experiments suggested that a putative source of active transcription initiation may be present at centromeres, and that it originates within CENP-A domains.
3.2.3. Discussion

Centromeres are resistant to polycomb repression by H3K27me3

By tethering the polycomb PRC2 component EZH2 to the HAC, I was able to establish enrichment of H3K27me3 and efficient recruitment of PRC1, effectively recapitulating conditions for endogenous polycomb repression. While the HAC chromatin was modified to a state less enriched with active transcription marks, consistent with polycomb repression, I was surprisingly unable to affect centromere transcription. Given that transcription-related chromatin marks across the HAC α-satellite array were severely reduced but HAC transcript levels did not change, this strongly suggests that transcription detected from the HAC is mostly centromere-driven (as opposed to passive transcription from the remainder of the array), as EZH2 tethering on a non-centromeric integration array (in HeLa 1F10) effectively causes silencing. My observation that transcription from the non-centromeric array can be silenced by TetR-EYFP-EZH2 tethering suggests that the HAC centromere possesses anti-silencing activity. This could potentially protect it from encroaching polycomb marks, or even other types of repression mechanisms, which could be an explanation for why centromeres can transcribe while being located within/adjacent to heterochromatin.

The changes observed in HAC chromatin upon EZH2 tethering cast into question the relevance of H3K4me2 and H3K36me2 for centromere activity, given that they can be mostly removed from the HAC with no loss of mitotic fidelity or transcription. Yet the nature of qPCR of repetitive regions, which averages signal enrichment across the whole region, might obscure sub-regions of resistant H3K4me2 and H4K36me2 (which could correspond to centrochromatin), that are not resolvable by ChIP. Therefore, it cannot be completely excluded that these active transcription marks are completely lost from the HAC centromere, only that the majority of them across the HAC (possibly in non-centromeric regions) are actually not required for centromere activity.
Given that the HAC in 1C7 cells shows, at the start, higher levels of H3K4me2 and transcripts than endogenous chromosome 21 centromere (fig. 3.4A, C, D), one can infer that the HeLa 1C7 HAC is inherently more euchromatic, probably because it was generated in low-heterochromatin HT1080 cells. It is therefore possible that the majority of H3K4me2 across the \( \alpha \)-satellite array of the HAC is redundant, but total removal of this mark (via LSD1) is deleterious. This could be confirmed at a latter date, by analysing HAC stretched chromatin fibers, for any remaining vestiges of H3K4me2 at the HAC CENP-A domain after EZH2 tethering.

Nonetheless, the HAC centromere was slightly affected by polycomb nucleation, with CENP-A, CENP-C and CENP-T being reduced in transient transfections of TetR-EYFP-EZH2 (which was clearly overexpressed in these cells). In the stable 1C7-EZH2 cell line, effects of polycomb on the HAC centromere were still observed, as CENP-A levels decrease slightly, but otherwise CENP-C and CENP-T levels did not show much alteration. Furthermore, HAC mitotic fidelity was not diminished, indicating that kinetochore protein levels can be reduced, with no deleterious effects.

In conclusion, the HAC centromere, its functions and associated transcriptional activity seems impervious to repression by polycomb silencing via H3K27me3 nucleation. Despite the fact that the HAC sports a fully functional centromere, its chromatin shows some differences from natural centromeres: it is more euchromatic and exhibits a slightly higher transcription level. It cannot be excluded, therefore, that endogenous centromeres (or possibly only a subset of them, depending on their chromatin signature) might be vulnerable to polycomb repression.
Centromeres show sensitivity to direct repression but not repression-associated "bookmarks", from both the heterochromatin and polycomb mechanisms

In an effort to understand why the HAC centromere resisted polycomb repression, I looked into the cell-cycle behaviour of polycomb repression, and confirmed that indeed PRC1 relocates away from chromosomes during mitosis, and only returns after mitotic exit. Given that centromeres have been shown to transcribe during mitosis \(^{560}\) and CENP-A assembly starts at telophase \(^{286}\), I hypothesized that inhibiting this relocation and retaining PRC1 at the HAC could likely result in centromere repression. In an interesting parallel, the same delocalization behaviour is also observed for HP1 proteins, as their H3K9me3 binding site is similarly affected by Aurora B-dependent H3S10 phosphorylation \(^{561–563}\). This raised the possibility that maybe centromere transcription requires this mitotic relocation (by either PRC1 or HP1) to alleviate repression.

While testing this hypothesis, I was able to prevent PRC1 relocation from chromatin, by inhibiting Aurora B, this had no effect on subsequent CENP-A assembly, indicating that PRC1 retention on the H3K27me3-rich domains of the HAC is not sufficient to affect CENP-A assembly. I have not yet assessed HAC transcript levels under conditions of PRC1 mitotic retention. Therefore, it is possible that transcriptional repression could have occurred, but with no loss of CENP-A assembly.

To further test whether polycomb repression could actually negatively affect centrochromatin, I performed direct tethering of PRC1, through transient transfection of its BMI1 subunit. This resulted in a CENP-A reduction and impairment of mitotic fidelity comparable to that of KAP-1 \(^{408}\), suggesting that polycomb repression factors can indeed impair the HAC centromere. Since direct tethering of PRC1 can affect the HAC centromere, but its recruitment by EZH2-promoted H3K27me3 cannot (even if forcibly maintained during mitosis), these results suggest two possibilities. First, maybe PRC1 was in fact never recruited onto the core centromere itself in EZH2-tethered HACs, despite being enriched
throughout the rest of the HAC. It is possible this is due to lack of H3K27me3 at that locus, either due to low H3 density (CENP-A replacing it instead) or because centrochromatin is refractory to modification by EZH2. I have not yet been able to show absence of H3K27me3 at the CENP-A domain of EZH2-tethered HACs, however, and demonstrating that would be key to proving this theory. Second, it may be that EZH2 tethering does indeed recruit PRC1 throughout centrochromatin, but the dynamics of PRC1 binding may be altered, preventing it from establishing a repressive state. The lesser effect of EZH2 tethering on HAC centromere activity could also be explained by the fact that its activity is dependent of local chromatin context. While the PRC2 complex has the potential to recruit DNA methyltransferases and histone deacetylases, EZH2 tethering to active chromatin has been shown to be unable to induce de novo DNA methylation, or reduce local H3K4me2 or histone acetylation. This indicates that EZH2's ability to remove active transcription marks as part of the repression process is limited under certain circumstances. It is possible that, although EZH2 tethering results in the majority of active chromatin marks at the HAC being reduced, as measured by ChIP, resistant foci might remain at a critical region within the centromere core.

Because polycomb and heterochromatin have similar chromatin-targeting mechanisms, I explored whether the same difference was observed for heterochromatin factors. Remarkably, I observed the same response of HAC CENP-A levels to Suv39h1-SET and HP1, as I did for EZH2 and BMI1, respectively. This could suggest that centromeres use a similar mechanism to cope with the two modes of repression. Thus, the initial complexes, which lay down the "repressive bookmarks" H3K9me3 or H3K27me3, have little effect, but direct targeting of effective repressors achieves efficient reduction of CENP-A. Interestingly, although Suv39h1-SET tethering has little impact on CENP-A levels, full-length Suv39h1 tethering severely affects CENP-A, possibly through H3K9 methylation-independent events such as direct HP1 recruitment. Such an observation had also been observed.
for KAP-1 targeting to the HAC \textsuperscript{408}, where an HP1-recruiting fragment of KAP-1 led to efficient centromere inactivation, but a fragment which bound only the NuRD complex (histone deacetylase) and SETDB1 (an H3K9 methyltransferase) \textsuperscript{566,567} did not. Together, these observations reconcile the previous finding that effective heterochromatin repression (whether by full-length Suv39h1 or HP1) can indeed inactivate centromeres with the fact that the natural genomic surroundings of centromeres are enriched in this chromatin state. Centromeres may have an active mechanism to prevent H3K9me3 nucleation and heterochromatin spreading.

Collectively, I infer from these results that while direct repression, through tethering of effective silencing complexes PRC1, HP1 and KAP-1, can achieve loss of centromere activity, nucleation of just the initial steps of silencing, the "bookmarks" H3K9me3 or H3K27me3, carries little negative effects. This has been confirmed for EZH2-dependent H3K27me3 nucleation, but for H3K9me3 only short-term effects were examined. In future experiments, long-term tethering of Suv39h1 could be studied to verify if indeed this response is in fact similar.

\textit{Pericentromeric transcriptional silencing marks are less enriched at the centromere core, instead promoter-associated marks can be found there}

I also confirmed the finding, first observed by Sullivan & Karpen \textsuperscript{456}, that centrochromatin seems to have low levels of silencing chromatin marks, but I have expanded this by showing that polycomb marks, in addition to heterochromatin, can also have this distribution. By performing immunostaining against H3K9me3 and H3K27me3 on chromatin stretched fibers, I observed that signals for both repressive chromatin types can exist near centrochromatin, but that their levels are lower or absent within the CENP-A
domain. This suggests that these marks don't readily spread to the centromere core (where CENP-A is located), where transcription-related chromatin marks can be found\textsuperscript{405,456,520}.

It is possible that the mechanisms (whatever they may be) that prevent full repression of centromeres by heterochromatin and allow active transcription within them, might also function the same way for adjacent polycomb chromatin, as these two repression systems have similar recruitment mechanisms\textsuperscript{568,569} and in some situations can functionally interact\textsuperscript{495–497}.

The filamentous fungus \textit{N. crassa} shows a similar distribution in its centromeres, which have no obvious sequence-dependent boundary between centromere and pericentromere. They present H3K9me3 enrichment across both regions, but exhibit a lack of DNA methylation at the centromere core\textsuperscript{329}. However, no analysis of chromatin marks associated with transcription was performed in that study. Mouse centromeres also show this same heterochromatin pattern, but the boundary is correlated with underlying DNA sequence, as mouse CENP-A sits on minor satellite arrays, and heterochromatin in flanking major satellite arrays\textsuperscript{126,381–383}.

Most pericentromeric regions are enriched in H3K9me3, but there is evidence for high heterogeneity of chromatin states between cell types\textsuperscript{519,525}. The HT1080 cell line used in these assays expresses lower levels of Suv39h1\textsuperscript{294}, which catalyzes H3K9 methylation and controls constitutive heterochromatin formation\textsuperscript{518,570}. Absence of both H3K9me3 and H3K27me3 at some pericentromeres in the HT1080-derived HeLa 1C7 could be an outcome of this initial low-heterochromatin state. Nonetheless, it cannot be excluded that residual levels of repressive chromatin marks could still be associated with discrete centromere locations, but undetectable by immunofluorescence staining.

H3K9me3 and H3K27me3 seem to be less dense across CENP-A domains at native centromeres. The HAC centromeric response to nucleation of either mark is quite similar, at least with regard to levels of CENP-A, suggesting a similar mechanism could be providing
resistance against both repressive chromatin states. Polycomb and heterochromatin have a similar recruitment method to chromatin, in which secondary repressors (PRC1 and HP1, respectively) recognize methylated “ARKS” motifs on the H3 tail through their chromobox domains. There could be several explanations for this apparent low level of repressive marks: lower H3 occupancy across the CENP-A domain (and therefore less substrate for methylation), the presence of active TSSs (which promote low level of repressive marks) or even sterical hindrance of repression by the presence of CCAN proteins.

**An active centromeric TSS as anti-silencing mechanism**

I have shown evidence for putative TSS-like activity at human centromeres, by observing that H3K4me3 and H2A.Z can be enriched at human centromeres. Can centromere-driven transcription initiation be the reason why repressive marks are decreased across centrochromatin? It is known that both H3K9me3 and H3K27me3 are highly depleted at the TSS and 5' region of active genes, thus keeping secondary repressors (such as PRC1 and HP1) away from promoter regions. Removal of such repressive marks, and delocalization of the complexes that promote them, is essential for prevention of promoter silencing, by forming a barrier against spreading of repressive chromatin. Additionally, higher histone turnover, generally associated with 5' regions of transcribing chromatin, is known to counter-act repressive chromatin states by remodelling out histones, thus removing any possible repressive marks on them. This may occur at centromeres as well, as canonical nucleosomes within centrochromatin are enriched for the H3.3 variant, which replaces replication-deposited H3.1 in a co-transcriptional manner. Curiously, newly-identified centromere-associated proteins Eaf6/CENP-28, MSL1v1/CENP-36 and PHF2/CENP-35 are subunits of transcription initiation complexes present at promoter regions. PHF2/CENP-35 has been shown to directly bind H3K4me3 in vitro, have H3K9 demethylation activity, and be part of a family of promoter-
binding JmjC-domain proteins, whose collective function is to remove repressive histone methylations such as H3K9me3, H3K27me3 and H4K20me3.  

All the above hypotheses rely on active transcription to maintain any putative barrier against repression: abolishing transcription initiation altogether could possibly break this defense, maybe irreversibly. LSD1 tethering, which caused transcriptional repression and centromere inactivation at the HAC, by removing H3K4me3/2 marks, led to total loss of centromere transcription within 3 days. It is possible that this effect might be irreversible, if chromatin marks associated with active transcription (including those normally found at TSSs) were removed, and memory of transcriptional initiation lost. It would be interesting to observe if the HAC centromere can recover from LSD1-mediated transcriptional silencing, and subsequently, resist H3K27me3 or H3K9me3 nucleation.

3.2.4. Conclusion and outlook

I have demonstrated in this work the persistence of centrochromatin within the polycomb repressive chromatin state. I have explored how chromatin marks associated with repressive chromatin states can distribute on and around centromeres, and I have directly probed centromere response to both heterochromatin and polycomb. Further confirmation is necessary to understand if this response is indeed true, for both these chromatin states, both at short-term and at long-term. If validated, this can represent a paradigm shift, in that it had been assumed that the presence of heterochromatin on centromere repeats fully antagonized de novo centromere formation. My data thus helps to reconcile previous observations, that centromeres are often localized within heterochromatin regions and that heterochromatin factors may even contribute to centromere formation, yet heterochromatin can antagonize de novo formation of centromeres.
Future work on centromere survival within repressive chromatin states, and which specific factors promote it, can shed light on how centromeres respond to genome-wide changes to chromatin regulation, and how these essential genomic regions remain durable and adapt across evolution.
3.3. Heterochromatin stabilizes the core centromere and prevents DNA damage

3.3.1. Background

Centromeres are associated with repetitive regions of the genome in many studied eukaryotes, but these repeats only show sequence conservation between closely related species. These regions generally possess a repressed transcriptional state mediated by canonical heterochromatin factors. The fact that centromere maintenance seems to require active transcription or a non-repressed chromatin state contrasts against its silencing heterochromatin surroundings. It can be asked then, what the evolutionary benefits are of having centromeres in such regions of the genome.

