



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„The conformation of replicated human chromosomes“

verfasst von / submitted by

Michael Mitter, B.Sc. MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 794685490

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on
the student record sheet:

Molekulare Biologie

Betreut von / Supervisor:

Dipl.-Biol. Dr. Daniel Gerlich

Contents

1	Introduction	1
1.1	DNA and its higher order structure	1
1.1.1	The nucleosome and the 30 nm chromatin fiber	1
1.1.2	Higher order chromosome structure	2
1.2	Factors shaping chromatin structure	5
1.2.1	Histone PTMs	6
1.2.2	Polycomb group proteins	8
1.2.3	SMC complexes and loop extrusion	9
1.2.4	CTCF	12
1.3	Functional importance of DNA structure	13
1.3.1	Gene expression	14
1.3.2	DNA repair	15
1.4	DNA structure throughout the cell cycle	17
1.4.1	Mitosis	20
1.4.2	S-Phase and sister chromatid cohesion	22
1.5	Open questions in sister chromatid biology	24
1.5.1	Methodological short-comings	25
1.6	Methods to describe DNA structure	26
1.6.1	Electron microscopy	26
1.6.2	FISH	26
1.6.3	Chromosome conformation capture techniques	27
1.7	Aim of this thesis	27
2	Methods	29
2.1	Cell culture	29
2.2	Generation of cell lines	29
2.3	FACS analysis of cell cycle stage	30
2.4	DNA damage assay	30
2.5	Viability assay	31
2.6	Western Blot	31
2.7	Metaphase congression assay	31
2.8	Cell synchronization for scsHi-C	32
2.9	Hi-C sample preparation	32
2.10	Sequencing	33
2.11	Quantification of 4sT incorporation into gDNA	33
2.12	Conversion analysis of 4sT on synthetic oligos	33
2.13	Conventional sequencing library preparation to estimate 4sT mutation rates	34
2.14	Iodoacetamide mediated conversion of 4sT	34
2.15	OsO ₄ / NH ₄ Cl-mediated conversion of 4sT	34
2.16	Melting curve analysis of hairpin oligo	35
2.17	Synthesis of a 4-thiothymidine (4sT) phosphoramidite building block	35
2.18	Solid-phase synthesis of 4sT containing DNA	36

2.19	Deprotection of 4sT containing DNA	36
2.20	Analysis and purification of 4sT containing DNA	37
2.21	Sequencing data analysis	37
2.21.1	Calling HeLa single-nucleotide polymorphisms	37
2.21.2	Mutation rate analysis	37
2.21.3	Correlation analysis of loci splitting frequency and FISH with scsHi-C	38
2.21.4	Hi-C data preprocessing	38
2.21.5	Hi-C genome scaling plots	39
2.21.6	Observed-over-expected transformation of Hi-C matrices	39
2.21.7	Hi-C aggregate maps at TAD-centers	39
2.21.8	Extraction of sample regions	39
2.21.9	TAD-calling	40
2.21.10	Pairing-score and contact-density calculation	40
2.21.11	Stack-up analysis of line profiles	40
2.21.12	LOLA-analysis of highly paired and highly unpaired TADs	41
2.21.13	HiCRep analysis	41
2.21.14	False-positive rate estimation of double labelled reads and trans-sister contacts	41
2.21.15	H3K27me3 enrichment analysis	42
2.21.16	<i>Prediction of pairing score</i>	42
2.21.17	<i>Generation of average trans-sister classes</i>	43
2.21.18	<i>ORI pileups</i>	43
2.21.19	<i>Saddle plot analysis</i>	43
3	Results	45
3.1	Concept of a sister chromatid sensitive Hi-C approach	45
3.2	4sT is a novel nucleotide label that allows sequencing based de- tection	47
3.2.1	4sT incorporation rate	48
3.2.2	4sT toxicity	49
3.2.3	Iodoacetamide mediated conversion of 4sT	51
3.2.4	OsO ₄ -conversion allows quantitative detection of 4sT	55
3.2.5	Accounting for HeLa SNPs increases signal-to-noise ratio of 4sT detection	56
3.3	A synchronization scheme for sister chromatid sensitive Hi-C	57
3.3.1	Thymidine/aphidicoline block for G1/S synchronization	57
3.3.2	4sT can be used in the proposed synchronization scheme	58
3.3.3	Optimization of synchronization schemes for different cell cycle stages	59
3.4	Quality control of scsHi-C	61
3.4.1	Feasibility of sister chromatid-sensitive Hi-C	61
3.4.2	Confidence of sister chromatid assignment	63
3.4.3	Biological controls for scsHi-C	65
3.5	Large scale organization of sister chromatids	67
3.5.1	Large scale organization in the G2 cell cycle stage	67

3.5.2	Large scale organization in prometaphase	70
3.6	TADs as units of trans-sister organization in G2	73
3.6.1	New metrics to analyze scsHi-C data	75
3.6.2	CTCF is enriched at trans-sister contact sites	76
3.6.3	TAD-boundaries are contact sites between sister chromatids	77
3.7	H3K27me3 determines trans-sister TAD contact state	80
3.8	Influence of Sororin on chromatin structure	84
3.8.1	Sororin degradation leaves cis-sister structures intact	85
3.8.2	Sororin is required for sister chromatid alignment in G2	87
3.9	Loop extrusion is required for local separation of sister chromatids	88
4	Discussion	93
4.1	4sT is a new DNA label with many applications	93
4.1.1	Potential applications	94
4.2	scsHi-C: Sister chromatid sensitive chromatin structure analysis	95
4.2.1	Strand-specific labelling allows high-yield, low-noise sister chromatid Hi-C	95
4.2.2	Principle of strand-specific labelling is transferable	96
4.3	The role of sister chromatid alignment in G2	97
4.3.1	Sister chromatid alignment and DNA damage repair	97
4.4	Implications of TAD-level organization	98
4.4.1	Consequence for gene expression	98
4.5	Functional role of trans-sister contact heterogeneity	99
4.5.1	Mechanisms that determine trans-sister TAD state	99
4.5.2	Functional implications	100
4.6	Sororin is necessary for sister chromatid alignment in G2	101
4.6.1	Role of sororin in gene expression regulation and HDR	101
4.7	Influence of loop extrusion on trans-sister contacts	102
4.8	Conclusion and outlook	103
5	References	104
6	Supplementary Material	120
6.1	Abbreviations	120
6.2	Supplementary Tables	122

Acknowledgements

In current times, every scientific undertaking is not the achievement of a single individual, but the result of extensive collaboration within and between laboratories. I thus want to thank the entire Gerlich laboratory for providing a great scientific atmosphere, many fruitful discussions and help in performing key experiments. Here, I especially want to thank Claudia Blaukopf for keeping the lab extremely organized and helping me with the proliferation and toxicity experiments shown in Fig. 3.4a and b. In addition, I want to thank Zsuzsanna Takacs for helping me prepare scsHi-C samples and interesting discussions about the conformation of sister chromatids. I also want to thank Christoph Langer, who engaged in many important discussions about data analysis and automated our scsHi-C preprocessing steps, thus saving us tremendous amounts of time. I further want to thank Paul Batty for critical reading of the manuscript and many interesting discussions. Also, I want to thank Manuel Alonso Y Adell, Isma Bennabi, Ana David, Sara Haering-Cuylen, Rudi Hoefler, Mina Petrovic, Matthias Samwer, Max Schneider, Christoph Sommer and Federike Teloni for many interesting discussions and valuable input. Finally, I also want to thank Daniel Gerlich, for allowing me to work on this - sometimes risky - project and guiding me scientifically in the past four years.

This project would not have been possible without the sophisticated conversion chemistry for 4sT developed by the lab of Ronald Micura at the University of Innsbruck. Here, I want to especially thank Catharina Gasser and Eva Neuner, who characterized the 4sT conversion, performed the first test conversions for me and taught me how to do this experiment myself. In addition, I want to thank Ronald for fruitful discussions and willingness to help us get the conversion chemistry to work. Experiments performed by the lab of Ronald Micura are shown in Fig. 3.7.

I also want to thank the laboratory of Jan-Michael Peters, especially Gregor Jessberger and Wen Tan, for providing me with the NIPBL-AID and Sororin-AID cell line. Additionally, I want to thank Jan-Michael Peters for fruitful discussions.

Additionally, I want to thank my thesis committee - Kikuë Tachibana-Konwalski, Jan-Michael Peters and Daniel Gerlich - for great input and stimulating discussions.

Then, I also would like to thank the VBC core-facilities, especially Thomas Köcher for helping me with 4sT mass spectrometry measurements and the NGS facility for swift sequencing and helpful discussions.

In addition, I want to thank the VBC PhD-program, especially Inês Crisóstomo and Christopher Robinson, for providing a great atmosphere and a sense of community on campus.

Then, I want to thank the VBC summer selection 2016 for being a great group of people and helping me get through difficult times.

Finally, I want to thank Viktoria for reading this manuscript, for providing constant support during these four years and being there for me, when things did not go as planned.

To Viktoria.

Abstract

German

Die dreidimensionale Konformation von Chromatiden hat einen enormen Einfluss auf eine Vielzahl verschiedener Aspekte der Zellbiologie, wie Genexpression, Chromosomensegregation und DNA-Reparatur. High-Throughput-Chromosom-Conformation-Capture (Hi-C) hat es Forschern ermöglicht, diese Konformation im Detail zu untersuchen, und hat zur Entdeckung zahlreicher Prinzipien der Chromosomentopologie geführt. Hi-C ist derzeit jedoch nicht in der Lage, replizierte Chromosomen zu unterscheiden, so dass Forscher Schlüsselaspekte der Genomfunktion nicht untersuchen können. Dazu gehören die Frage, wie DNA-Reparatur das Kopieren von Information des Schwesterchromatids koordiniert und wie die Resolution der Schwesterchromatiden während der Mitose mit deren Kohäsion koordiniert wird. Um dieses methodische Rätsel zu lösen, verwendete ich 4-Thio-Thymidin, ein neuartiges Nukleotidanalogen, das durch Sequenzierung nachweisbar ist, um die Schwesterchromatide differenziell zu markieren. Dies ermöglichte es mir, eine Hi-C-Variante zu entwickeln, die in der Lage ist, Schwesterchromatide zu unterscheiden. Mit dieser Methode habe ich dann die ersten 3D-Genomkarten, die Schwesterchromatide unterscheidet, erstellt, wobei sich zeigte, dass die Trans-Schwester-Interaktionen eine bemerkenswerte Heterogenität aufweisen. Ferner zeige ich, dass die Schwesterchromatide an den Grenzen topologisch assoziierender Domänen zusammengehalten werden und dass Trans-Schwester-Interaktionen mit dem epigenetischen Zustand von genomischen Regionen korrelieren. Schließlich verwende ich Perturbationen des Cohesin-Komplexes, um zu zeigen, dass zwei Klassen von Cohesin die Schwesterchromatide unterschiedlich organisieren: Kohäsives Cohesin hält die Schwesterchromatide zusammen, während schleifenextrudierendes Cohesin sie auf einer kleinen genomischen Skala trennt. Zusammenfassend lässt sich sagen, dass scsHi-C eine neuartige Technologie ist, die schwesterchromatidenspezifische 3D-Strukturuntersuchungen ermöglicht und ich diese Technologie benutzt habe, um den Einfluss von Cohesin auf diese Konfiguration zu untersuchen.

English

The three-dimensional conformation of chromatids has tremendous impact on a variety of different aspects of cell biology such as gene expression, chromosome segregation and DNA repair. High-throughput chromosome conformation capture (Hi-C) has enabled researchers to examine this conformation in great detail and has led to the discovery of numerous principles of chromosome topology. However, Hi-C is currently not able to distinguish replicated chromosomes, thus precluding investigators from examining key aspects of genome function. These include how DNA repair coordinates copying information from sister chromatids and how sister chromatid resolution during mitosis is coordinated with sister chromatid cohesion. In order to resolve this methodological conundrum, I used 4-thio-thymidine, a novel nucleotide analog that is detectable by sequencing, to mark sister chromatids differentially. This enabled me to develop a Hi-C variant, which I name sister chromatid sensitive Hi-C (scsHi-C), that is capable of distinguishing sister chromatids. I then used this method to generate the first sister chromatid resolved 3D-genome maps, revealing that trans-sister interactions exhibit remarkable heterogeneity. I further show that sister chromatids are held together at the boundaries of topologically associating domains and that trans-sister interactions correlate with epigenetic state. Finally, I use perturbations of the cohesin complex to show that two classes of cohesin organize sister chromatids differentially: Cohesive cohesin holds sister chromatids together, whereas loop extruding cohesin separates them on a small genomic scale. In conclusion, scsHi-C is a novel technology that allows sister chromatid specific 3D-structure examinations that I showcased by examining the influence of the cohesin complex on this configuration.

1 Introduction

1.1 DNA and its higher order structure

The human genome codes for roughly 20,000 genes and an even larger number of non-coding transcripts, along with the regulatory sequences that are needed to activate transcription of these units at appropriate time points [1, 2]. This complexity is encoded on $3.2 \cdot 10^9$ bp of DNA that together stretch out to roughly 2 m of DNA that each cell needs to fit inside its nucleus. A typical human cell measures roughly 100 μm in diameter suggesting that the genetic material needs to be compacted 20,000-fold just in order to fit inside. To complicate this matter, the DNA fiber is highly negatively charged due to its phosphodiester bonds, and thus opposes compaction by electrostatic repulsion.

But fitting a large bio-molecule into a restricted space is not the only problem that DNA folding needs to solve: The information stored in the genome needs to be readily available to ensure proper gene expression, imposing the need for a tight order of the 3D-genome structure. Additionally, genetic information needs to be read dynamically, with different regions being accessible at different time points. For example, during development of multicellular organisms, different genes need to be switched on sequentially with their genetic information readily accessible. Additionally, when cells react to environmental queues such as heat, they must access a set of specialized genes that would otherwise not be available. Thus, the challenge of genome folding is two-fold: Not only must enormous DNA molecules be fit into a tiny cell nucleus, but this organization also needs to follow a tight order to ensure that the genetic information can be read at the appropriate time-points and undergo dynamic changes.