In *S. pombe* it has been demonstrated that heterochromatin factors recruit cohesin. This provided a good explanation for why such regions would have been conserved on chromosomes in the vicinity of the centromere: to provide robust sister chromatid cohesion adjacent to the kinetochores, thereby maximizing tension under bipolar attachment. Indeed, it has been shown that depletion of heterochromatin-associated factors in cells of a variety of organisms can lead to mitotic defects. Furthermore, knock-out mice without heterochromatin-depositing enzymes have a predisposition for cancer and suffer synapse defects during meiosis which lead to oocyte apoptosis.

Oddly, at least in both mouse and human cells, loss of heterochromatin (by Suv39h1/h2 knock-out or HP1 RNAi, respectively) does not lead to loss of centromeric cohesin in mitosis. In these cases, only a mild separation (but not disjunction) of sister chromatid centromeres could be observed, yet these cells still suffered significant chromosome mis-segregation. In *S. cerevisiae*, it has been suggested that cohesin-enriched silenced domains (such as the HMR mating-type locus) do not contribute to sister-chromatid cohesion. Instead, the authors have put forward that cohesin accumulation at these loci is
intra-chromatid and not between sister chromatids, possibly contributing to chromatin compaction \(^{333,608-610}\). However, because the repressive chromatin state in \(S. cerevisiae\) is mediated by \(Sir\) proteins instead of canonical heterochromatin \(^{611-613}\), it remains to be tested whether the same is true for heterochromatin-based cohesin recruitment. A clue to answer this question may come from experiments in mice without heterochromatin factors Suv39h1/h2, where chromosomes mis-segregate and show some loosening of sister chromatid cohesion in mitosis, but remain cohered and still retain the cohesin complex at mitotic centromeres \(^{498,606}\). This suggests that indeed heterochromatin-dependent cohesin recruitment may not play a significant role in the maintenance of sister chromatid cohesion.

Experiments in \(Drosophila\) cells have shown that neocentromeres formed by transient CENP-A overexpression are often situated adjacent to heterochromatin \(^{322,367}\). Moreover, neocentromere formation in \(S. pombe\) is impaired in heterochromatin mutants \(^{348}\). In fact, in both \(Drosophila\) and \(S. pombe\), ectopic nucleation of heterochromatin can cause accumulation of centromere proteins and formation of a functional kinetochore \(^{322,352,353}\). In both cases, these centromeres are epigenetically maintained even in subsequent absence of heterochromatin. This suggests heterochromatic regions may have a stabilizing effect during centromere establishment, rather than being essential for its maintenance. In \(S. cerevisiae\) and human cells it has been shown that high levels of transcription through the centromere domain can evict CENP-A, causing the centromere to be lost \(^{335-337,405,407}\). Human neocentromeres have been described to shift position and lose CENP-A density after treatment with a deacetylase inhibitor, which causes chromatin to become more open \(^{470}\). Together, these findings encourage the idea that somehow heterochromatin (or overall silenced chromatin) might provide a local dampening of transcription, or a critical chromatin compaction level, which may promote formation and/or stabilization of centromeres.

On the other hand, several studies present the opposite view: that heterochromatin is dispensable or even incompatible with centromere function. HAC-generating experiments in human cells show that the efficiency of \textit{de novo} centromere formation is counteracted by the
cellular levels of heterochromatin. Additionally, inactive centromeric repeats (such as integrated HAC arrays or inactive former centromeres in fused dicentric chromosomes), can be reactivated by antagonizing chromatin silencing, such as through inhibition of deacetylases, local targeting of acetyltransferases (in human cells), or by deleting heterochromatin pathway components (in S. pombe). Furthermore, direct targeting of heterochromatin factors to HAC centromeres also leads to their silencing and inactivation. Perhaps more striking is the fact that some centromeres can exist in apparently heterochromatin-free regions of the genome. Three centromeres of chicken, a few characterized human neocentromeres, and recently an S. pombe neocentromere all have been described as being heterochromatin-poor, although they retain full kinetochore activity and mitotic fidelity. This highlights that, even if there might be establishment or evolutionary benefits of having pericentromeric heterochromatin, it is not strictly essential for centromere formation and maintenance.

The evidence in support of the hypothesis that heterochromatin is important for the formation and maintenance of centromeres, as well the evidence opposing it, comes from several different organisms. It is therefore likely that the observations and conclusions found are coloured both by non-conservation of functions between organisms (in both regulation of chromatin and centromere protein function) as well as the caveats of the methods used for each. Yet both sides present evidence that chromatin silencing can, at the very least, influence centromere integrity and/or formation (for better or worse), as opposed to playing only a structural role through cohesion recruitment (as proposed before).

Consequently, there is a question that needs solving: how exactly do centromeres interact with heterochromatin? The answer will increase our understanding of how centromeres network with their genomic surroundings, how they respond to changes in whole chromatin reprogramming, and their evolutionary history in relocating along chromosomes. Ultimately, this will allow improvements in the reliability of generating artificial centromeres, in a controlled fashion. More importantly, given that disruption of
genome-wide chromatin regulation is so frequent in cancer cells, due to misregulated expression of key regulatory proteins \(^{616,617}\), it is likely that centromeres might not escape such changes unscathed. It has been shown that tumorigenic transformation of primary human cells can cause a decrease in pericentric heterochromatin, probably through upregulation of H3K9 demethylases, leading eventually to a CIN phenotype \(^{618}\). CIN cells frequently show evidence of (peri)centromeric translocations \(^{619,620}\), further implicating defects at the pericentromere with loss of mitotic fidelity. Given that acquisition of the high-aneuploidy CIN behaviour is a driver for tumour formation and adaptation \(^{25,621–624}\), it is crucial to understand if heterochromatin deregulation can indeed contribute to this process, and if so, through which mechanisms.

### 3.3.2. Results

**JMJD2D tethering removes heterochromatin from the HAC**

In order to gain a better understanding of the role of heterochromatin at centromeres, I decided to study the effects of heterochromatin depletion in an active (peri)centromere. To do this, I turned to the targetable alploid\(^{\text{TetO}}\) HAC system in human cells \(^{294,339,405,407,408}\). This system allows me to affect (peri)centromeric heterochromatin locally on a single chromosome, thereby avoiding deleterious effects of whole-genome heterochromatin depletion. The conditional nature of this system is a powerful advantage, because it allows for step-by-step characterization of the effects of (peri)centromeric heterochromatin loss over time, possibly recapitulating events during tumorigenesis.

In this study I utilized HeLa 2-4 cells, which carry a HAC consisting of an \(\alpha\)-satellite\(^{\text{TetO}}\) repetitive array based on those of chromosome 21 (\(\alpha\)-21-I satellite repeats). This HAC shows a clear enrichment for H3K9me3 immunofluorescence staining, as visualized by TetR-EYFP binding in interphase cells (fig. 4.1A). This makes HeLa 2-4 cells a good tool to visualize depletion of HAC heterochromatin, and avoids any artefacts from other HAC cell
lines, such as AB2.2.18.21 and Hela 1C7, which were derived from low-heterochromatin HT1080 cells.

In order to remove heterochromatin from the HAC, I used the JMJD2D demethylase, which has been shown to remove the H3K9me3 modification. This enzyme also seems to remove H1.4K26 and H3K56me3. But because both of these are also heterochromatin marks, I considered JMJD2D as a reasonable tool for effective removal of the heterochromatic state, especially given that other reported H3K9 demethylases also affect euchromatic marks such as H3K36me2. In addition to wild-type JMJD2D, I also used a catalytically-null mutant version of the protein (JMJD2D D195A) as a negative control. This ensured that any phenotypes that I observed are demethylation-specific, and helped account for effects the enzyme might cause in addition to demethylation (such as activities of other domains of the protein or of possible binding partners). Both sequences were cloned into mammalian expression vectors to generate fusion proteins TetR-EYFP-JMJD2D and TetR-EYFP-JMJD2D D195A.

I transiently transfected the aforementioned TetR-EYFP constructs into HeLa cells for 48h, to allow them to tether to the HAC, and then performed immunofluorescence staining against heterochromatin markers. Transfection of JMJD2D resulted in significantly reduced H3K9me3 levels on the HAC, to very close to nuclear background level, compared to TetR-EYFP control (fig. 4.1C). Transfection of JMJD2D D195A also reduced H3K9me3 levels, but clearly not to the same extent as the wild-type enzyme (fig. 4.1C). Analysis of HP1α staining shows that JMJD2D also efficiently delocalizes this heterochromatin protein from the HAC, however JMJD2D D195A has a similar effect on HP1α levels (fig. 4.1D).

This shows that JMJD2D tethering to the HAC successfully removes canonical heterochromatin markers from the HAC. On the other hand, the effects of JMJD2D D195A
on HAC heterochromatin markers H3K9me3 and particularly HP1α reveals that care must be taken in interpreting further results.

**Short-term HAC heterochromatin removal does not affect centromere protein levels**

To assess the effect of heterochromatin loss on the HAC centromere, I analyzed the levels of centromere proteins CENP-A and CENP-C by immunofluorescence. After 48h transient transfection, of TetR-EYFP-JMJD2D or TetR-EYFP-JMJD2D D195A, no significant changes were detected in the levels of either protein (fig. 4.1E, F). As such, I conclude that loss of heterochromatin seems to have no immediate impact on core centromeric proteins, indicating that it is not a necessary active component of centromere maintenance (i.e., every cell cycle), as opposed to critical factors like Mis18 or HJURP.

205,261,287,292,297,301,632-634.
Fig 4.1 – Short-term tethering of TetR-EYFP-JMJD2D efficiently removes heterochromatin from the HAC, but does not affect centromere proteins or mitotic fidelity

A. Analysis of interphase HeLa 2-4 cells, 48h after transient transfection with plasmids expressing TetR-EYFP, TetR-EYFP-JMJD2D and TetR-EYFP-JMJD2D D195A (green). Cells were fixed and stained with antibodies against CENP-A (red) and H3K9me3 (blue) and DNA was stained with hoescht 333342 (gray). Arrowheads locate the HAC. Scale bar: 5 µm.

B. Insets, focusing on the HAC, of HeLa 2-4 interphase cells, transiently transfected as in A. Cells were fixed and stained with antibodies against CENP-C (red) and HP1α (blue), and DNA (gray) stained as in A.

C. Quantification of mean HAC-associated H3K9me3 immunofluorescence signal, with subtraction of local background, in images of A. Total of three independent replicates, n>35 interphase cells each, blue bar indicates median, red dotted line indicates background level. Mann-Whitney U statistical test showed significant differences between results: *** = p<0.0001.

D. Quantification of HAC-associated HP1α, as described in C, in images of B. Total of two independent replicates, n>36 interphase cells each. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

E. Quantification of mean HAC-associated CENP-A immunofluorescence signal, normalized to endogenous centromeres, in images of A. Total of two independent replicates, n>29 interphase cells each, blue bar indicates median. Mann-Whitney U statistical test showed no significant differences between results.

F. Quantification of HAC-associated CENP-C, as measured in E, in images of B. Total of two independent replicates, n>26 interphase cells each, blue bar indicates median. Mann-Whitney U statistical test showed no significant differences between results.
**A** TetR-EYFP
2 days post transfection

CENP-A
H3K9me3
merge + DNA

**B** TetR-EYFP-JMJD2D
2 days post transfection

CENP-A
H3K9me3
merge + DNA

**C** TetR-EYFP-JMJD2D D195A
2 days post transfection

CENP-A
H3K9me3
merge + DNA

---

**C** HAC H3K9me3

mean HAC A.F.U.
(background-subtracted)

**D** HAC HP1α

mean HAC A.F.U.
(background-subtracted)

**E** HAC CENP-A

HAC A.F.U. normalized to endogenous centromeres

**F** HAC CENP-C

HAC A.F.U. normalized to endogenous centromeres

---

TetR-EYFP construct

TetR-EYFP construct

TetR-EYFP construct

TetR-EYFP construct
Loss of HAC heterochromatin disrupts HAC centromere integrity and mitotic fidelity

To investigate the effects of long-term loss of (peri)centromeric heterochromatin, I established two HeLa 2-4 HAC cell lines, that stably express TetR-EYFP-JMJD2D or TetR-EYFP-JMJD2D D195A (henceforth called 2-4-JMJD2D and 2-4-JMJD2D-D195A, respectively). These cell lines were isolated and cultured in the presence of Doxycycline, to prevent tethering of the TetR constructs prior to initiating the experiments.

To control for any long-term effects of tethering TetR to the HAC, I used a previously established HeLa 2-4 cell line that stably expresses TetR-EYFP (2-4-EYFP-only). The effects of TetR-EYFP tethering, by itself, on the HAC centromere have not been thoroughly studied for a period longer than 5 days (in transient transfections \cite{294,339,405,407,408}), so this was an important opportunity to perform essential quality controls for the alphoid\textsuperscript{TetO} HAC system. TetO arrays with bound TetR have been reported to interfere with replication progression in \textit{E. coli} \textsuperscript{635}, and slightly affect transcription levels and histone modifications at human HACs \textsuperscript{339,405}. Therefore, properly assessing TetR-derived perturbations is essential for validation and future use of this system.

After washing out doxycycline in each cell line and allowing the constructs to tether to the HAC for 8 days, I performed immunofluorescence staining for centromere proteins and quantified HAC mitotic phenotypes in fixed asynchronous cell cultures. After only a short tethering period (~1h), there was no significant reduction of mitotic fidelity (fig. 4.2B) or CENP-A protein levels (fig. 4.2A, C) of the HAC, in either the 2-4-JMJD2D or the 2-4-JMJD2D-D195A cell line, compared to the 2-4-EYFP-only control cell line. This was consistent with observations from transient transfections. However, after 8 days of JMJD2D tethering, there was a marked reduction of CENP-A (fig. 4.2A, C), and an increase in HAC mis-segregation (fig. 4.2B), as seen by the presence of co-segregated or lagging HAC chromatids, and HAC-containing micronuclei \textsuperscript{407}. Surprisingly, these JMJD2D-tethered...
HACs showed few metaphase defects apart from a slightly increased number of misaligned HAC chromatids at 8 days of tethering (fig. 4.2B): the vast majority congressed and bi-oriented correctly. HACs of the 2-4-JMJD2D-D195A cell line behaved similarly to HACs on the 2-4-EYFP-only control cell line (fig. 4.2B), confirming that the segregation defect observed upon JMJD2D tethering is indeed dependent on its catalytic activity. As such, the reduction in H3K9me3 and HP1α caused by JMJD2D D195A tethering, in transient transfections, may not contribute much to any centromeric defects caused by wild-type JMJD2D. It should be noted however, that both TetR-EYFP and JMJD2D D195A cause some reduction of CENP-A levels (fig. 4.2C), even though this does not seem to impact HAC mitotic fidelity very much (fig. 4.2B).