1.1.1 The nucleosome and the 30 nm chromatin fiber

The first level of compaction in eukaryotic cells is achieved by wrapping DNA around nucleosomes. Nucleosomes are octameric protein complexes that consist of two copies each of the four core-histones: H2A, H2B, H3 and H4 (Fig. 1.1a) [5]. These core histones are highly conserved throughout eukaryotes and consist of a globular histone-fold domain that makes up the bulk of the histone and a highly flexible and unstructured tail that is thought to be a regulatory hub for gene expression (see 1.2 for more details) [6, 7]. Each core-histone is highly positively charged and therefore can compensate for some of the negative charge of DNA, thus facilitating compaction. Specifically, the nucleosome level compaction is realized by wrapping DNA in a left-handed helix around the disk-like histone octamer, on average 1.67 times, which partitions the genome into DNA that is engaged with nucleosomes and so-called "linker-DNA" [5]. The resulting structure is often called a "beads-on-a-string" organization or referred to as the 10 nm chromatin fiber [8, 9]. While this process does compact DNA, it barely scratches the surface of what needs to be achieved to enable efficient packaging of the genetic material.

The 30 nm chromatin fiber has historically been a way to explain the next

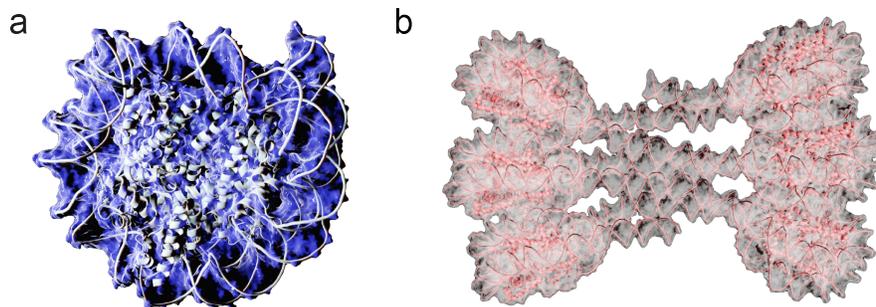


Figure 1.1: **Structure of the nucleosome and nucleosome arrays.** (a) Crystal structure of the nucleosome. Surface representation in blue and cartoon inlay in white. Drawn from PDB entry 2CV5 [3]. (b) Crystal structure of a 6-mer nucleosome array with the linker Histone H1 showing a two-start helix. Surface representation in white, cartoon inlay is shown in red. Drawn from PDB entry 6HKT [4]

level of compaction. This structure has been obtained for isolated chromatin fragments and artificial chromatin arrays harboring strong nucleosome positioning sequences *in vitro* under certain buffer conditions [10, 11]. Depending on different parameters such as the spacing of nucleosomes and whether or not the linker histone H1 was present, two different arrangements of the 30 nm chromatin fiber were inferred from the data: A solenoid organization where the 10 nm fiber was arranged in a one-start helix and a "zig-zag" organization, where the 10 nm fiber was arranged in a two-start helix (Fig. 1.1b) [12]. Between the two models, more convincing structural evidence has been accumulated for the latter organization, with a recent cryo-EM study resolving a large array of artificial chromatin to 11 Å, exhibiting a "zig-zag" organization [13–15]. However, the variability observed *in vitro* already hints at the fact that the organization might be more complex and diverse inside living cells. Indeed it was shown by small-angle X-ray scattering (SAXS) as well as by electron microscopy that an ordered 30 nm fiber likely does not exist *in vivo* [16, 17]. Instead, these studies and others suggest that the 10 nm fiber is the largest ordered assembly *in vivo*, which is then arranged in heterogeneous ways to achieve higher order compaction [18].

1.1.2 Higher order chromosome structure

The higher order folding of chromosomes has sparked enormous research interest in recent years, not only because its inherent complexity and heterogeneity is intriguing, but also because it is starting to become apparent that it has tremendous influence on a multitude of biological functions (see 1.3).

Historically, it was believed that hierarchical, higher order compaction of the 30 nm fiber into large loops leads to an increasing compaction of the gen-

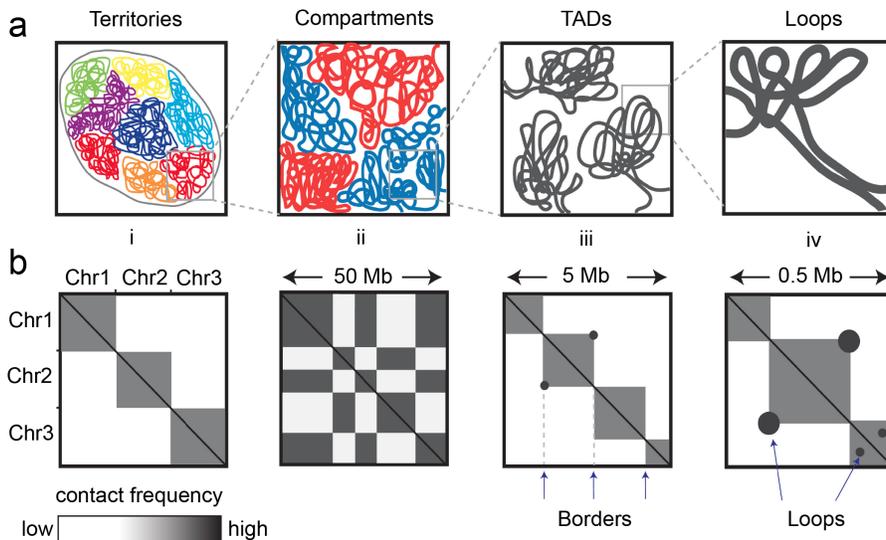


Figure 1.2: **Higher order chromatin folding.** (a) Chromatin folding at different length scales from chromosome territories (i), compartments (ii), TADs (iii) to chromatin loops (iv). (b) Hi-C patterns that are thought to be signatures of the chromatin folding motifs in panel (a). Figure adapted from Fig. 1 of [29]

ome culminating in the highly compacted mitotic chromosome (see 1.4.1) [19]. While the specifics of this model have been challenged in the past, recent high resolution chromosome structure analyses do suggest a "pseudo-hierarchy" of higher order genome folding, where the genome is organized into elements of different length-scales [20] (Fig. 1.2).

Chromosome conformation capture techniques (see 1.6.3) were a revolution for the analysis of genome structure and have enabled significant insight into the higher order structure of chromosomes in recent years [21–28]. These techniques yield a measurement of contact frequency between different genomic locations (called "contact maps"; see Fig. 1.2b). This means that these techniques do not measure distance between different parts of the genome, but rather how often they are in contact. Chromosome conformation capture methods allow both experiments examining small-scale details of nucleosome neighborhoods [26] as well as studies probing entire chromosomes [27, 28], thus bridging the gap between nucleosome level *in vitro* experiments and large scale microscopy studies.

The smallest structures that can be observed in contact maps are so-called "loops", which can be as small as a few kilobases (kb), but may range up to several megabases (Mb) in size [22]. It should be noted that the nomenclature regarding loops has caused some confusion in the Hi-C literature: Loop is often used to refer to both the contact map pattern as well as the model of chromatin

folding into a loop [22, 30]. In this thesis, I will adopt the term loop for the model of chromatin folding and looping interaction for the pattern in Hi-C contact maps. These structures are focal interactions between two genomic loci - called loop anchors - and are frequently thought of as representations of loops in 3D-space. However, the techniques used to study them are population techniques, measuring average conformations of millions of cells. We therefore do not know whether single, stable chromatin loops are formed in all cells, or whether looping interactions represent the tendency of two loci to contact each other more frequently in a dynamically changing chromatin environment. However, several lines of evidence, such as single particle tracking of proteins that are known to associate with loop anchors and polymer modeling [31, 32], suggest that the latter interpretation might be true. Hi-C maps revealed that looping interactions frequently co-occur with another pattern, called TADs.