Together, these results suggest that, by tethering of JMJD2D, long-term removal of H3K9me3 from the HAC results in reduction of its core centromere proteins and causes a mitotic defect that leads to HAC mis-segregation upon anaphase onset. However, this does not prevent the HAC from congressing to the metaphase plate. It seems that, although core centromere proteins are reduced when TetR-EYFP-JMJD2D is tethered for longer periods, a large enough pool remains at centromeres to support kinetochore function. These results also suggest that the primary reason for HAC mis-segregation might not be loss of the centromere/kinetochore (given that the HAC still attaches to the mitotic spindle), but a defect in centromere structure and/or alteration of a mitotic regulatory pathway (such as kinetochore attachment error-correction or cohesin release), with consequences that only become obvious upon anaphase onset. These results are in line with previous observations of heterochromatin loss/depletion causing defects in centromere function 9,128,139,329,349,603,604.
Fig 4.2 – Long-term tethering of TetR-EYFP-JMJD2D causing HAC mis-segregation and reduction of CENP-A

A. Representative pictures of divided pairs of 2-4-JMJD2D cells, at 1 day and 8 days after doxycycline washout to allow TetR-EYFP-JMJD2D tethering. Cells were fixed and stained with antibodies against CENP-A (red) and DNA was stained with hoescht 33342 (blue). Arrowheads denote HACs, and insets show CENP-A reduced after 8 days of tethering. Scale bar: 5 μm.

B. Quantification of mitotic phenotypes in fixed cultures of 2-4-EYFP-only, 2-4-JMJD2D and 2-4-JMJD2D D195A cells. Total of two independent replicates, metaphase defects n>100 metaphase cells each; segregation defects n>97 telophase or early G1 cells each. Fisher’s exact test was used to evaluate significance between differences observed. Metaphase defects: EYFP 1d vs 8d ***p=0.0001, JMJD2D 1d vs 8d ***p=0.0001, JMJD2D D195A **p=0.0013. Segregation defects: EYFP 1d vs 8d *p=0.0135, JMJD2D 1d vs 8d ***p=0.0001, JMJD2D D195A **p=0.0013; EYFP 8d vs JMJD2D 8d ***p=0.0001, EYFP 8d vs JMJD2D D195A 8d n.s. p=2426.

C. Quantification of mean HAC-associated CENP-A immunofluorescence signal, normalized to endogenous centromeres. Single experiment, n>14 interphase cells each condition, blue bar indicates median. Mann-Whitney U statistical test shows significant difference between results: * p=0.026, ** p=0.0027, *** p<0.0001.
**HAC H3K9me3, even at residual levels, is required to prevent HAC centromere disruption and DNA damage**

To understand the long-term effect of TetR-EYFP tethering constructs on HAC H3K9me3 levels, I performed a second long-term tethering experiment, culturing cells either in the presence or absence of doxycycline, with the objective of visualizing HAC H3K9me3 levels in each condition. Because growing HAC cells under doxycycline prevents HAC visualization using TetR-EYFP (by inhibiting TetR binding), I used a strategy to locate the HAC by tethering TetON-tdTomato. This construct tethers to the HAC only in the presence of doxycycline, the opposite of TetR , allowing me to locate the HAC by tdTomato fluorescent signal , when TetR constructs cannot be tethered.

The results observed after 8 days of tethering recapitulated those obtained by transient transfection. JMJD2D efficiently removes H3K9me3 (to close to background levels) (fig. 4.3A, B), while JMJD2D D195A reduces H3K9me3 levels somewhat but these still remain significantly above background (fig. 4.3A, B). Importantly, tethering of TetR-EYFP by itself, for 8 days, led to some reduction of H3K9me3 levels as well, yet still significantly above background levels (fig. 4.3A, B). This indicated that TetR-EYFP on its own might interfere slightly with heterochromatin structure and/or maintenance.

In 2-4-JMJD2D cells, however, I discovered that initial H3K9me3 levels on the HAC were quite low (fig. 4.3A, B). There was only a residual level of HAC H3K9me3 before doxycycline washout, much lower than control cell lines, which is then further decreased after JMJD2D tethering. Other heterochromatin regions throughout the nucleus of these cells still seem significantly enriched in H3K9me3 (fig. 4.3A), suggesting that it was not due to an overall reduction of H3K9me3 throughout the genome. Nevertheless, this low initial level of heterochromatin is significantly above background levels, and correlates with maintenance of correct chromosome segregation and centromere function, as well as CENP-A levels comparable to control cell lines (fig. 4.2B, C), as long as TetR-EYFP-JMJD2D is not tethered to the HAC.
Loss of heterochromatin on repetitive regions has been reported to lead to accumulation of unresolved recombination intermediates from DNA repair, when replication fork collapse occurs. This can cause segregation defects and DNA damage. Immunofluorescence staining for 53BP1, a DNA double-strand repair signalling protein, revealed that after 8 days of tethering, HACs in 2-4-JMJD2D cells indeed showed enrichment for 53BP1. This enrichment was only visible in early G1 cells, consistent with damage on these regions being caused during or at the end of mitosis, as has been demonstrated previously. 53BP1-enriched G1 foci appear occasionally, in natural cell populations, on regions of replication stress, such as repetitive regions, to allow for repair of DNA damage, so it would not be completely unexpected to find 53BP1 foci at the HAC. However, control 2-4-EYFP-only or 2-4-JMJD2D-D195A HACs do not show significant increase in HAC 53BP1 over background levels, thus confirming that tethering of TetR-EYFP-JMJD2D specifically causes DNA damage signalling to increase at the HAC.

Therefore, preservation of HAC heterochromatin, even if only at low levels, correlates with avoidance of active DNA damage signalling (and presumably DNA damage) on the HAC, in addition to maintenance of CENP-A levels and preventing mitotic segregation problems. Moreover, the timing of the hypothetical DNA damage (early G1, after mitotic exit) points to breakage events caused during a mitotic or G1-specific event (such as CENP-A assembly itself).
Fig 4.3 – HACs of 2-4-JMJD2D cells have only residual H3K9me3 before JMJD2D tethering, but it is sufficient to prevent post-mitotic DNA double-strand breaks; HACs centromere can recover from heterochromatin removal.

A. 2-4-EYFP-only, 2-4-JMJD2D and 2-4-JMJD2D D195A cells were washed of doxycycline, transiently transfected after 7 days with a plasmid expressing TetON-tdTomato, then fixed the next day, for immunofluorescence with antibodies against H3K9me3 (blue) and DNA was stained with Hoescht 333342 (gray). Cultures of each cell line, but still under doxycycline, were equally processed in parallel. Initial HAC H3K9me3 levels can be observed, in each cell line, by TetON-tdTomato binding, under doxycycline conditions. Scale bar = 5 µm.

B. Quantification of mean HAC-associated H3K9me3 immunofluorescence signal, with subtraction of local background, in images of A. Total of three independent replicates, n≥13 interphase cells each, blue bar indicates median, red dotted line indicates background level. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

C. Quantification of mean HAC-associated 53BP1 immunofluorescence signal, with subtraction of local background, in images of C. Total of two independent replicates, n>31 early G1 cells each, blue bar indicates median, red dotted line indicates background level. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

D. Analysis of DNA damage signalling in early G1 cells of 2-4-EYFP, 2-4-JMJD2D and 2-4-JMJD2D D195A. Cells were washed out of doxycycline for the indicated periods and then fixed for immunofluorescence: antibodies against CENP-A (red) and 53BP1 (blue)(a DNA-damage signalling factor) were used, and DNA was stained with Hoescht 333342 (gray). Early G1, recently-divided cell pairs were located by presence of cytoplasmic connection through midbody remnants. Arrowheads denote the HAC.
A) TetR-EYFP cell line

- EYFP only
- JMJD2D
- JMJD2D D195A

8 days of tethering

B) HAC H3K9me3

- Mean HAC A.F.U. (background-subtracted)

Tethering duration
- None
- 8 days

TetR-EYFP cell line
- EYFP only
- JMJD2D
- JMJD2D D195A

C) HAC 53BP1 levels

- Mean HAC A.F.U. (background-subtracted)

TetR-EYFP cell line
- Early G1 cells

D) TetR-EYFP

- DNA
- Merge
- CENP-A
- 53BP1

- 8 days of tethering

TetR-EYFP-JMJD2D

- DNA
- Merge
- CENP-A
- 53BP1

- 8 days of tethering

TetR-EYFP-JMJD2D D195A

- DNA
- Merge
- CENP-A
- 53BP1

- 8 days of tethering
**JMJD2D removes heterochromatin marks from the HAC, but causes it to have acquired polycomb marks**

Tethering of JMJD2D the HAC had a significant negative impact on its mitotic behaviour, centromere protein pool and genomic stability. As such, I sought to identify what other chromatin changes had occurred at the HAC, to gain insight into possible causes that may be at the root of these observed defects. To that end, following JMJD2D tethering, I performed ChIP-qPCR for different chromatin marks on 2-4-EYFP-only and 2-4-JMJD2D cells (fig. 4.4B).

After 5 days in the absence of doxycycline, in 2-4-JMJD2D cells, I observed that JMJD2D successfully removed H3K9me3. I also observed removal of H4K20me3, confirming successful loss of the heterochromatin state. Initial levels of HAC H3K9me3 in 2-4-JMJD2D HACs, in the presence of doxycycline, were indeed low, as observed by immunofluorescence, but still considerably higher than those found on a constitutively active gene (PABPC1) (fig. 4.4B). It is thus reasonable to hypothesize that the levels of H3K9me3 on the HAC, although low, may be biologically significant. Furthermore, H3K9me3 levels at the centromere of chr. 21 (α21-I) are comparable to those in control cell line 2-4-TetR-EYFP, either in the presence or absence of doxycycline (fig. 4.4B). This supports that TetR-EYFP-JMJD2D, which is overexpressed in these cells, has little effect on loci to which is not directly tethered, as was suggested by microscopy observations of H3K9me3 in 2-4-JMJD2D cells (fig. 4.3A). Therefore, it is unlikely that the reduced initial levels of H3K9me3 on the HAC in these cells are caused by untethered TetR-EYFP-JMJD2D overexpression.

Loss of heterochromatin, which promotes silencing, can cause underlying loci to be reactivated and transcribed \(^{349,649-652}\). Furthermore, CENP-A stability is known to be compromised when local transcriptional activity increases excessively \(^{335-337,339,407}\). To assess whether HAC centrochromatin had acquired euchromatic characteristics, I analyzed levels of H3K4me2, which is found in regions of active transcription \(^{459-461}\). I observed no change in
HAC H3K4me2 levels by ChIP-qPCR, following H3K9me3 removal (fig. 4.4A). This suggested that transcriptional activity, even after heterochromatin depletion, might not increase significantly.

To test whether removal of HAC heterochromatin did not necessarily lead to an excessive increase in HAC (peri)centromeric transcription, I quantified HAC-derived transcripts after JMJD2D tethering. After 5 days of doxycycline removal, I extracted RNA from 2-4-EYFP-only and 2-4-JMJD2D cells, and performed RT-qPCR to quantify transcripts. There was no observable change in HAC α-satellite$^{\text{TetO}}$ transcripts, after tethering either TetR-EYFP or TetR-EYFP-JMJD2D (fig. 4.4B). Overall this indicates that the HAC segregation defects induced by JMJD2D tethering are unlikely to be caused by over-transcription of its centromere domain.

Unexpectedly, ChIP-qPCR measurements showed a considerable reduction of CENP-A levels after 5 days of control TetR-EYFP tethering, comparable to that observed on JMJD2D-tethered HACs (fig. 4.4A). This reduction was less pronounced than that measured by immunofluorescence staining, where JMJD2D induced much greater CENP-A reduction (fig. 4.2C). Given that ChIP-qPCR measures CENP-A presence only on the α-satellite$^{\text{TetO}}$ repeats of the HAC, and immunofluorescence quantification shows that total HAC CENP-A levels are higher, this suggests that there may be additional CENP-A on these HACs, away from the α-satellite$^{\text{TetO}}$ repeats, to explain this paradox. Centromere redistribution has been reported previously in human cells, after perturbation with TSA$^{470}$. As such, it is possible that long-term TetR-EYFP tethering can induce a slight redistribution of centriochromatin on the HAC (possibly to HAC backbone sequences), while still maintaining centromere integrity.
**Fig 4.4** – JMJD2D efficiently removes heterochromatin marks from the HAC, with no increase in transcription; HAC of 2-4-JMJD2D cells possesses a low-heterochromatin, high-polycomb profile before tethering

A. Chromatin analysis by ChIP of 2-4-EYFP-only and 2-4-JMJD2D cell lines. Cells were washed of doxycycline and grown for 5 days before harvesting and processing for ChIP and qPCR. Cells grown in doxycycline were processed equally, in parallel. Mouse antibodies against H3K9me3, H4K20me3, H3K4me2, CENP-A and H3K27me3 were used; mouse unspecific IgG was used as negative pulldown control. Pulldown DNA was quantified by qPCR. Primer sets used are shown at the top. Single experiment, statistical tests could not be performed due to absence of experimental replicates.

B. Quantification of transcripts from 2-4-EYFP-only and 2-4-JMJD2D cell lines. Cells were grown as in A, harvested for RNA extraction, cDNA was retrotranscribed and transcripts were quantified by qPCR. Expression level is normalized to genomic copy number (for repeats) and normalized to β-actin. Single experiment, statistical tests could not be performed due to absence of experimental replicates.
Initial levels of heterochromatin marks in HACs of 2-4-JMJD2D cells were lower than in those on control and parental HeLa 2-4 HACs. This raised the possibility that its chromatin was somehow different, from the start, possibly as a result of clonal selection during stable cell line establishment. Remarkably, ChIP-qPCR analysis showed that 2-4-JMJD2D HACs, before JMJD2D tethering, are polycomb-rich rather than heterochromatin-rich, as observed by high levels of H3K27me3 (fig. 4.4A). This was not observed in control 2-4-EYFP-only HACs, suggesting that HACs in 2-4-JMJD2D cells might have adapted during the process of selection.

Double knock-out of the Suv3-9h1/h2 gene in mouse, which abolishes classical heterochromatin, has been shown to allow polycomb repression to take over previously heterochromatic regions \(^{126,493,498,499}\). A similar response was also seen in paternal-derived chromatin, in early mouse embryos, which initially carries no heterochromatin marks \(^{499}\). A similar phenomenon seems to have occurred in the 2-4-JMJD2D HACs. The presence of polycomb chromatin on the HAC may therefore explain why loss of the majority of HAC heterochromatin does not cause a increase in HAC transcription. If true, one can image that such a compensatory response could potentially silence otherwise disruptive transcription on centromeres that have suffered loss of heterochromatin. However, despite some reduction of HAC H3K27me3 levels after 5 days of JMJD2D tethering, its levels still remained well above background levels (when compared to other loci) (fig. 4.4A). This suggests that even if polycomb repression can provide some compensation in the absence of heterochromatin, in terms of transcriptional silencing \(^{499}\), such activity is not enough to prevent the HAC mitotic defects observed when heterochromatin is completely removed by JMJD2D.

I conclude therefore that HACs of 2-4-JMJD2D cells possess distinct chromatin from parental 2-4 HeLa cells, with the majority of its heterochromatin replaced by a repressive polycomb domain: a chromatin configuration which surprisingly does not disrupt
HAC centromere function or mitotic behaviour, as long as residual H3K9me3 levels are maintained. As such, it seems that for the majority of the HAC structure, the heterochromatin state is mostly dispensable, at least when replaced by polycomb chromatin. Nevertheless, I have determined that, at the very least, residual H3K9me3 levels at the HAC are indeed present, and their presence correlates with maintenance of HAC centromere stability and mitotic fidelity.

**Removal of HAC H3K9me3 its and associated phenotypes is ultimately reversible, but causes lasting centromeric defects**

Long-term binding of TetR-EYFP to the HAC by itself induced changes in CENP-A levels and a mild increase in mis-segregation. I hypothesized that perhaps the mis-segregation phenotype observed in HACs of 2-4-JMJD2D cells could be due to a synergistic effect between H3K9me3 removal and constant TetR binding. This implied that the consequences of heterochromatin removal, without having a TetR construct permanently tethered, might not be as severe. To confirm that low levels of H3K9me3 on the HAC are required for proper centromere function, in the absence of TetR binding, I decided to perform recovery assays. After tethering the constructs for a suitable period of time to allow for efficient H3K9me3 removal, I added doxycycline to prevent TetR binding, and followed the HAC to verify if there was recovery of H3K9me3 levels, centromere proteins and mitotic defects.

Firstly, to verify if heterochromatin removal from the HAC was permanent, constructs were tethered for 4 days, on both 2-4-EYFP-only and 2-4-JMJD2D cells, and then released from the HAC by addition of doxycycline. Tethering of TetR-EYFP by itself, as observed previously, caused some reduction in H3K9me3 levels, but after 5 days of recovery, H3K9me3 levels returned to initial values, showing that this effect is transient (fig. 4.5A). In HACs of 2-4-JMJD2D cells, where initial residual H3K9me3 was removed after 4 days of tethering, recovery for 5 days indeed allowed H3K9me3 levels to rise back to their
(low) initial state (fig. 4.5A). This confirms that low-level H3K9me3 was indeed still present at the HAC before tethering. Additionally, this suggests that these low levels of heterochromatin were being actively maintained, given that they return in the absence of JMJD2D tethering, even in the polycomb-rich conditions of 2-4-JMJD2D HACs (fig. 4.4A).

As the HAC can recover its peri(centromeric) heterochromatin levels over time, after their removal, I sought to discover if the HAC centromere and its mitotic fidelity could also recover from long-term JMJD2D tethering. After 8 days of JMJD2D tethering, when HAC CENP-A levels were shown previously to be severely depleted and mitotic fidelity severely decreased (fig. 4.2B, C), doxycycline was added back to the medium, and cells analysed after 2 or 4 days. Strikingly, HAC CENP-A levels began to recover, reaching their initial levels after 4 days of recovery (fig. 4.5B). Metaphase misalignment defects decreased to near initial values after only 2 days of recovery, while segregation fidelity progressively recovers, albeit much more slowly. Even after 4 days, when HAC CENP-A is back at initial levels, HAC mitotic fidelity has still not fully recovered (fig. 4.5C).

Collectively, these observations suggest that HAC kinetochores are quite robust, recovering the ability to orient properly just within 2 days (~2 cell cycles), after 8 days of heterochromatin removal, even while still not having recovered normal levels of CENP-A. Yet, despite CENP-A levels having recovered to initial values after 4 days of JMJD2D release, segregation fidelity is not yet fully restored at that point (fig. 4.5C). This indicates that long-term JMJD2D tethering causes lasting perturbations to the HAC centromere, which still persist somewhat after heterochromatin has been restored. Furthermore, this lasting effect on mis-segregation persists in the absence of TetR binding, thus excluding my initial hypothesis that TetR was contributing synergistically to the H3K9me3 removal phenotype. Finally, recovery of H3K9me3 at the HAC, following its removal by JMJD2D, correlates with the recovery of HAC CENP-A levels and mitotic fidelity, suggesting indeed that they might be causally related.
Fig 4.5 – JMJD2D-dependent removal of HAC heterochromatin and associated centromeric defects is reversible

A. Quantification of HAC H3K9me3 recovery, after JMJD2D release, in 2-4-EYFP-only and 2-4-JMJD2D cells. Doxycycline was washed out of the cell medium and cells were allowed to grow for 4 days; a fraction of these were then allowed to grow for 5 more days while another had doxycycline added to the medium to prevent JMJD2D binding, and also allowed 5 days to recover. On the penultimate day, cells were transiently transfected with TetON-tdTomato, to allow visualization of HAC under doxycycline conditions. Cells were fixed and processed for immunofluorescence microscopy. Mean HAC-associated H3K9me3 immunofluorescence signal was quantified, with subtraction of local background. Single experiment, n>37 interphase cells per condition, blue bar indicates median, red dotted line indicates background level. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

B. Quantification of CENP-A recovery after JMJD2D release, in 2-4-JMJD2D cells. Doxycycline was washed out of cell medium and cells were allowed to grow for 8 days; a fraction of these were allowed to grow for 2 more days, while another had doxycycline added to the medium to prevent JMJD2D binding, for 2 or 4 more days. After the respective recovery time had elapsed, doxycycline was washed from the medium to allow constructs to tether for 1h only, to allow HAC visualization, before fixation for immunofluorescence. Antibodies against CENP-A were used for detection. Single experiment, n>33 interphase cells per condition, blue bar indicates median. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001; except for the last timepoint (4 days of JMJD2D release) where CENP-A levels are similar to initial values: p=0.5049.

C. Quantification of HAC mitotic defects after JMJD2D release, in 2-4-JMJD2D cells (see C). Single experiment, metaphase defects n>41 metaphase cells; segregation defects n>101 telophase or early G1 cells. Fisher’s exact test was used to evaluate differences between results observed.
Loss of HAC heterochromatin over 4 days or more destabilizes HAC metaphase alignment and causes frequent mis-segregation

Because HAC mitotic defects are characterized by a failure to segregate chromatids, despite seemingly successful HAC congression and bi-orientation in metaphase, it seemed possible that the altered HAC centromere might still be able to make kinetochore-microtubule attachments, but fails to properly segregate due to perturbed kinetochore/centromere mitotic regulation. I decided to characterize this particular phenotype, to understand the possible mechanisms by which loss of heterochromatin may affect centromere dynamics in mitosis.

To understand how loss of heterochromatin is affecting the HAC's mitotic behaviour over time, I followed 2-4-EYFP-only and 2-4-JMJD2D cells after doxycycline washout, and analyzed HAC mitotic phenotypes at several time points, up to 16 days. I observed that JMJD2D-tethered HACs show progressively more severe segregation problems from 4 days of tethering onwards, to a maximum of around 55% mis-segregation at 8 days, with no further increase up to 16 days (fig. 4.6A). The predominant phenotype was that mis-segregating HAC chromatids failed to disjoin, and co-segregated to a single daughter cell (fig. 4.6B). A smaller fraction of HACs lag during anaphase, causing HAC chromatin bridges and micronuclei. As before, I observed few defects in metaphase, apart from a slight increase in misaligned and uncongressed HACs after a very long tethering period (13 and 16 days) (fig. 4.6A). Control TetR-EYFP-tethered HACs also show some minor defects from 6 days onward, though never higher than 20%, in either metaphase behaviour or segregation efficiency (fig. 4.6A). This indicates that long-term TetR binding to the HAC does indeed cause some minor mitotic defects, but still far from those induced by JMJD2D.

Remarkably, this experiment shows that the HAC centromere, when depleted of heterochromatin and despite suffering reduced levels of centromeric proteins, still seems
capable to correctly orient in metaphase and properly segregate half of the time, even after 16 days of heterochromatin removal.

My previous quantifications of HAC mitotic defects all derive from fixed microscopy samples. It is possible that transient behaviours of the HAC during metaphase might have been missed. To observe such transient dynamic behaviour of JMJD2D-tethered HACs, I performed microscopy imaging of live mitotic cells after 6 days of construct tethering, on both cell lines. I found that JMJD2D-tethered HACs indeed showed differences in metaphase behaviour: they were more likely to become transiently misaligned or suffer congression problems throughout the duration of mitosis than TetR-EYFP-tethered HACs, and those that did so most often mis-segregated at anaphase (fig. 4.6B, C).

Additionally, I observed brief stretching between HAC sister chromatids, in a fraction of 2-4-EYFP-only (16.67%) and 2-4-JMJD2D cells (27.03%) (fig. 4.6C, D). Cohesion between HAC chromatids did not seem to be lost, however, for this stretching seemed to be only transient, with the chromatids "snapping" back into bi-oriented tension afterwards (fig. 4.6D). In 2-4-EYFP-only cells, segregation still occurred normally after HACs exhibited this behaviour, thus ruling out any significant contribution to HAC mis-segregation. Close analysis of mitotic phenotypes of fixed cells, in the long-term tethering time course, revealed a small percentage of this phenomenon among mitotic HACs in both 2-4-EYFP-only and 2-4-JMJD2D cells, which increased in both cell lines to around 10% after 4 days of tethering, but no further (fig. 4.6E). There seems to be no difference between 2-4-EYFP-only and 2-4-JMJD2D cells in this respect. The fact that the effect increases with how long TetR-EYFP construct were tethered suggests that the effect is dependent on the tethering, and is not a behaviour inherent to the untethered HACs.

To address whether removal of heterochromatin was affecting HAC centromeric cohesion, I measured inter-kinetochore distance between the two HAC chromatids, 3 days after tethering (which ensures full HAC heterochromatin removal by this time). I found no
significant increase in overall inter-kinetochore distance (fig. 4.6F), in either cell line, regardless of tethering duration. This suggested that stretching events are indeed too transient to contribute to these measurements, and that for the most part HAC cohesion is preserved. All the above results therefore seem to imply that the stretching observed between HAC chromatids is TetR-dependent, the effect is transient, does not seem to significantly impair their cohesion and does not contribute significantly to HAC mis-segregation.

I have showed, therefore, that tethering of JMJD2D to the HAC, and consequent heterochromatin removal, impairs mitotic fidelity over time, by destabilizing the efficiency of HAC bi-orientation, leading often to mis-segregation. On the other hand, long-term TetR-EYFP tethering by itself can cause transient stretching between HAC sister chromatids in a small percentage of cells, but without significant effects on mitotic fidelity.
Fig 4.6 – Removal of HAC heterochromatin causes loss of stable orientation on the metaphase spindle and loss of segregation fidelity

A. Time-course of 2-4-JMJD2D cells for analysis of mitotic defects. Cells were washed of doxycycline and grown for several days, and samples were taken in intervals and fixed for microscopy visualization. Total of two independent experiments, metaphase defects n>80 metaphase cells each; segregation defects n>100 telophase or early G1 cells each. Numbers for stretched HAC chromatids were included under 'congressed and bioriented' due to their lack of effect on segregation efficiency (see D).

B. Imaging of live 2-4-JMJD2D cells 6 days after doxycycline washout. Images were acquired every 4 min.TetR-EYFP-JMJD2D (green), phase contrast (gray). Arrowheads indicate HAC sister chromatids: note changes in orientation and position relative to the metaphase plate.

C. Quantification of mitotic behaviour and segregation fate tracking of B. Total of two independent experiments, 2-4-EYFP-only cells n=36, 2-4-JMJD2D cells n=37. Chi-squared (two-tailed with 2 degrees of freedom) was used to evaluate the significance between the distribution of metaphase phenotypes, between correct segregation and defects (mis-segregation or mitotic delay). The distribution was analyzed for the following conditions: bioriented HACs ("bioriented" and "congressed not under tension"), stretched HAC chromatids and HACs showing defective phenotypes (misaligned, stretched and misaligned or uncongressed). For correct segregation results: chi-square value 0.497, n.s. p=0.7799. For defective segregation results: chi-square value 10248.706, ***p=0.0001.

D. Representative image of transient HAC interchromatid distance stretching, in 2-4-JMJD2D cells, as obtained in B.

E. Quantification of HACs with increased interchromatid distance, from A, as assessed by HAC-associated EYFP signal. HAC sister chromatids separated by 1μm or more were considered "stretched". Two independent experiments, error bars are standard deviation. Fisher's exact test was used to evaluate significance between differences observed.

F. Quantification of HAC interkinetochore distance, after 3 days of tethering, in 2-4-EYFP and 2-4-JMJD2D cells. Cells were fixed, metaphase bi-oriented HACs (only) were imaged, and distance between the two HAC EYFP chromatid spots was measured. Single experiment, n≥41 cells per conditions. Mann-Whitney U statistical test showed no significant differences between results.
Heterochromatin removal causes reduction of Aurora B on the HAC, but the HAC is remarkably resistant to loss of error-correction

The increase in HAC misalignment on the metaphase plate (i.e., congressed but not bi-oriented) and subsequent mis-segregation, induced by TetR-EYFP-JMJD2D tethering, can indicate that both HAC sister chromatids are attached to microtubules, but their orientation is aberrant. One explanation could be that one or both kinetochores are making merotelic attachments, and that the unbalanced pulling forces could cause HAC chromatid orientation to be disrupted and cause segregating HACs to lag in anaphase.\textsuperscript{653–655}

Disengagement of merotelic attachments has been shown to be dependent on the action of Aurora B, thus contributing to attachment error-correction.\textsuperscript{655–660} To test whether heterochromatin removal affects Aurora B recruitment to the HAC, I allowed constructs to tether for 6 days, in both 2-4-EYFP-only and 2-4-JMJD2D cells, before analysing mitotic HACs. TetR-EYFP-tethered HACs show clear Aurora B recruitment (fig. 4.7A), but in 2-4-JMJD2D cells Aurora B levels were much reduced on the HAC (fig. 4.7A). To quantify this, I arrested cycling cultures of both cell lines with the microtubule-depolymerizing agent TN-16 for 2h, after a 6-day tethering period. Prometaphase-arrested HACs of 2-4-JMJD2D showed a significant decrease in Aurora B at the HAC compared to those of 2-4-EYFP-only (fig. 4.7B), confirming that loss of HAC heterochromatin leads to Aurora B delocalization.