Topologically associating domains (TADs) are visible in contact maps as "triangles" with increased contact frequency, emanating from the main diagonal, often containing a looping interaction at their tip [22, 24]. The triangular shape of TADs suggests that loci inside a TAD interact more frequently with themselves than with loci outside the TAD and that the boundaries of TADs are depleted for contacts that cross them. This phenomenon has been named "insulation" and is frequently used in computational algorithms that determine the location of TADs [33–35]. Interestingly, the location of TADs is highly similar in the genome of different tissues and is even highly conserved between different species, suggesting that they have an important functional role (see 1.3.1 for more details) [24, 36]. On a mechanistic level, polymer modelling suggests that TADs in contact maps arise through the averaging of many small loops that occur within the TAD genomic region and thus TADs likely do not represent static "globules" of interaction, but dynamic chromatin looping [32]. Experimental support for such a dynamic organization comes from OligoPaint FISH experiments, where individual cells show a wide array of different looping interactions within TADs, but when averaged, the signal coalesces into the familiar TAD pattern [37].

The next level of chromatin organization has been named "compartmentalization". Compartmentalization manifests itself as blocks that tend to associate with other, similar blocks over vast genomic distances, forming a "checkerboard pattern" [21]. Initially, coarse-grained analysis suggested two types of compartments, which seemed to differ in gene expression state. One type - termed A-compartments - is associated with active histone marks (see 1.2.1) and gene rich, whereas the other type - termed B-compartments - is associated with gene repression [21]. The mechanism by which the different compartment types interact with each other is most likely through weak intermolecular forces between chromatin, possibly as a phase-separation phenomenon (see 1.2.1 for details) [38, 39]. It is noteworthy that the TAD-compartment relationship is not a strict hierarchy: There are "compartment domains" that are in the size-range of TADs and are primarily formed based on a compartment mechanism, based on their persistence in degradation conditions that remove TADs [40, 41]. However, in general, compartments do exhibit longer range interactions, whereas TADs are

predominantly short-range phenomena.

The largest scale of chromosome organization in interphase (see 1.4 for consideration in other cell cycle stages) are chromosome territories. This term refers to the finding that chromosomes tend to not intermix and occupy a specific subvolume of the nucleus [42]. This phenomenon can be seen in contact maps as large triangles surrounding entire chromosomes [21], but it has also been studied using microscopy-based techniques such as FISH (see 1.6.2) [43]. While the existence of chromosome territories has been confirmed in multiple cell types and organisms, their exact functional roles are still obscure. One potential function that has been proposed is that holding chromosomes separate in interphase will allow easier individualization in subsequent mitoses [42]. However, this does not explain why non-cycling cells maintain chromosome territories and thus much more work needs to be done to elucidate this principle of chromosome organization.

High resolution chromosome conformation capture approaches have not only revealed a wealth of different interaction patterns on various length scales, but they have also allowed to merge predictions from polymer modeling with experimental evidence. The problem of fitting large amounts of a (bio)-polymer into a small space has led to theoretical modeling efforts to explain how such an arrangement could work [44]. Based on the packing density of DNA within cells, the polymer is thought to assume a dense globule organization. However, most dense globules exhibit extensive knotting and with that a transition to a glass-like state. This would hamper biological function severely since many processes need access to different parts of the genome. Therefore, it was theorized that the genome would assume a fractal globule configuration that exhibits non-penetrating folding of globular units, which is self-similar on multiple scales. These structures are knot-free and would thus solve the problems mentioned before. High resolution chromosome conformation techniques were then able to confirm that indeed interphase chromatin assumes a fractal globule organization by measuring the decay of contact frequency with genomic distance, which was shown to follow mostly an exponential decay with an exponent characteristic for fractal globules [21, 45]. To sum up, the genome folds into a complex mixture of different structures on multiple length-scales, suggesting that a multitude of different control mechanisms have evolved to establish, maintain and change this organization.

1.2 Factors shaping chromatin structure

The biochemical state of a genomic region - defined as the set of associated protein factors and their post-translational modifications (PTMs) - can change its biological properties dramatically: It can permit entry of protein machineries that read genetic information, or package DNA tightly to avoid association of transcription factors. Moreover, this organization is subject to highly dynamic regulation to allow the cell to respond to environmental or developmental queues. Within this highly complex organization, the post-translational modification of histones is probably one of most important regulatory actors.

1.2.1 Histone PTMs

The flexible linkers of histones that emanate from each nucleosome (see 1.1.1 for details) are subject to a multitude of different post-translational modifications (PTMs) that can change the biochemical and biological nature of a given nucleosome substantially [7]. These modifications exert their effects either by changing the biochemical properties of nucleosomes directly - for example by altering its net charge - or by influencing the interaction with chromatin associated proteins that in turn can elicit an effect. It has been proposed in the past that these PTMs constitute a "histone-code", where PTMs of histones encode combinatorial information similar to the genetic code that is the basis of protein encoding [46, 47]. While the existence of such a "histone-code" is disputed, it is clear that different histone PTMs can exert distinct effects on chromatin structure.

The earliest histone PTM to be discovered is the acetylation of lysine residues on tails of histones [48]. Here, an acetyl group is attached on the ϵ -amino group of lysine by enzymes called histone acetyl transferases (HATs). These enzymes exist in an equilibrium with histone deacetylases that catalyze the removal of histone acetylation, often in the context of gene repression and chromatin compaction [49]. The immediate effect of this modification is to change the net charge of the modified lysine from positive to neutral, which is thought to weaken the electrostatic interaction between DNA and histones, resulting in a looser DNA association. Additionally, acetylation of histone H2A and H4 is thought to interrupt the inter-nucleosomal interaction of these proteins, resulting in disruption of nucleosome higher-order association [50]. In addition, histone acetylation serves as a binding platform for proteins that contain a so-called bromodomain [51]. This domain mediates binding to acetylated lysines and can be very specific to a given histone context, e.g. lysine 14 of histone H3 [52]. The recruited proteins often mediate amplification of histone acetylation in a feedback manner and are therefore involved in maintenance of chromatin domains that are rich in histone acetylation. Moreover, the recruited proteins can exert their effect through different means as for example many activating transcription factors contain binding domains for histone acetylation [51]. To sum up, histone acetylation is therefore thought to exert an activating effect on gene expression since it results in open chromatin, allowing transcription factors to access the underlying DNA, and mediates recruitment of protein factors that have a similar effect.