To verify whether JMJD2D-dependent loss of Aurora B at the HAC is sufficient to impair its error-correction capability, I performed a monastrol arrest-and-release assay after 6 days of tethering. Briefly, monastrol causes the cell to form a single microtubule aster, causing all centromeres in the cell to form attachments to one pole only (syntelic). Upon monastrol washout and separation of poles, chromosomes require the Aurora B-dependent error-correction pathway to release from syntely, so they can form new bipolar attachments. When Aurora B function is impaired, chromosomes are unable to recover from monastrol-induced syntelic attachments,\textsuperscript{446,663} making this a useful test for Aurora B function in error correction.
Cell lines 2-4-EYFP-only and 2-4-JMJD2D were washed of doxycycline and allowed to tether constructs to the HAC for 6 days, then arrested for 3h with monastrol, after which they were washed and incubated for 45min in fresh medium to allow the cells to reach metaphase alignment. As a positive control for loss of Aurora B activity, additional samples had the Aurora B-specific inhibitor ZM447439 added to the medium upon release from monastrol.

Surprisingly, I observed that the HAC, in both cell lines, was quite insensitive to loss of Aurora B activity (fig. 4.7C). At very short tethering times, release from monastrol in the presence of Aurora B inhibition, despite causing widespread congression problems for most chromosomes, did not noticeably hinder HAC congression and bi-orientation to the spindle midzone (fig. 4.7C). One potential explanation could be that the HAC makes fewer syntelic attachments under monastrol conditions in the first place. This could be due to its small size and circular chromosome configuration (which could confer additional structural rigidity), granting it a bias towards bi-orientation.

When TetR-EYFP by itself was allowed to tether for a longer period (6 days as opposed to 3h), together with Aurora B inhibition, then I could observe a slight increase in HAC metaphase alignment defects in comparison with TetR-EYFP control (fig. 4.7C). This implies that long-term tethering increases the HAC’s reliance on Aurora B-dependent error-correction. Because prolonged tethering of TetR-EYFP can cause the HAC chromatids to stretch apart from each other (fig. 4.6D, E, F), it may be possible that improved flexibility between the two HAC chromatids could lead to a higher probability of making syntelic microtubule attachments.

When JMJD2D was tethered to the HAC for 6 days, surprisingly, the HAC mostly congresses correctly after monastrol release (fig. 4.7C), despite the observed reduction in Aurora B levels (fig. 4.7B). When Aurora B is inhibited after monastrol release, 2-4-JMJD2D HACs after 6 days of tethering show slightly increased congression defects, although these are not significantly different when in the absence of ZM447439. This
indicates that residual levels of Aurora B still exist at the HAC, actively contributing to error-correction, after heterochromatin removal from the HAC (peri)centromere.

Therefore, although heterochromatin removal might lead to reduction of Aurora B from the inner centromere of the HAC, the fact that the HAC is not very reliant on error-correction together shows that it is unlikely that HAC Aurora B reduction, by itself, can explain the mis-segregation phenotype caused by JMJD2D tethering.
Fig 4.7 – Depletion of HAC heterochromatin causes delocalization of HAC Aurora B and loss of mitotic checkpoint signalling

A. Metaphase cells of 2-4-EYFP-only and 2-4-JMJD2D, after 6 days of doxycycline washout. Antibodies were used to detect Aurora B and DNA was stained with hoescht 333342 (blue). Scale bar = 5 µm.

B. Quantification of HAC-associated Aurora B signal, detected with antibodies as in A. Doxycycline was washed out from cell medium, and 6 days after the cells were arrested in prometaphase for 2h with TN-16, before fixation for immunofluorescence staining. Single experiment, blue bar indicates median, n>18 prometaphase arrested cells. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

C. Quantification of HAC error-correction capabilities after monastrol arrest. 6 days after doxycycline washout from cell medium, cells were arrested for 3h with 20 µM monastrol, and then washed out, allowing them to progress to metaphase for 45min. Cells were then fixed and stained for immunofluorescence microscopy with antibodies against α-tubulin (to confirm spindle recovery) and hoescht 333342 (to visualize DNA). As a positive control for loss of centromeric error-correction, 2 µM ZM447439 was added to cell medium after release from monastrol arrest, to inhibit Aurora B function (on all chromosomes). Total of 3 independent experiments, n=200 metaphase HACs each. Chi-square (2 tailed and with 2 degrees of freedom) test was used to evaluate statistical differences between results observed.

D. Quantification of mitotic duration, to visualize any mitotic checkpoint-induced arrest, on live cycling cells of lines 2-4-EYFP-only and 2-4-JMJD2D, 6 days after doxycycline washout. Images were acquired every 5 min. Total of two independent experiments, 2-4-EYFP-only n=39, 2-4-JMJD2D n=32. Blue bar indicates median. Mann-Whitney U statistical test showed no significant differences between results: p=0.6960.

E. Immunofluorescence staining of BubR1, in metaphase cells, 2-4-EYFP-only and 2-4-JMJD2D, 6 days after doxycycline washout. Antibodies were used against BubR1, CENP-C (to visualize kinetochores) and DNA was stained with hoescht 333342 (gray). Scale bar = 5 µm.

F. Quantification of BubR1 signal on prometaphase-arrested HACs, detected as in E. Doxycycline was washed out from cell medium, and 6 days after the cells were arrested in prometaphase for 2h with TN-16, before fixation for immunofluorescence staining. Lack of kinetochore-microtubule binding (caused by TN-16-dependent depolymerization of tubulin) should ensure all centromeres (including the HAC) activate mitotic checkpoint signalling. Total of two independent experiments, n=50 prometaphase cells each. Chi-square (2 tailed and with 2 degrees of freedom) test was used to evaluate statistical differences between results observed.
Loss of heterochromatin causes the HAC to lose its mitotic checkpoint signalling

Since heterochromatin-depleted HACs showed bi-orientation problems in metaphase, which did not seem to be due to critical loss of error-correction, the possibility remained that the problem was being caused by a reduced attachment to microtubules or mis-regulation of the attachment itself \(^{214,218,220,224,225,258,668-670}\). The latter could be due to a critical reduction in levels of kinetochore proteins, which could also explain the increase in uncongressed HACs. If so, unstable attachments would be expected to cause the HAC kinetochore to signal to the mitotic checkpoint and delay mitotic progression \(^{46,54,55,671-674}\).

To test whether heterochromatin-depleted HACs can correctly activate the mitotic checkpoint, I performed live microscopy imaging of cycling cells, of both 2-4-EYFP-only and 2-4- JMJD2D lines, after 6 days of doxycycline washout, and quantified the duration of mitosis. There was no delay in anaphase onset in 2-4-JMJD2D cells in comparison with control 2-4-EYFP-only cells, despite HAC mis-segregation in the former (fig. 4.7D). This suggests that either HAC kinetochores are stably attached to microtubules and satisfy the mitotic checkpoint (despite their subsequent mis-segregation), or that the checkpoint machinery of the HAC is impaired and thus unstable/incorrect attachments are not detected by the cell (increasing the chance that incorrect attachments will eventually cause HAC chromatid mis-segregation).

To distinguish between these possibilities, I performed immunofluorescence staining of mitotic checkpoint signalling protein BubR1 \(^{46,53,674-677}\), on both cell lines, after 6 days of tethering. I found that JMJD2D-tethered HACs had reduced BubR1 recruitment, evident in uncongressed (and thus not yet bi-oriented) HACs. To quantify this, I arrested 2-4-EYFP-only and 2-4-JMJD2D cells in TN-16 for 2h, to depolymerize microtubules and thus activate checkpoint signalling on all kinetochores. HACs tethered for 6 days with JMJD2D clearly show a significant decline in their BubR1 levels (fig. 4.7E, F), suggesting that these HACs might not be able to properly signal to the checkpoint and delay progression to anaphase.
Most BubR1-depleted HAC kinetochores after 6 days of JMJD2D tethering also show reduced levels of CENP-C (fig. 4.7E), suggesting that loss of a proper kinetochore recruitment platform might be the cause for insufficient checkpoint machinery. Despite that, this apparent minimal kinetochore still allows for a level of microtubule attachment and congression to the metaphase plate.

These results revealed that JMJD2D tethering to the HAC impairs effective mitotic checkpoint signalling from its kinetochore. If, in addition, the HAC’s kinetochore-microtubule attachments are in fact unstable, due to reduced levels of microtubule-attachment proteins, loss of its ability to activate the mitotic checkpoint can likely render it vulnerable to mis-segregation, as the cell will progress to anaphase without the HAC being able to correct/stabilize those attachments. However, to properly test this hypothesis, it will be necessary to specifically disrupt kinetochore-microtubule attachments at the HAC, under JMJD2D tethering, to observe whether or not there is any residual mitotic checkpoint signalling that can effectively prevent progression into anaphase.

**Heterochromatin removal causes destabilization and partial reduction of core centromere proteins and outer kinetochore proteins**

As shown before, JMJD2D tethering to the HAC causes it to lose centromere proteins, become less able to bi-orient on the metaphase plate, and ultimately be unable to generate a mitotic checkpoint signal, leading to HAC mis-segregation. A reduction of centromere/kinetochore proteins, enough to severely disrupt kinetochore functions but still allow a degree of microtubule attachment, could in principle explain this behaviour. To determine the extent and rate of centromere/kinetochore depletion, I quantified HAC levels of basal centromere proteins, in 2-4-JMJD2D cells, over time, by immunofluorescence microscopy. Both CENP-A and CENP-C levels are decreased by 4 days of tethering, reaching their lowest at 6 days, but then their levels stabilize and do not decrease further (fig.
4.8A, B). Extending the experimental timescale to 16 days showed no further change on CENP-A or CENP-C levels, or any evidence of a resistant subpopulation within the cell culture (fig. 4.8A, B). HACs tethered with JMJD2D for 6 days showed a marked loss of Hec1, part of the outer kinetochore KMN microtubule-attachment protein network, (fig. 4.8C), suggesting that attachment to microtubules may indeed be impaired.

Thus, while the HAC kinetochore is significantly compromised following heterochromatin removal, it is not completely lost. The presence of such an attenuated kinetochore could explain why the HAC is able to congress to the metaphase plate but its attachments and mitotic fidelity are apparently unstable.
Fig 4.8 –HAC heterochromatin removal reduces centromere and kinetochore proteins

A. Time-course quantification of HAC CENP-A, in 2-4-JMJD2D cells. Cells were washed of doxycycline and grown for several days, and samples were taken in intervals and fixed for microscopy visualization. Detection was performed with antibodies against CENP-A. Mean HAC-associated immunofluorescence signal was measured, and normalized to that of endogenous centromeres. Total of two independent replicates, n>22 interphase cells each. Blue bar indicates median. Mann-Whitney U statistical test showed significant differences between results: *** p<0.0001.

B. Quantification of HAC CENP-C in 2-4-JMJD2D cells, as in A, using antibodies against CENP-C. n>23 interphase cells each. Mann-Whitney U statistical test showed significant differences between results: * p=0.0158, *** p<0.0001.

C. Quantification of outer kinetochore protein Hec1 at the HAC centromere, in prometaphase cells. After 6 days of doxycycline washout, cells were arrested for 2h in TN-16, before fixation for immunofluorescence. Antibodies against Hec1 (magenta) were used for detection, and sum of HAC-associated Hec1 signal (background subtracted) was quantified. Blue bar indicates median, red line represents background level, n>14 prometaphase arrested cells. Mann-Whitney U statistical test showed significant differences between results: ** = p=0.0015. DNA was stained with hoescht 333342 (blue). Scale bar = 5 µm.
Heterochromatin removal on the HAC causes reduced assembly or fixation of new CENP-A, but does not affect HAC transcription

To understand how heterochromatin loss compromises the HAC centromere, I wished to determine whether the defect was due to destabilization of local CENP-A or to a flaw in assembly of new CENP-A. For this purpose I used the SNAP-tag protein labelling strategy to study CENP-A turnover and assembly. This method allows me to fluorescently label and visualize specific CENP-A populations in time, by transiently expressing CENP-A-SNAP in cells, and fluorescently labelling the SNAP-tag when desired.

To assess CENP-A stability, a pulse-chase of labelled CENP-A-SNAP was performed, and levels of fluorescent CENP-A-SNAP at the HAC centromere were assessed after 3 days of construct tethering (the point just before 2-4-JMJD2D HACs start showing significant CENP-A and CENP-C loss, fig. 4.9A). I observed that TetR-EYFP tethering for 3 days, by itself, causes a minor decrease in the levels of labelled CENP-A-SNAP remaining on the HAC, when compared to only 1h of tethering (fig. 4.9B). This indicates that TetR-EYFP binding to the HAC by itself increases CENP-A turnover. However, this loss of stability is unlikely to compromise centromere integrity in the long run, because I have shown that TetR-EYFP-tethered HACs present only a few mitotic defects and their total CENP-A levels are only slightly reduced. JMJD2D tethering, however, showed the same effect as TetR-EYFP control (fig. 4.9A, B). This indicates that JMJD2D tethering has no additive effects on CENP-A stability (beyond that of TetR binding). Consequently, this strongly suggested that JMJD2D impairs the HAC centromere not through destabilization of local CENP-A, but by interfering with assembly of new CENP-A.

To test the effects of JMJD2D tethering on CENP-A assembly, I performed a SNAP-labelling assay to measure incorporation of new CENP-A on the HAC, after 3 days of construct tethering. Using a quench-chase-pulse strategy which labels only newly-expressed SNAP-tag, I could visualize assembly of specifically newly-assembled CENP-A-SNAP at...
centromeres, after cells enter G1. The results showed that, while TetR-EYFP tethering by itself slightly reduces CENP-A assembly (which together with the loss of stability can account for the decrease in total CENP-A levels observed previously), tethering of JMJD2D severely reduces CENP-A assembly (fig. 4.9C, D). Given that total levels of CENP-A show significant decrease only from 4 days of tethering onward (fig. 4.8A), this drop in CENP-A assembly, at 3 days of tethering, strongly suggests that this is how JMJD2D-tethered centromeres progressively lose CENP-A.

I conclude that while TetR-EYFP tethering decreases HAC CENP-A stability (which had not been described before), the HAC centromere can seemingly compensate and persist, maintaining good mitotic fidelity. However, when heterochromatin is removed, assembly of new CENP-A (or its fixation at the centromere) is reduced, which explains the loss of other centromere and kinetochore proteins. My results therefore suggest that heterochromatin promotes the processes that allow CENP-A to assemble into centromeres, or aids in recruitment of CENP-A assembly factors.
**Fig 4.9** – TetR-EYFP tethering to the HAC increases CENP-A turnover, but JMJD2D tethering causes CENP-A assembly to be reduced.