The methylation of lysine residues on histone tails is another very prevalent histone PTM. In this modification, a methyl group is attached to the ϵ -amino group of lysine by enzymes called histone methyl transferases (HMTs) [53]. As with histone acetyl transferases, these enzymes exist in a balance with histone demethylases that catalyze the removal of methyl groups [54]. In contrast to histone acetylation, methylation does not change the net positive charge of lysine and therefore has no effect on the association between DNA and nucleosomes. Rather, histone methylation exerts its effects through recruitment of protein factors that harbor a so-called chromodomain [55]. This protein domain medi-

ates binding of protein factors to methylated lysines in a specific histone context. Interestingly, the recruited protein factors can have either an activating or repressive effect on gene expression, depending on where exactly the modification residues: Trimethylation of lysine 9 of histone H3 (H3K9me3) is considered a purely repressive mark since it recruits heterchromatin protein 1 (HP1) that is capable of compacting chromatin (see below) and mediating recruitment of other repressive factors [56]. On the other hand, trimethylation of lysine 4 of histone H3 is considered an activating mark since it recruits activating transcription factors including the general transcription factors (GTF) required for transcription of most genes [57, 58]. Additionally, histone methylation has been shown to persist across cell divisions and it is therefore thought to be involved in mediating gene expression memory, the inheritance of gene expression state [59, 60]. In summary, histone methylation is a more complex mark than histone acetylation since it exerts its effect mostly indirectly and is capable of mediating both repression and activation depending on the precise location [61].

On a larger scale, different reports have indicated that histone PTMs can influence the coalescence of extended stretches of chromatin to compact, phase separated assemblies [62–65]. The earliest report of this behavior comes from electron microscopy studies of polycomb repressive complex 1 (PRC1; see 1.2.2 for details) [62]. This complex is capable of binding to trimethylation of lysine 27 of histone H3 (H3K27m3) via the chromodomain of its polycomb subunit and is thought to exert a repressive effect on gene expression [66]. Interestingly, this study showed that addition of PRC1 to artificial chromatin is sufficient for its compaction, a finding that was corroborated and extended by recent work that suggests that PRC1 forms phase-separated condensates *in vitro* and *in vivo* [63]. A further example of this principle is HP1 [56], which is thought to mediate gene repression upon binding to H3K9me3. Here, studies showed that a major mechanism for this repressive effect might be direct chromatin compaction since HP1 was shown to coalesce into compact droplets both *in vitro* and *in vivo* [64]. Interestingly, studies of unmodified chromatin arrays suggest that the propensity to form phase separated droplets might be the "ground-state" of chromatin since these assemblies readily partition into compact condensates [65]. These studies further suggest that histone acetylation disrupts this tendency and suggests that one of the mechanisms through which histone acetylation exerts its activating effects might be through disruption of phase-separated condensates. In summary, histone PTMs and the proteins they recruit have a dramatic effect on the long range interactions of chromatin via modulating phase separation propensity.

To sum up, the "chromatin state", which in this context is taken to be the collective histone PTMs and associated protein factors of a certain genomic location, has great influence on the overall structure of chromatin and can influence whether a given locus is permissive to transcriptional activity or not.

1.2.2 Polycomb group proteins

The polycomb system was originally discovered in *D. melanogaster* by experiments examining the maintenance of gene expression in developing embryos. The body of the developing fly is patterned very early into different segments by the combinatorial activation of key enhancers through the action of transcription factors [67, Chapter21]. However, the expression of these factors ceases after a short burst and yet the segments maintain their identity throughout the entire life of the fly. The factors that mediate this gene expression memory were elusive until polycomb group (PcG) gene mutants were isolated that caused the ectopic expression of HOX genes - the master regulators of segment identity - in a wrong body segment, after the initial patterning cues disappeared [68, 69]. In the past 50 years, orthologous proteins were found in a wide array of multicellular organisms, suggesting a highly conserved role for this system [70].

Polycomb group proteins exert their biological function through a multitude of biochemical mechanisms, most of which are connected to the post-translational modifications of histones (see 1.2.1) [71]. These modifications are catalyzed by two enzymes that perform their function in the context of large, multisubunit protein complexes: Polycomb repressive complex 1 and 2 (PRC1 and PRC2). In the case of PRC1, the main modification is ubiquitylation of lysine 119 of histone H2A (H2AK119ubi), catalyzed by the E3-ligase dRing (Ring1b in mammals) [72]. PRC2 - on the other hand - catalyzes the trimethylation of lysine 27 of histone H3 via its methyltransferase E(z) (EZH2 in mammals) [73]. It is thought that the establishment of epigenetic memory requires the concerted action of both complexes with the exact recruitment mechanism being the subject of active research [74, 75]. It is however clear that their interaction happens - at least in part - through the ability of subunits of one complex to bind the modifications installed by the other. A prime example here is the ability of *polycomb*, a subunit of PRC2, to bind H3K27me3, the modification installed by PRC1, via its chromodomain [76].

The main effect of polycomb action at a certain genomic location is the repression of gene expression. This effect is thought to be mediated both via the modifications directly and by recruiting accessory factors: It is known that H2AK119ubi is able to block elongation of RNA polymerase II [77], whereas the mechanism of gene repression by H3K27me3 is thought to be more indirect. Here it is believed that the main way H3K27me3 exerts gene repression is by recruiting PRC1 via interaction of *polycomb* (CBX proteins in humans) with this modification. This both leads to more H2AK119ubi at the recruiting site and increased PRC1 recruitment, which is thought to lead to further repression since PRC1 is able to compact chromatin *in vitro* and *in vivo* [62, 78], thus blocking access to DNA. In addition, a variant of PRC2 containing EZH1 was recently shown to also be able to compact chromatin, suggesting that both polycomb group complexes are able to perform this function [79]. Interestingly, recent work has also shown that PRC1 is able to form phase separated droplets that might contribute to the establishment of a compact, repressive environment [63]. These findings suggest that polycomb group proteins might not only have

an influence on the state of gene expression, but also on the structure of the genome.

Indeed, polycomb repressive elements (PREs) - the binding sites of polycomb group proteins - associate in 3D in the nucleus of *D. melanogaster* and form repressive hubs, termed "polycomb bodies" both in flies and in mammals [80, 81]. Interestingly, these interactions are able to cross chromosomes and exhibit synergistic silencing, which might explain the involvement of polycomb group proteins in pairing sensitive silencing, where PREs on homologous chromosomes exhibit synergistic repression [82–84]. While the mentioned studies demonstrate a strong correlation between polycomb group protein action and 3D-chromatin architecture, they do not establish causality. In a seminal recent work, Rhodes et al. show that polycomb group proteins are able to mediate 3D-interactions in the mammalian genome independently of cohesin - the other major driver of interphase 3D-interactions (see 1.2.3) - and sensitive to PRC1 inhibition [85]. Taken together, this suggests that polycomb group proteins have a tremendous influence both on gene expression as well as on the 3D-architecture of the genome.

1.2.3 SMC complexes and loop extrusion

SMC complexes are one of the most important classes of chromatin associated architectural proteins. They have important functions in gene expression control, DNA repair and are thought to be largely responsible for most of the higher order 3D-architectural changes that happen both in interphase and mitosis (see 1.1.2 and 1.4).

Protein complexes that include structural maintenance of chromosome (SMC) subunits, have traditionally been referred to as SMC complexes [87]. To date, we know three different classes of SMC proteins in eukaryotes: Condensin, cohesin and SMC5/6, of which I will discuss the first two here (Fig. 1.3). Condensin was originally discovered as a factor in *X. laevis* egg extract that was necessary and sufficient to induce chromosome compaction in the context of mitosis [88] (see 1.4.1 and 1.4 for more details). Cohesin, on the other hand, was discovered in a *S. cerevisiae* mutagenesis screen for factors involved in sister chromatid cohesion [89]. Once both proteins had been isolated and imaged using rudimentary electron microscopy, it came as a surprise that proteins that performed such seemingly different functions have very similar composition and overall structure [90, 91].