**A.** Assessment of HAC CENP-A stability using a SNAP-tag pulse-chase assay. 2-4-EYFP and 2-4-JMJD2D cells were transiently transfected with a plasmid expressing CENP-A-SNAP; 24h later they were labelled with SNAP-binding TMR-Star and doxycycline was washed out. After 3 days, cells were fixed and mean HAC-associated CENP-A-SNAP signal was quantified (normalized to endogenous centromeres) by microscopy. Total of 2 independent experiments, n>25 each. Blue line represents median. Mann-Whitney U statistical test showed significant differences between results: * p=0.0277, ** p=0.0013, *** p<0.0001.

**B.** Representative images of A. Antibodies against CENP-C (blue) were used as reference for total centromere protein levels, and DNA was stained with hoescht 333342 (gray).

**C.** Quantification of new CENP-A assembly using a SNAP-tag quench-pulse-chase assay. 2-4-EYFP-only and 2-4-JMJD2D cells were wash ed of doxycycline and allowed to grow for 2 days, before transiently transfecting a plasmid expressing CENP-A-SNAP. 12h later, SNAP-tag inside the cells were labelled with BTP, and cells were chased for 9h to allow new CENP-A-SNAP expression, which was then labelled with TMR-Star. Cells were fixed and quantification was performed as in A. Mann-Whitney U statistical test showed significant differences between results: * p=0.0416, ** p=0.0014, *** p<0.0001.

**D.** Representative images of C. Antibodies against CENP-C (blue) were used as reference for total centromere protein levels, and DNA was stained with hoescht 333342 (gray).
**A**

CENP-A-SNAP stability

<table>
<thead>
<tr>
<th>Tethering duration</th>
<th>HAC A.T.U. normalized to endogenous centromeres</th>
</tr>
</thead>
<tbody>
<tr>
<td>1h</td>
<td>1.27</td>
</tr>
<tr>
<td>3 days</td>
<td>1.20</td>
</tr>
<tr>
<td>1h</td>
<td>1.20</td>
</tr>
<tr>
<td>3 days</td>
<td>1.20</td>
</tr>
</tbody>
</table>

**B**

TetR-EYFP

3 days of tethering; 3 days of CENP-A pulse-chase

**C**

2 days

Dox washout

Transfected CENP-A

SNAP

Quench SNAP with EYP

Label (new) SNAP with TMR-Star

New CENP-A-SNAP is synthesized, and assembled onto centromeres upon entry into G1

**D**

TetR-EYFP

3 days of tethering; CENP-A quench-pulse-chase 9h
3.3.3. Discussion

**Loss of HAC heterochromatin disrupts its centromere and mitotic behaviour**

I have shown that, by tethering the H3K9 demethylase JMJD2D to the alphoid\textsuperscript{TetO} HAC, in HeLa cells, heterochromatin on this artificial chromosome can be efficiently removed, affecting both H3K9me3 and H4K20me3 histone marks. Tethering of JMJD2D to the HAC seems to prompt a DNA damage signalling response after mitotic exit and leads to reduction of its centromere, by impairing the CENP-A assembly process. Other centromere and kinetochore proteins show significant reduction as well, likely due to their dependency on CENP-A for localization \[197-199\]. This includes CENP-C but also subunits of the microtubule-binding outer kinetochore KMN network (Hec1) and components of the mitotic checkpoint pathway (BubR1). Loss of these proteins can lead to reduced ability of the HAC to properly attach to microtubules as well as render it sensitive to mis-segregation if it cannot stabilize those attachments quickly enough, for it will not be able enforce a mitotic checkpoint. In summary, my results suggest that heterochromatin at (peri)centromeres contributes to stability of the centromere/kinetochore structure and function itself. This knowledge can prove useful when trying to understand how the myriad factors that contribute to mitotic fidelity are undermined in cancerous cells with defective heterochromatin regulation, and how they interact with known mitotic control pathways.

The HAC centromeres also revealed peculiar differences in mitotic responses, compared to endogenous centromeres. However, the HAC revealed little reliance on Aurora B-dependent error-correction, because of its apparently poor ability to make syntelic attachments, even under conditions which promote them on all other chromosomes. This peculiarity has allowed me to probe into underlying factors that contribute independently to mis-segregation, such as the inability of heterochromatin-depleted HACs to signal to the mitotic checkpoint, which may have been otherwise missed. Nevertheless, it should be taken into account that, unlike on the HAC, heterochromatin depletion from endogenous...
centromeres could possibly present a phenotype which impacts much more severely on centromeric error-correction, thus resulting in a slightly different (perhaps more penetrant) mitotic phenotype.

**Lack of disjunction of heterochromatin-depleted HACs: possible causes**

Since the majority of HAC segregation defects after JMJD2D tethering result in non-disjunction and co-segregation of HAC chromatids, it is plausible that in addition to weakened kinetochore-microtubule attachments, the HAC chromatids have difficulty in separating from each other. This could be caused either by a defect in cohesin release, or due to physical entanglement of chromatids themselves, such as chromatin bridges.

Upon anaphase onset, cohesin is ultimately released from chromosomes via proteolytic cleavage, catalyzed by separase \(^{682}\). Separase preferentially cleaves chromosome-bound cohesin, and this preference is promoted by Plk1-dependent phosphorylation of cohesin, in *S. cerevisiae* \(^{683,684}\). Plk1 \(^{685-687}\) is retained at kinetochores until anaphase, via the concerted action of the CPC and subunits of the mitotic checkpoint apparatus \(^{688-690}\). Given that I observe a clear loss of both Aurora B and BubR1, at JMJD2D-tethered HACs, it is possible that such HACs might suffer from inefficient cohesin cleavage, thus causing non-disjunction and explaining the subsequent mis-segregation.

On the other hand, defects in resolution of chromosomal bridges in anaphase can account for increased chromatid non-disjunction, independently of cohesin \(^{691-693}\). These entanglements have been shown to cause breakage of DNA during anaphase, which activate the DNA damage response locally, as cells enter G1 \(^{693-696}\). My observation, that heterochromatin-depleted HACs exhibit higher levels of DNA damage signalling as they enter G1, fits this hypothesis. I aim to support this evidence using γH2AX immunofluorescence staining, as an alternative marker for DNA damage signalling, which unlike 53BP1, can localize to mitotic chromosomes \(^{644}\), thereby allowing me to determine if the damage occurs during this time.
There are at least three demonstrated possible causes for sister chromatid entanglement in mitosis: lack of DNA decatenation, unresolved recombination intermediates after DNA repair via sister-chromatid exchange, or persistent under-replicated DNA hemicatenanes (ultra-fine bridges). Both unresolved DNA-repair recombination intermediates and excessive ultra-fine DNA bridges (which otherwise occur physiologically at low-level) can arise in the wake of replication fork collapse. Regions of highly repeated DNA, including centromeres, present significant replicative stress, suffering DNA double-strand breaks at a higher rate and are prone to be under-replicated.

Repetitive regions in particular seem to rely mostly on the SMC5/6 complex for the DNA damage response, in several organisms. During the DNA repair process, SMC5/6 promotes and resolves sister-chromatid homologous recombination, thereby protecting repetitive DNA from deleterious excisions and translocations between different chromosomes. Accumulation of deleterious recombinalional events due to absence of SMC5/6 eventually leads to inefficient mitotic disjunction of repetitive DNA, and DNA damage after mitosis, likely due to breakage. The SMC5/6 complex associates with heterochromatin in Drosophila and S. pombe, and loss of heterochromatin on repetitive DNA has been shown to increase the incidence of unresolved recombination intermediates. If indeed loss of heterochromatin on the HAC precipitates delocalization of SMC5/6, deregulation of repair-induced recombination and subsequent disjunction problems, then this could be one possible cause of HAC mis-segregation.

I have not yet been able to confirm whether it is lack of cohesin release, or replication stress and subsequent chromatid entanglement, which contribute to mis-segregation in JMJD2D-tethered HACs. Additionally, I also do not know whether any putative entanglement is caused by recombination defects (SMC5/6 dependent) or by DNA ultra-fine bridges due to under-replication (which are PICH and BLM-dependent). PICH and BLM themselves are recruited to centromeres. Given that JMJD2D-tethered HACs
show kinetochore disruption, localization of either of these proteins might be impaired, which could potentially cause the segregation defects observed.

**Flexibility of repressive chromatin at the HAC pericentromere?**

Heterochromatin at the HAC, at least at a residual level, seems to be surprisingly robust. My observation that H3K9me3 levels recover after 2 days, after their removal by JMJD2D, indicates that intrinsic factors promote H3K9me3 enrichment at the HAC α-satellite repeats. It would be interesting to speculate that centromeric activity could be the promoting factor for this activity, but without a non-centromeric region to compare to, or study whether H3K9me3 levels depend on centromere proteins, at this point it is highly uncertain.

Surprisingly, in 2-4-JMJD2D cells, HAC chromatin was already changed before experimentation, possibly due to unforeseen tethering events (even under doxycycline conditions) or an artefact of clonal selection during cell line establishment. This caused the HAC chromatin to become heterochromatin-poor, but polycomb-rich (two mutually exclusive types of chromatin). Such a change of chromatin state had been observed previously in mouse knock-out models which lack Suv39h1/h2, where heterochromatic regions convert to polycomb chromatin, particularly at the pericentromere. Cells of these mice showed elevated aneuploidy and a slight increase in pericentromeric satellite transcription. In 2-4-JMJD2D cells, the conversion of HAC heterochromatin to polycomb did not significantly affect levels of HAC centromere proteins nor its mitotic segregation fidelity. However it should be noted that, unlike the mouse cells mentioned, where H3K9me3 has been completely eliminated, this mark is not completely absent from the HAC in 2-4-JMJD2D cells, before tethering of the construct. Only after removal of residual, low-level H3K9me3 remaining on the HAC did I observe the defects described, highlighting the correlation of maintaining HAC heterochromatin with effective centromere function. However, to show that the events that lead to the chromatin signature of 2-4-
JMJD2D are reproducible, I will require independent confirmation that removal of heterochromatin from the HAC repeats can indeed prompt its subsequent enrichment with polycomb chromatin. Confirmation of this transition, in HACs of HeLa 2-4 cells (the HAC in its initial state) by transient transfection with TetR-EYFP-JMJD2D, can hopefully recapitulate the observed state of 2-4-JMJD2D HACs. If true, it would validate that such an transition occurs in vivo and that my studies in 2-4-JMJD2D cells are not invalidated as model for (peri)centromeric heterochromatin removal.

The replacement of heterochromatin with polycomb chromatin, on the majority of HAC repeats of 2-4-JMJD2D cells, surprisingly seemed to carry no observable deleterious effects. There are at least two ways of how polycomb could contribute to this, as a heterochromatin substitute: by dampening transcription on centromeres (as high-level transcription at the centromere is deleterious), or by promoting pericentromeric cohesin enrichment. Polycomb has been shown to associate with cohesin subunits and requires active cohesin to perform its silencing functions. This however does not necessarily mean that such cohesin contributes to sister-chromatid cohesion because, as mentioned before, cohesin loading onto chromatin does not necessarily imply that it envelops both sister chromatids. I have not yet analysed cohesin levels at the HAC: only interchromatid distance as a surrogate for effective cohesion. Given that mice without heterochromatin are viable, but suffer from mitotic and meiotic chromosomal instability, it begs the question: how genomically stable are the polycomb-rich HACs of 2-4-JMJD2D cells? Are satellite repeats protected from replication stress-derived abnormalities in the absence of heterochromatin? Evidence from mouse embryonic stem cells indicates that at least a fraction of repetitive mobile genomic elements are under polycomb repression. Additional confirmation of the genomic stability of 2-4-JMJD2D HACs, before JMJD2D tethering, is therefore required.
**Effects of long-term binding of TetR to the HAC**

With an established HAC cell line expressing TetR-EYFP, I was able to characterize the HAC centromere response to long-term tethering by TetR, which had not yet been performed thoroughly, remaining a standing question for our experimental system. I was able to show that indeed TetR tethering for longer than 2 days can slightly reduce HAC levels of centromere proteins, by destabilizing local CENP-A, thus reducing total kinetochore protein levels, and consequently leading to slightly lower levels of CENP-A assembly. Yet this assembly is efficient enough to counterbalance losses, and net CENP-A levels decrease only slightly. The HeLa cell cycle lasts ~22-23h. It is possible that CENP-A assembly during G1 can compensate for some loss in CENP-A stability, by "repopulating" the centromere with new CENP-A each cell cycle, up to a certain "baseline" level. If it can, such dynamics could possibly help add resiliency to centrochromatin in the face of perturbations. Tethering also generates a mild (less than 20%) increase in HAC mitotic defects, such as misalignment on the metaphase plate and mis-segregation, and transient interchromatid stretching in a small fraction of cells.

Overall, despite the measured HAC segregation defects being sufficient to generate appreciable HAC aneuploidy over a long-term culturing period, these defects are not enhanced further after 4-6 days. Therefore, 2-4-EYFP-only cells still remain a *bona fide* control for testing TetR constructs that can effectively cause severe HAC segregation defects. Only if a TetR-targeted factor has but a mild, but biologically significant, effect on the HAC, can this control's reliability be put into question. Furthermore, it is not beyond the realm of possibility that whatever changes TetR-EYFP causes on the HAC might have a synergistic effect with the fusion protein of interest. This would produce a HAC defect phenotype that might not be seen at all via action of the protein of interest by itself. The recovery experiments by doxycycline addition, despite having cleared TetR steric interference from the observed defects in 2-4-JMJD2D HACs, for example, do not exclude
long-lasting effects that TetR-binding (in a heterochromatin-free substrate) might have caused to the HAC.

### 3.3.4. Conclusion and outlook

In conclusion, I was able to show that critical levels of HAC heterochromatin correlate with effective centromere function. One possibility is that it might confer stability to the CENP-A assembly process, perhaps through enhancing recruitment of assembly factors, or aiding in the stabilization of newly-assembled CENP-A. Another interpretation could be that only the HAC centromere architecture was disrupted, with little initial loss of centromere proteins. This could subsequently reduce the efficiency with which new CCAN proteins are assembled into the structure over time. If true, a reduction in amount and density of CENP-C would be expected, which would eventually translate in less recruitment of the Mis18 complex and subsequent decrease in CENP-A assembly. Redistribution and reduction of the CENP-A domain has been reported before in *N. crassa* deletion mutants for heterochromatin components or in human cells after TSA treatment (which impairs heterochromatin silencing). Revisiting the early kinetics of centromere protein loss on HACs tethered with JMJD2D is required, as is verification of where in the pathway the assembly defect occurs: loss of Mis18 complex or HJURP recruitment, inefficient CENP-A incorporation or inability to "fix" new CENP-A in place. If confirmed that one of these steps is impaired by loss of heterochromatin (either directly or indirectly), it will allow us to understand better not only the process of CENP-A assembly (and possible steps of *de novo* assembly and neocentromere formation), but the nature of centrochromatin and its relationships with the rest of the genome.
4. Concluding remarks and future perspectives

4.1. Do centromeres require heterochromatin?

Centromeric activity in chicken cells can seemingly subsist in the absence of heterochromatin, and centromeres can apparently be enriched in polycomb chromatin instead. However, in my observations (and also in mice heterochromatin knock-out cells) complete depletion of heterochromatin can destabilize mitotic performance and lead to defects in meiotic recombination. This is accompanied by an increase in polycomb marks and proteins, but that does not seem to be enough to rescue mitotic and meiotic defects when heterochromatin is fully depleted. It should be noted these defects are observed in (peri)centromeres that are themselves located on repetitive DNA. As such, the defects observed may be mostly due to recombinational defects caused by repetitive DNA, and not from any positive effect that heterochromatin would have on the CCAN or CENP-A. A hypothesis can be made, in that centromeres not based on tandem repeats do not require heterochromatin, because the stability of their genomic environment does not depend on it. It would be interesting to test this hypothesis in chicken cells, by depletion of heterochromatin genome-wide, and asking if there is a difference between the segregation efficiency of heterochromatic and polycomb-based centromeres.