Both complexes harbor a heterodimer of SMC proteins that form a V-shaped coiled-coil structure [90, 92, 93]. Cohesin contains both SMC1 and SMC3 whereas condensin harbors SMC2 and SMC4. These SMC proteins fold in on themselves with each SMC subunit starting with its N-terminus at the so-called head-domain, making its way to the hing-domain and reversing back to the head-domain, forming a coiled-coil structure with itself on the way. Both cohesin and condensin have been thought of as a ring-structure in part due to imaging results and since they associate with a third type of subunit - called kleisins - that close the tripartite ring [87]. Cohesin's kleisin subunit is called

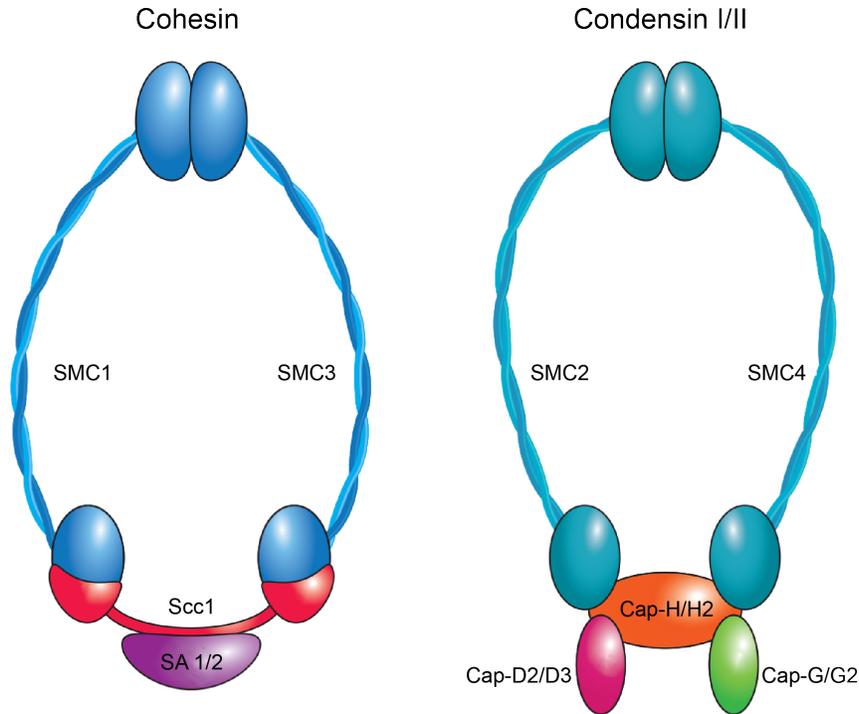


Figure 1.3: **SMC complexes.** Subunit organization of the SMC complexes cohesin as well as condensin I and II is depicted. Figure adapted from Fig. 1 of [86]

Scc1 and is thought to bridge the head domains of SMC1 and SMC3. The situation is more complex for condensin, where vertebrates contain two types of condensin complexes - condensin I and II - that share SMC2 and SMC4, but differ in accessory subunits [94]. In the case of condensin I the kleisin subunit is called Cap-H and the kleisin subunit of condensin II is called Cap-H2.

In addition to the core subunits that constitute the tripartite ring of cohesin and condensin, both proteins contain additional subunits called HAWK (**H**EAT repeats **a**ssociated with **k**leisins) proteins [95]. In the case of cohesin, HAWK proteins - with the exception of SA 1/2 - are not thought to be part of the core complex, but associate transiently to preform various regulatory functions (see 1.4 for more details on some of these subunits) [95]. Perhaps the best known factors here are NIPBL (Scc2) and Mau2 (Scc4) that are thought to be necessary to load cohesin onto chromatin and were recently shown to be essential co-factors for cohesin's loop extrusion activity (see below) [96, 97]. Wapl/Pds5 are the other important HAWK proteins associated with cohesin: It is thought that Wapl/Pds5 mediate unloading of cohesin from chromatin to limit its residency time in interphase and purge cohesin from chromosomal arms

in mitosis (see 1.4 for more details) [98, 99]. Condensin - on the other hand - is thought to permanently associate with its HAWK subunits, which constitute parts of the core complex: Cap-D2 and CAP-G for condensin I and CAP-D3 and CAP-G2 for condensin II [94]. While the biochemical composition of cohesin and condensin has been clear for many years, the exact mechanism of how they perform their function has been elusive until very recently.

The ring shape of cohesin offered an initial, naive explanation of how it might mediate sister chromatid cohesion, namely by holding the sister chromatids together either in its lumen or as a concatamer of rings (see 1.4.2 for more details). However, it also seemed to be involved in chromosome condensation - at least in some organisms - as disruption of the cohesin kleisin subunit in yeast not only showed defects in sister chromatid cohesion, but also severely affected chromosome compaction in mitosis [100]. Additionally, ChIP-seq studies suggested that cohesin associates with the boundaries of topologically associating domains and was necessary for the insulating function of CTCF [101] (see 1.2.4). Indeed, later perturbation studies suggested that degradation of Scc1 caused a complete loss of TADs and looping interaction in multiple organisms [23, 41, 102]. These studies are complemented with experiments disrupting the cohesin loading factor NIPBL [25], together suggesting that cohesin is responsible for organising the interphase genome into the familiar structures seen in Hi-C experiments. However, the mechanism of how ring shaped SMC protein complexes might mediate chromosome compaction and interphase chromatin structuring was completely unclear.

Crystallographic and enzymatic studies suggested that the head-domain of both complexes is an ABC-type ATPase and thus might provide the energy input needed for the compaction process, but how this energy is used was not known [91, 103]. First hypotheses about a potential mechanism came from electron microscopy studies of mitotic chromosomes (see 1.4.1) [104]. These studies suggested that - under certain denaturing conditions - large chromatin loops emanated from a central proteinaceous scaffold. Although the precise nature and function of this scaffold has been challenged in recent years [105], the notion that condensin might be responsible for creating these loops arose. Similarly, in lampbrush meiosis, loops emanating from a central scaffold were observed in electron microscopy studies [106]. This, and the fact that *S. cerevisiae* mutants lacking Rec8 - the meiosis specific kleisin subunit of cohesin - do not assemble proper meiotic chromosomes, led to the hypothesis that cohesin might organize meiotic chromosomes via formation of loops [107, 108]. The observation of such loops in processes associated both with condensin and cohesin laid the foundation for explaining the mechanism of SMC complex action on chromosomes using a unified theory.

Specifically, the theory emerged that SMC-complexes might perform their function through binding and possible expanding chromatin loops (a process termed "loop extrusion") [30, 108, 109]. Theoretical modelling suggested that the existence of a loop extrusion factor (LEF) - a protein that is capable of binding chromatin and extruding loops with a certain binding affinity and processivity - is sufficient to explain both individualization of sister chromatids

in mitosis as well as chromosome compaction to cylindrical bodies (see 1.4 for more details) [30]. Later modelling studies then also showed that a potential loop extrusion function of cohesin can explain how it shapes TADs and loops in interphase chromosomes and how its perturbation might explain the complete lack of these signatures [32, 39]. Therefore, given its elegance and explanatory power, it is not surprising that the loop extrusion hypothesis was quickly accepted in the community.