4.2. Is centromeric association with a repressive state necessary?

Is the alternative association with polycomb chromatin, which I observed in chicken in non-repetitive centromeres, a functional one? Do centromeres always require proximity to an environment of silenced chromatin? Given the prominent role that CENP-A plays in cell viability, its stability on chromosomes is paramount, and a more silenced chromatin state
(with a lower histone turnover) could contribute to that. Additionally, enrichment of cohesin on those silenced domains could aid in strengthening sister chromatid cohesion. However, I have shown in my work that CENP-A stability can be decreased without severely compromising its ability to reconstitute a full centromere in subsequent cell cycles. Additionally, increasing histone acetylation at human centromeres does not compromise its CENP-A levels, although it may instigate redistribution of CENP-A across the locus. While a silenced chromatin state surrounding centromeres is a conserved feature in eukaryotes, that could simply be a product of favouring centromere establishment in gene-free (and thus often repressed) region of the genome. The work of Shang et al has shown that centromeres seem to indeed be able to form anywhere on the chicken genome, but when formed on genes many of those events are selected out due to viability issues. Testing whether the stability of centromeres of chr. Z and 5 of chicken is dependent or not on the surrounding polycomb state would help answer this question.

At first glance, current understanding of cohesin regulation, and its accumulation on silenced regions, would argue in favour of centromeres requiring repressed chromatin to sustain an efficient pericentromere function. However, the dependency of sister-chromatid cohesion on heterochromatin has been contested in the past. On second thought, retention of cohesion at the pericentromere, in the absence of heterochromatin, can be explained by its substitution by polycomb chromatin (which may also recruit cohesin (fig. 5). Therefore, to fully answer the question of whether centromeres have any need for silent chromatin (for either core of pericentromere function), both polycomb and heterochromatin marks must be removed.
4.3. "Open chromatin" at the centromere: fighting against repression?

To date, many independent observations have been made regarding the distinct, transcription-related chromatin signature at centromeres, as well as derived transcripts, on several different organisms. However, while the presence of centromere transcripts (to a larger or lesser degree) seems to be conserved, there does not seem to exist a strict conservation of chromatin signature in this respect. Yet, the evidence comes from several different organisms, including human cancer cells. One common pattern that can be identified is the reduced level of repressive marks, and the apparent presence of open chromatin at centromeres.

So far, the open chromatin pattern at centromeres has been ascribed to the presence of active transcription. In human cells, this was shown directly through observation of both elongating RNA polymerase II and nascent transcripts. My observation of H3K4me3 at the centromere cores of human centromeres would certainly lend strength to this claim. Given that in vertebrates the transcription-associated factors RSF, CHD1 and FACT, as well as an open chromatin pattern, have been shown to be required for CENP-A assembly, the current model holds that the process of CENP-A incorporation is somehow transcription-dependent.

In this work, I have put forward the hypothesis that centromeres can resist silencing induced by repressive chromatin "bookmarks", such as H3K9me3 and H3K27me3. As the ultimate function of these marks is realized by the proteins they themselves recruit, if centromeres possess a mechanism that can remove or block them, that could provide an equivalent to a "barrier" function against repressive chromatin. Thus, in this way, vertebrate centromeres would generate their own "open chromatin" environment, thus preventing their inactivation by surrounding heterochromatin (fig. 5), in a comparable way to other eukaryotes, such as S. pombe.
Surprisingly, the general transcription initiation process can provide a possible explanation for this resistance. I have shown preliminary evidence that such a chromatin signature exists at human centromeres, however not in chicken centromeres. I have not yet verified whether the resistance to repressive "bookmarks" at the centromere depends on transcription initiation (or its chromatin signature). Future tethering experiments, through removal of transcription-related chromatin marks before EZH2/Suv39h1 tethering, can likely address this question.

4.3. Caveats and considerations of the $\alpha^{\text{TetO}}$ HAC tethering system

The $\alpha^{\text{TetO}}$ HAC is a unique system, that allows tethering of fusion proteins across an array of DNA repeats, a fraction of which contains the HAC's centromere. The system has been extensively used in the context of modifying the chromatin of the array, via tethering of chromatin regulators. However, there are several considerations that must be taken into account when interpreting experimental results, some of which I have addressed in this work.

The constitutive tethering of TetR, to a DNA repeat array, over the course of several cell divisions, had previously raised concerns that it could potentially affect HAC chromatin and centromere function. It had already been demonstrated that tethering could induce a small increase in HAC transcription $^{339}$. Here I not only confirmed this observation, but also demonstrated that long-term tethering of TetR-EYFP by itself can increase CENP-A turnover at the HAC centromere, and cause a slight decrease in its segregation fidelity. These defects are significant, but not very penetrant like those from bona fide centromere inactivators, such as KAP1$^{408}$ or LSD1$^{405}$. Therefore, TetR-EYFP can still be used as a negative control, at least for constructs that can severely impair centromere function.

Another consideration is the effect of overexpression of the TetR-EYFP construct (as a fusion with any given chromatin regulator) on the whole genome, which by itself could
de-regulate processes that are themselves required to sustain HAC centromere function and mitotic fidelity.

Finally, perhaps the most difficult caveat to tackle is that of causality of phenotypes. Tethering a chromatin regulator to the HAC can theoretically, in addition to TetR-specific effects (as discussed above), cause modifications to the HAC other than those expected, due to our limited knowledge of chromatin biology. These may involve effects on chromatin marks other than those which the fusion protein was believed to affect, but also possible direct effects on HAC centromere proteins. Controlling for this requires confirmation by non-tethering approaches, likely affecting the whole genome, which carries its own caveats. Although it is possible that observed HAC phenotypes may be caused by direct effects of the construct on centromere proteins, there is evidence that may render this unlikely. It has been measured that the majority of CCAN proteins turn over several times during a cell cycle, with mostly CENP-A and members of the CENP-T/W/S/X being unusual exceptions \(^{227,252,277}\). Thus, any modification of CCAN proteins induced by a TetR-EYFP construct would be unlikely to contribute significantly to centromere/kinetochore malfunction, unless this unsuspecting modification specifically affected CENP-A or the CENP-T/W/S/X complex. But perhaps the most significant piece of evidence against this possibility is the fact that defective HAC phenotypes take time to develop over several cell divisions of constant tethering. If the direct effect of a TetR-EYFP chromatin regulator on centromere proteins was the cause of any observed HAC mitotic defects, one would expect these to be fully observable in the first cell divisions, and not rise over time. As such, we can conclude that the TetR-EYFP constructs studied in this work are unlikely to have a direct negative effect on the mitotic performance of centromere proteins. However, there is one final consideration. If any putative modifications of centromere proteins affect not their mitotic functions, but their contribution to the CENP-A assembly process (which occurs only once per cell cycle), then indeed the possibility remains that the phenotypes observed are not caused by the changes in chromatin state. If true, this becomes a serious experimental
artefact that must be investigated, to segregate its contribution to the phenotype from any other chromatin-derived effects.

To circumvent the limitations outlined above, I propose the following solutions. To control for TetR-specific effects, one can use CENP-B$^{DBD}$-EYFP constructs, which use the DNA Binding Domain (DBD) of CENP-B to bind to the α-satellite repeats of every human centromere. This targeting approach circumvents the use of TetR (and any associated effects it may cause) and mimics the same level of recruitment, to the same chromosomal region, observed in the HAC. The trade-off is, however, that single-chromosome specificity (the HAC) is lost.

In order to control for effects of overexpression of TetR-EYFP-chromatin regulators, comparison of short and long-term tethering durations, for each construct, can be used. If overexpression of the chromatin regulator contributes to some degree of HAC mis-segregation, but if then in addition its direct tethering significantly increases these defects when compared to a non-tethered control, a direct effect on the HAC can be considered valid. Additional confirmation can be obtained by comparing levels of chromatin marks by ChIP, at the HAC and other genomic regions, with those in control cell lines, to verify if the untethered chromatin regulator has any effect.

Finally, to control for direct effects on centromere proteins caused by tethering of chromatin modifiers to the HAC, more conventional methods of affecting the chromatin state must be employed, such as RNAi of enzymes which maintain particular chromatin states, or use of small molecule inhibitors to those enzymes. Although more unspecific, if these methods result in similar phenotypes to those observed in the HAC, then it is possible to have greater confidence in the results yielded by this system.
Fig. 5: Preliminary proposed model of chromatin state balance at centromeres:

Centromere-driven "open chromatin" pushes against pericentromeric repression. Centromeres might subsist within regions of repressive chromatin by generating "open chromatin" through transcription initiation. Such a locus could potentially provide an anti-silencing effect, by counteracting encroachment of silencing marks.

On the other hand, by associating with Mb-long domains of repressive chromatin, centromeres might acquire a strong platform for cohesin enrichment, granting them robust sister-chromatid cohesion and additional structural constraints which may aid mitotic kinetochore geometry.
5. References


## 6.1. Chromatin state signatures

Distribution of various chromatin marks throughout annotated genomic features in the human genome.
6.2. HAC & CRaQ imageJ macro

```java
//########## HAC & CRaQ VERSION 1.00
//## based on CRaQ code version = "v1.06":
//########## //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //####### //#######
#
run("Set Measurements...", "area mean min center feret's redirect=None decimal=3");
run("Colors...", "foreground=black background=black selection=green");

if (isOpen("PRJ") == 1){
    selectWindow("PRJ");
    close(); }

waitForUser("Image", "Please open a .dv file with multiple Zplanes, and click OK.");

PrimaryImage=getTitle();
setTool("rectangle");
waitForUser("Select cell", "Please draw a square around a single cell to analyze.");
run("Duplicate...");
selectWindow(PrimaryImage);
close();

Dialog.create("Set Channels");
    Dialog.addMessage("\nMethod:
Projects .dv file by Max. intensity projection.\nSelect HAC signal and draws square around it, measures it, and paints it black. \nThen detects endogenous centromeres, and performs the same.\nMeasurements are made by measuring the mean signal in the square (to account for unfolding), and subtracting the lowest signal to control for local background signal.\nOutputs the data in a log window, and leaves the PROJection image open for confirmation of detection.\n\nAcquisition mode:", newArray("Z then wavelength(Fast Acquisition)", "Wavelength then Z"));
    Dialog.addMessage("\nDV files:");
    Dialog.addNumber("Data channel number",2,0,0,"");
    Dialog.addNumber("Reference channel number",2,0,0,"");
    Dialog.addNumber("DAPI channel number",4,0,0,"enter 0 if no DAPI was used");
    Dialog.addNumber("Total channels",4,0,0,"");
    Dialog.addMessage("\n\nChange default parameter settings?",0);
    Dialog.addCheckbox("Cropped cells?",0);
    Dialog.addChoice("Projection method", newArray("Max Intensity","Sum Slices"));
    Dialog.addMessage("\nAcquisition method:");
    Dialog.addCheckbox("Change default parameter settings?",0);
    Dialog.addCheckbox("Cropped cells?",0);
    Dialog.addChoice("Projection method", newArray("Max Intensity","Sum Slices"));
    Dialog.show();
    FileType=Dialog.getChoice();
    DataCh=Dialog.getNumber();
    RefCh=Dialog.getNumber();
    DapiCh=Dialog.getNumber();
    TotCh=Dialog.getNumber();
    Change=Dialog.getCheckbox();
    CroppedCells=Dialog.getCheckbox();
    ProjectionType=Dialog.getChoice();
    if (((RefCh-DapiCh)*(DataCh-DapiCh)) == 0){
        exit("Reference and Data channels should be different from DAPI channel");beep();}
    if ((RefCh-DataCh) == 0){
```
Dialog.create("WARNING");
    Dialog.addMessage("Reference and Data channels are the same");
    Dialog.show();
}

Dialog.create("Change parameter settings");
    Dialog.addMessage("Make sure that box is big enough to contain some background
pixels.");
    Dialog.addNumber("Square size",7,0,0,"pixels");
    Dialog.addMessage("Decreasing circularity will allow you to pick up more bright
centromeres, but will also increase the chance of picking up doublets, clusters or non-
centromeric regions.");
    Dialog.addNumber("Minimum Circularity",0.95,2,4,"a.u.");
    Dialog.addMessage("Max Feret's Diam. is the longest diameter of a spot. Increasing
will have similar effect to lowering Min. Circularity.");
    Dialog.addNumber("Max Feret's Diameter",6.5,1,3,"pixels");
    Dialog.addMessage("Min/Max Centromere size: lowering excludes more doublets,
but also some brighter cents.; increasing will discard more false positives but also true
positives.");
    Dialog.addNumber("Min Centromere Size",4,0,2,"pixel");
    Dialog.addNumber("Max Centromere Size",35,0,2,"pixel");
    Dialog.addMessage("Increasing Threshold will improve detection of dimmer spots
and decrease that of brighter ones.");
    Dialog.addNumber("Threshold Factor",1.0,2,4,"pixel intensity");
    Dialog.addMessage("If known, set the chromatic aberration of the reference channel
compared to the data channel, increase if Ref has signal more to top/right, or vice versa.");
    Dialog.addNumber("Chromatic aberration (horizontal)": ",0,0,2,"pixels to right");
    Dialog.addNumber("Chromatic aberration (vertical)": ",0,0,2,"pixels down");
if (Change == 1)
    Dialog.show(); //####################keeps defaults if "Change default"
is unchecked
    SquareSize=Dialog.getNumber();
    MinCirc=Dialog.getNumber();
    MaxFeret=Dialog.getNumber();
    MinCentro=Dialog.getNumber();
    MaxCentro=Dialog.getNumber();
    OtsuUp=Dialog.getNumber();
    xCor=Dialog.getNumber();
    yCor=Dialog.getNumber();
if (MinCirc >= 1) exit("Minimum circularity should be smaller than 1");
if (MinCentro >= MaxCentro) exit("Minimum centromere size should be smaller
than maximum centromere size");

corner=(SquareSize-1)/2;
roiManager("reset");
print("Log Reset");
print(\"Clear\");
Ch=newArray(RefCh,DataCh,DapiCh);
RDM=newArray("Ref","Data","Mask");
TotSl=nSlices;
ImageFolder=getDirectory("image");
ImageName=getTitle();
run("Properties...", "unit=pixel pixel_width=1
pixel_height=1");
run("Rename...", "title=dvFile");
if(nSlices>1)
    {
        if (FileType="Z then
            wavelength(Fast Acquisition") {Order = 'xyzct'; }
        else if (FileType="Wavelength
            then Z") {Order = 'xyct(default)'; }
        run("Stack to Hyperstack...",
            "order="+Order+," channels="+nSlices/TotCh+" slices="+nSlices/TotCh+" frames=1
display=Grayscale");
        waitForUser("Choose slice interval ", "Check which Z plane interval you want to analyze, and click OK.");
        Dialog.create("Set Z plane interval");
        Dialog.addNumber("First slice",1,0,0,"");
        Dialog.addNumber("Last slice",nSlices/TotCh,0,0,"");
        Dialog.show();
        FirstSlice=Dialog.getNumber();
        LastSlice=Dialog.getNumber();
        run("Z Project...",
            "start="+FirstSlice+" stop="+LastSlice+" projection=["+ProjectionType+"],
            "title="+ImageName+"Proj.tif");
        } else
            run("Duplicate...",
                "title=PRJ");
        run("Rename...", "title=PRJ");
        selectWindow("dvFile");
close();
for (k=0; k<Ch.length; k++)
    {
        if(Ch[k]>0)
            selectWindow("PRJ");
        setSlice(Ch[k]);
        run("Duplicate...", "title=+RDM[k]");
        run("Brightness/Contrast...");
        resetMinAndMax();
    }
selectWindow("PRJ");
setSlice(Ch[1]);
run("Channels Tool...");
run("Brightness/Contrast...");
resetMinAndMax();
setSlice(Ch[2]);
run("Channels Tool...");
run("Brightness/Contrast...");
resetMinAndMax();
print (ImageName);
print ("HAC signal");

measure();

function measure(){
    count=1;
    if(DapiCh>0 && CroppedCells==0){
        selectWindow("Mask");
        run("Duplicate...", "title=blur");
        run("Gaussian Blur...", "sigma=75");
        imageCalculator("Subtract", "Mask","blur");
        selectWindow("blur");
        close();
        run("Invert");
        getStatistics(AREA,MEAN,MIN,MAX);
        for(i=0;AREA>=getWidth*getHeight;i+=10){
            makeRectangle(0,0,0,0);
            setThreshold(MIN,MAX-i);
            for(j=0;j<getHeight;j+=100) doWand(0,j);
            getStatistics(AREA,a,b,c);
        }
        run("Convert to Mask");
        run("Fill Holes");
        run("16-bit");
        run("Multiply...", "value=257.000");
    }
    if(DapiCh>0 && CroppedCells==1){
        selectWindow("Mask");
        setAutoThreshold("Default");
        getThreshold(AAA,BBB);
        setThreshold(AAA, BBB*2/3);
        run("Convert to Mask");
        run("16-bit");
        run("Multiply...", "value=257.000");
        run("Invert");
    }
}

selectWindow("PRJ"); //##Shift back to the PROJ to select the HAC centromere on the Ref channel
setSlice(Ch[1]);
resetMinAndMax();
setTool("point");
waitForUser("Select HAC centromere", "Please click on the center of the HAC centromere signal");
//##Select HAC centromere.

    //##selects the signal area and adds it to ROI manager, on the Data image
    getSelectionCoordinates(x, y);
x=round(x[0]);
y=round(y[0]);
cx=x-xCor;cy=y-yCor; //##This is the chromatic aberration correction
selectWindow("Data");
makeRectangle(x-corner, y-corner, SquareSize, SquareSize);
getStatistics(area, no, no, no);
makeRectangle(cx-corner, cy-corner, SquareSize, SquareSize);
roiManager("Add");
getStatistics(no, mean, min, max);
if (min>0 && max<65000 && area==(SquareSize*SquareSize)){
    if (max>0)
        print (round(mean-min));
    else
        print ("ND");
    fillRect(cx-corner, cy-corner, SquareSize, SquareSize);  //!!!!!!!!!!!! puts black box over spots, these are then disregarded in the next cycle due to "if(min>0)"
    makeRectangle(cx-corner, cy-corner, SquareSize, SquareSize);
    print ("Endogenous centromeres");
}

//##Select Ref image to detect endogenous centromeres
selectWindow("Ref");
run("Bandpass Filter...", "filter_large=10 filter_small=1 suppress=None tolerance=5 autoscale");
if(DapiCh>0) imageCalculator("AND", "Ref","Mask");
run("Invert");
if(is("Inverting LUT")) run("Invert LUT");
run("MultiThresholder", "otsu");
getThreshold(lower, upper);
setThreshold(lower, upper*OtsuUp);
run("Analyze Particles...", "size="+MinCentro+"-"+MaxCentro+" circularity="+MinCirc+"-1.00 show=Nothing exclude clear");

//##Change to the Data image and go through the analyzed particles and draw rectangles around them, and measure them
selectWindow("Data");
for (l=0;l<nResults;l++) {
    if (getResult("Feret", l)<MaxFeret){
        x=round(getResult("XM", l));
y=round(getResult("YM", l));
cx=x-xCor;cy=y-yCor; //##This is the chromatic aberration correction
makeRectangle(x-corner, y-corner, SquareSize, SquareSize);
getStatistics(area, no, no, no);
makeRectangle(cx-corner, cy-corner, SquareSize, SquareSize);
getStatistics(no, mean, min, max);
if (min>0 && max<65000 && area==(SquareSize*SquareSize)){
    if (max>0)
        print (round(mean-min));
    else
        print ("ND");
    count++;
    fillRect(cx-corner, cy-corner, SquareSize, SquareSize);  //!!!!!!!!!!!! puts black box over spots, these are then disregarded in the next cycle due to "if(min>0)"
}
}
selectWindow("Data");
close();
selectWindow("Ref");
close();
if(DapiCh>0){
    selectWindow("Mask");
close();
}

waitForUser("Run again", "Do you want to run the macro again?");
rnnMacro("HAC & CRaQ_");

6.3. HAC Area quantifier ImageJ macro

//####starts up the windows and asks user for an image
run("Brightness/Contrast...");
run("Threshold...");
run("Channels Tool...");
run("ROI Manager...");
roiManager("reset")
run("Measure");
run("Clear Results");
run("Set Measurements...", "area mean min integrated redirect=None decimal=3");
waitForUser("Image", "Please open a .dv file with multiple Zplanes, and click OK.");

//++++++++ acquires the image title, and creates a dialog for user to input the number of
channels in this initial stack and the order of his channels as taken by the microscope, and
assigns the number to 'Channels' and HyperstackMode, respectively. Also asks how many
Zplanes to project, how to select the measurement area and what values to measure.//
Dialog.create("Channels and image order")
Dialog.addNumber("Number of Channels", 4)
Dialog.addChoice("Acquisition mode:", newArray("Z then wavelength", "Wavelength then
Z"))
Dialog.addNumber("Number of Zplanes to project (HAC 'height')", 5)
Dialog.addChoice("Projection mode:", newArray("Max Intensity", "Sum Slices"))
Dialog.addChoice("Area selection mode:", newArray("Thresholded to EYFP area", "Free
draw"))
Dialog.show()
Channels = Dialog.getNumber();
HyperstackMode = Dialog.getChoice()
NumberZplanes = Dialog.getNumber();
ProjectionMode = Dialog.getChoice();
AreaSelectionMode = Dialog.getChoice();

//++++++++ assigns the chosen mode of hyperstack order to 'Order' //
if (HyperstackMode=="Z then wavelength") {
    Order = 'xyzct';
}
else if (HyperstackMode=="Wavelength then Z") {
    Order = 'xyczt(default)';
}

//++++++++ counts the total images in the stack and assigns that number to 'Slices'.//
Slices = nSlices/Channels;
//######## runs 'Stack to Hyperstack' and uses the channel number to separate the images.
It automatically chooses 'order' as xyzt and 1 frame only.//
run("Stack to Hyperstack...", "order=+Order+" channels="+Channels+" slices="+Slices+" frames=1 display=Color");
ImageName=getTitle();
run("Rename...", "title=dvFile");

//##asks user which channels are which
Dialog.create("Identify Channels");
Dialog.addNumber("Select EYFP channel", 1);
Dialog.addNumber("Select data channel A", 3);
Dialog.addNumber("Select data channel B (optional, leave 0 if not required)", 0);
Dialog.show();
HACchannel = Dialog.getNumber();
Datachannel = Dialog.getNumber();
DatachannelB = Dialog.getNumber();
Stack.setChannel(HACchannel);

///selects the center slice of channel 1
SliceCorrection = round(((Slices-1) * 4) / 2) - 1;
setSlice(SliceCorrection);

//######## selects a specific channel, then applies a color to it. It does this for each channel
in this order: 1=Green, 2=Red, 3=Grays, 4=Blue.
Stack.setDisplayMode("color");
Stack.setChannel(1);
run("Brightness/Contrast...");
resetMinAndMax();
run("Green");
Stack.setChannel(2);
run("Brightness/Contrast...");
resetMinAndMax();
run("Red");
Stack.setChannel(3);
run("Brightness/Contrast...");
resetMinAndMax();
run("Grays");
Stack.setChannel(4);
run("Brightness/Contrast...");
resetMinAndMax();
run("Blue");
Stack.setChannel(HACchannel);

///wait for user, find center of signal to project a chosen number of Zplanes with that plane
as the center.
waitForUser("Select center of projection", "Please select the Z plane where your signal is
focused sharpest, and click OK.");
CenterSlice = round((getSliceNumber()/4)+1);
NumberZplanesUpandDown = round(NumberZplanes/2);
StartSlice = CenterSlice-NumberZplanesUpandDown;
EndSlice = CenterSlice+NumberZplanesUpandDown;
run("Z Project...", "start=\"+StartSlice+\" stop=\"+EndSlice+\" projection=[\"+ProjectionMode+\"]");
run("Rename...", "title=PRJ");
selectWindow("dvFile");
close();

//######## goes through each channel in the projection, opens the 'Brightness/Contrast'
window and resets the values to make the user's life easier. //
for(k=1;k<=nSlices;k++){
  setSlice(k);
  run("Brightness/Contrast...");
  resetMinAndMax();
}

//######### chooses the area selection and measurement modes //
if (AreaSelectionMode=="Free draw") {
  QuantifybyFreeDraw();
} else if (AreaSelectionMode=="Thresholded to EYFP area") {
  QuantifybyEYFPThresholdedArea();
}

//######### checks if user has required measurements for another channel and waits for
data collection before proceeding
if (DatachannelB>0) {
  Datachannel = DatachannelB;
  waitForUser("Collect your data", "Collect your data for Data Channel A, then press OK
to acquire from Data Channel B.");
} else if (DatachannelB==0) {
  exit();
}

//######## for datachannelB chooses the area selection and measurement modes //
if (AreaSelectionMode=="Free draw") {
  QuantifybyFreeDraw();
} else if (AreaSelectionMode=="Thresholded to EYFP area") {
  QuantifybyEYFPThresholdedArea();
}

//############################################################################FUNCTIONS
BEGIN
function QuantifybyEYFPThresholdedArea(){
  //#########This function creates an area thresholded to the EYFP HAC signal, and adds to
  the ROI the region in both HAC and data channels. Then the user has to drag the area
  around to select background regions, by clicking 'Add' on the ROI manager or pressing 't'.
  Then it measures all regions into the results window.
  //##automatically tries to set the treshold correctly to select only the signal area. User can
correct and click OK.
  Stack.setChannel(HACchannel);
  Stack.setDisplayMode("Grayscale");
  setAutoThreshold("MaxEntropy dark");
  waitForUser("Adjust threshold to HAC", "Please adjust the threshold level to your HAC signal
  area, and click OK.");
  run("Create Selection");
  //##selects the signal area and adds it to ROI manager, both on the EYFP channel and the
data channel
  Stack.setChannel(HACchannel);
  run("ROI Manager...");
```javascript
//##user moves selection to add more backgrounds, click OK afterwards
waitForUser("Add background regions", "Please move the selection to a background area and click add on the ROI manager, or press 't'. Repeat three times. Then click OK.");
roiManager("Deselect");
run("Set Measurements...", "area mean center redirect=None decimal=3");
roiManager("Measure");
roiManager("Delete");

print("Log Reset");
print("Clear");
print(ImageName);
print("Mean, HAC");
print(round(100*getResult("Mean", 1)));
print("Min signal in HAC area");
print(round(100*getResult("Min", 1)));
print("Mean, 3 backgrounds");
print(round(100*getResult("Mean", 2)));
print(round(100*getResult("Mean", 3)));
print(round(100*getResult("Mean", 4)));
print("Sum in area, HAC");
print(round(100*getResults("Area", 1)*getResult("Mean", 1))); 
print("Sum in area, 3 backgrounds");
print(round(100*getResults("Area", 2)*getResult("Mean", 2))); 
print(round(100*getResults("Area", 3)*getResult("Mean", 3))); 
print(round(100*getResults("Area", 4)*getResult("Mean", 4))); 
```
run("Brightness/Contrast...");
resetMinAndMax();

//##user moves selection to add more backgrounds, click OK afterwards
waitForUser("Add background regions", "Please move the selection to a background area and click add on the ROI manager, or press 't'. Repeat for as many as you wish. Then click OK.");
roiManager("Deselect");
run("Set Measurements...", "area mean min center redirect=None decimal=3");
roiManager("Measure");
roiManager("Delete");

print("Log Reset");
print("\nClear");
print (ImageName);
print ("Mean, HAC");
print (round(100*(getResult("Mean", 1))));
print("Min signal in HAC area");
print(round(100*(getResult("Min", 1))));
print("Mean, 3 backgrounds");
print( round(100*(getResult("Mean", 2))));
print( round(100*(getResult("Mean", 3))));
print( round(100*(getResult("Mean", 4))));
print("Sum in area, HAC");
print ( round( 100*( ( getResult("Area", 1)*getResult("Mean", 1) ) ) ));
print("Sum in area, 3 backgrounds");
print ( round(100*(getResult("Area", 2)*getResult("Mean", 2))));
print ( round(100*(getResult("Area", 3)*getResult("Mean", 3))));
print ( round(100*(getResult("Area", 4)*getResult("Mean", 4))));

//###########################################################################FUNCTIONS
END
6.4. Appended collaboration article

Shang et al 2013 – Chromosome engineering allows the efficient isolation of vertebrate centromeres.

(overleaf)