However, the loop extrusion hypothesis remained theoretical for a considerable amount of time since no loop extrusion activity had been shown for neither condensin nor cohesin. Several studies had shown that cohesin can translocate on DNA *in vitro* and that condensin is able to compact and exert force on DNA held with optical tweezers [110–112]. These studies were in support of cohesin and condensin being molecular motors, but did not suffice to accept the loop extrusion hypothesis. Finally, in 2018, loop extrusion was shown *in vitro* for purified condensin on DNA curtains [113] and in 2020 using *X. laevis* extract [114]. Cohesin followed shortly after with two independent laboratories showing its ability to extrude DNA loops *in vitro* in 2019 [97, 115]. Interestingly, while condensin extrudes asymmetric chromatin loops - with one side fixed and the other "reeled in" - cohesin extrudes symmetric loops [114]. Modelling studies suggest that asymmetric loop extrusion would not provide the degree of compaction that we observe in mitotic chromosomes, hinting at still unresolved details of the process [116]. To sum up, the loop extrusion hypothesis elegantly explains the principle of how both cohesin and condensin shape 3D-chromatin architecture of interphase and mitotic chromatin. Although some of the details are still missing, it is very likely that it will be able to explain a large amount of the characteristic functions associated with SMC biology.

1.2.4 CTCF

CTCF is - next to SMC complexes (see 1.2.3) - probably the most important factor that shapes interphase genome architecture. It has been associated both with specific interactions between promoters and enhancers that influence gene expression as well as with a general role of providing genome-wide road-blocks for the architectural functions of SMC-complexes.

CTCF was originally discovered as a suppressor of the oncogene *c-myc* and is a Zinc-finger DNA-binding protein. CTCF has an unstructured N- and C-terminus and contains 11 Zinc-fingers that specify its cognate DNA motif consisting of three repeats of the core-sequence "CCCTC" with different spacers [117]. CTCF is a highly conserved protein that is found in most animals with the notable exceptions of *C. elegans* and *S. cerevisiae* [117], suggesting an essential function.

The biological role of CTCF that is perhaps best known is as an insulator protein [118]. Insulators are DNA-binding proteins that are thought to limit the interaction of promoters and enhancers across their binding sites and are thus thought to be capable of preventing aberrant regulatory interactions. Here, the role of CTCF has been extensively studied at the β -globin locus, where deletion

of a CTCF binding sites can lead to ectopic expression of multiple genes [118]. Another example of CTCF's role as an insulator protein is in the context of imprinting - the homolog specific expression of certain genes. Here, CTCF only binds the unmethylated, maternal allele of the *H19* imprinting control region and thus blocks the association of a downstream enhancer with the *Igf2* gene. The paternal allele is not methylated and thus permits interaction of the enhancer with *Igf2*, leading to its expression [119].

Different mechanisms have been proposed for how CTCF might perform this insulation function. One theory was that CTCF might limit the spread of certain chromatin marks, thus "insulating" the gene from an activating or repressive environment spreading from regulatory regions. This notion is corroborated by the fact that CTCF is capable of limiting the spread of H3K27me3 in embryonic stem cells [120]. A different theory was that CTCF might block the physical association of promoters with regulatory regions in 3D-space. Support for this notion comes from studies that examine the interaction between CTCF and cohesin, a protein that has a tremendous impact on shaping the 3D-genome (see 1.2.3 for more details). Here, researchers showed that CTCF binding sites in the genome overlap with the binding sites for cohesin and that CTCF is required for positioning cohesin on DNA [101]. Importantly, this study also showed that cohesin in turn is required for the insulator function of CTCF. Later studies then revealed that CTCF marks chromatin looping interactions genome-wide and that CTCF perturbations leads to a marked loss of looping interactions [22, 121]. This suggests that CTCF performs its insulating activity mainly through shaping the 3D-genome via its interaction with cohesin.

Interestingly, the interaction between these two proteins seems to be more complex than CTCF posing a simple boundary for cohesin: The binding sites of CTCF that mark the loop anchors of a looping interaction are oriented in a "convergent" orientation more often than expected by chance [22]. Indeed, inverting a CTCF motif was sufficient to completely abolish looping interactions at some genomic locations [122]. This phenomenon has puzzled researchers for some time and although some progress has been made recently [123, 124], the mechanism of how this orientation preference is achieved is still largely unknown. To conclude, CTCF has a tremendous impact on the 3D-structure of animal genomes, mainly through its interaction with the cohesin complex.

1.3 Functional importance of DNA structure

In general, all biochemical processes that act on DNA inside living cells do not act on bare DNA, but on chromatin, which is an assembly of protein factors and DNA. These processes therefore highly rely on the specific accessibility of the DNA substrate and it is thus not surprising that organisms have evolved the ability to exploit this organization for a multitude of different processes such as gene regulation, DNA repair and chromosome condensation in the context of mitosis.

1.3.1 Gene expression

The most intuitive and historically best researched cellular role of 3D chromatin changes is in the context of regulating gene expression. The principle behind this type of regulation is simple: Parts of the genome that should be read in a given cell type are accessible whereas parts that should not be read are inaccessible. The first hint that the genome is compacted differentially in distinct cell types came from early electron microscopy (EM) studies that showed that a large portion of the genetic material of each cell appears to be packaged tightly in regions that are known as heterochromatin (due to their dense staining in EM images), whereas other regions - known as euchromatin - appeared to be packed much less tightly [125]. Interestingly, these regions are dependent on the specific cell types and environmental conditions, suggesting that differential compaction is associated with gene expression changes. Notably, it was often the case that regions of heterochromatin occupied areas close to the nuclear periphery, a milieu that is now known to be associated with repression of transcriptional activity [125]. Since then, multiple lines of study have established a strong correlative link between compaction of chromatin areas and transcriptional activity.

The first phenomenon that connected heterochromatin with gene repression was the inactivation of the second X-chromosome of female mammals. Here, it was found that the inactive X-chromosome is much more compact and densely staining than the actively transcribed copy [126, pp. 252–278]. More recently, OligoFISH experiments examining active and inactive chromatin loci of *D. melanogaster* showed that active loci occupied a much larger volume than inactive regions [127], suggesting that gene activity associates with chromatin decondensation. Further evidence for a correlation between transcriptional activity and chromatin structure comes from Micro-C, a Hi-C variant (see 1.6.3), where it was found that *S. cerevisiae* genes show a pronounced anticorrelation between transcriptional activity and compaction state [26]. In addition to correlative hints that chromatin compaction might influence transcriptional output, a number of studies suggest a causal relationship. Gcn5, a major transcriptional activator in *S. cerevisiae*, was found to be a histone acetyltransferase thought to promote opening up of chromatin [128] (see 1.2 for more detail). Moreover, PRC1, a transcriptional repressor, was found to be able to compact chromatin [62] (see 1.2 for more detail). Furthermore, in the above mentioned Micro-C study [26], treatment of cells with diamide - a stressor that leads to a change in expression of 20 % of genes - led the decompaction of the newly activated genes. One can thus conclude that a wealth of evidence suggest a causal link between chromatin structure and transcriptional activity and that modulating chromatin structure is a prime means to alter gene expression state.

Compaction as a direct tool to modulate gene expression is not the only way chromatin structure exerts its function. An emerging principle in the last years has been that chromatin structure may have a profound influence on the control of gene expression. The first idea in this regard where promoter-enhancer loops to explain the communication between genes and their regulatory elements. Specifically, the theory was that enhancers - which are often located

many hundreds of kb away from their controlled promoters - would exert their action by coming into physical proximity to the promoters via a chromatin loop. Current advances in methodology have largely confirmed this notion on a structural level with Hi-C and, more recently, Micro-C revealing several hundred focal interactions between promoter and enhancer loci in the human genome [26, 129, 130]. On a functional level, the validation for this model mostly comes from perturbations of factors involved in loop formation such as cohesin (see 1.2.3 for more detail) and CTCF (see 1.2.4 for more details). Here, a screen in *D. melanogaster* to determine factors important for long range regulatory interactions yielded the cohesin loading factor Nipped-B, suggesting loop formation to be a prerequisite for long-range regulatory interactions [131]. Perhaps the strongest evidence for this model, however, comes from recent single molecule imaging studies, where promoter-enhancer looping interactions were directly imaged and correlated with transcriptional output [132]. Thus, it is no surprise that the looping model for promoter-enhancer interactions is widely accepted in the community [133].

On a larger scale, TADs (see 1.1.2) have also been suggested to have an influence on gene expression output. Here, a number of studies reported that TADs constitute gene regulatory domains, with genes residing in the same TAD to often be co-regulated [134]. Indeed, this is reflected by chromatin state as TADs often delineate regions of active or inactive histone modifications [22, 134], with their boundaries often being marked by active genes and activating histone modifications. Functional support for this model comes from studies examining the *shh* locus and its regulation in mouse development. Here, genetic perturbations that move the enhancer activating *shh* into a different TAD lead to loss of activation of *shh* and developmental defects [135]. Mechanistically, this behavior can be explained by considering TADs as contact domains that insulate a chromatin environment. In this model, interactions between regulatory elements within a TAD are more likely than interactions of regulatory elements across TAD boundaries. This would explain how genes are often co-regulated within a TAD, carry similar histone modifications and how perturbations of TAD-boundaries would lead to aberrant gene activation.

However, although this model is very attractive, a number of studies suggest that the real situation is more complex. First, a study performed on *D. melanogaster* mutants with genome-wide 3D-structure changes only found minor effects on gene expression, although many TADs were altered [23, 25]. In addition, more recent and thorough studies on the *shh* locus, suggests that *shh* expression is remarkably resilient to alterations of TAD structure [136]. Taken together, while 3D-genome conformation is undoubtedly important for gene expression, the precise relationship and mechanisms remain to be elucidated.

1.3.2 DNA repair

Within cells, DNA is subject to a multitude of different lesions that happen with surprising frequency: The most common lesion - single strand DNA breaks - happens 50,000 times per cell per day in an average human [137], whereas the

rarest lesion - double strand DNA breaks - still occurs 10 times per cell per day [138].

Given this high frequency of DNA lesions, it is not surprising that cells have evolved many different repair pathways that are custom tailored to repair specific DNA damage types. DNA damage can be broadly classified into single strand damage (SSD) and double strand breaks (DSB). Cells use different pathways to repair each of these lesions. If a given instance of a SSD falls into a narrow class of certain lesions, cells can employ direct damage reversal, where a specific enzyme catalyzes the removal of the lesion. If this is not possible, the damaged piece of DNA is either removed via base excision repair or nucleotide excision repair and resynthesized via the second strand as template. While SSD is relatively common and easy to repair, DSB pose a more serious issue to cells. There are two pathways that can be employed to repair these lesions. Non-homolog end joining (NHEJ) is the preferred pathway in human cells if there is no second sister chromatid. Here, the broken DNA ends are repaired by random nucleotide synthesis and then ligated back together. This process is inherently error prone and thus - if the second sister chromatid is available - human cells choose to repair DSB via homologous recombination (HR). This process uses the DNA information on the second sister chromatid to repair the lesion exactly and requires that sister chromatids to come into close proximity [67].

In cells, DNA damage repair does not happen on bare DNA, but on nucleosome containing chromatin that exhibits complex 3D-folding. Many DNA repair pathways are thus dependent on the specific structure of chromatin and often use enzymes and structural proteins that can manipulate chromatin folding to perform DNA repair.

A prime example for this principle is the extensive use of histone modifications that can modulate accessibility to the chromatin template (see 1.2) by a variety of DNA repair pathways: Histony acetylation accompanies some forms of DNA damage and Gcn5 (a histone acetyltransferase) is frequently present at sites of DNA damage [139, 140]. This modification causes chromatin to open up and thus grants repair factors access to the DNA substrate. Along these lines, nucleotide excision repair (NER) is stimulated by the nucleosome remodelers ACF and SWI/SNF that are associated with displacing nucleosomes to make the DNA template accessibly [141, 142]. Thus, compact chromatin poses an obstacle to DNA damage repair and different pathways allow conversion to open chromatin in order to allow access for DNA repair factors.

On a larger scale, the organization of the genome into topologically associating domains (TADs; see 1.1.2) provides a natural partitioning of the genome that is used by some repair pathways to limit the extent of repair factor binding. Specifically, 53BP1 and RIF1 - two factors involved in the repair of double strand breaks (DSB) - accumulate in TAD-sized domains around DNA damage sites. Perturbation of TAD structure by depleting cohesin (see 1.2 for more details) caused loss of this organization and spreading of DSB associated factors to larger regions, resulting in mutagenic hyper-resection of DNA ends [143]. This suggests that TAD-level organization by the cohesin complex is required to confine DNA repair associated events to a limited region around DSBs.

Cohesin has been implicated in DNA repair not only through its ability to shape TADs, but also via its involvement in sister chromatid cohesion. The fact that cohesin is involved in DNA repair was initially shown by its requirement for survival of *S. cerevisiae* in the presence of several genotoxic agents [144, 145]. It was then found that cohesin is needed to rescue stalled replication forks through template switching, where the newly synthesized sister chromatid is used as a repair template. Interestingly, this effect is likely due to cohesin's role in bringing sister chromatids closer together since the rescue defect caused by cohesin depletion could be rescued by artificially tethering sister chromatids together [146, 147]. Moreover, cohesin likely affects the choice of DNA repair pathway upon DSB, with depletions leading to increased usage of the error prone NHEJ pathway [148, 149]. Other studies also suggest that cohesin promotes the use of the sister chromatid as a donor for recombination upon DSB, with perturbation favoring the use of the homologous chromosome [150]. In conclusion, cohesin and its effect on DNA structure have a large effect on DNA repair, especially in the context of double strand breaks.

To conclude, the 3D-structure of the genome has a multitude of functional roles in interphase cells, ranging from controlling and regulating gene expression to affecting the way cells repair DNA lesions. However, perhaps the most important cellular role of 3D-structural changes is in the context of permitting ordered segregation of the genome in the context of the cell cycle.

1.4 DNA structure throughout the cell cycle

It is a central principle that new cells can only arise from old ones and thus cell division is a very basal requirement for the propagation of life. Unicellular organisms employ simple cell divisions to multiply, whereas multicellular organisms rely on cell division for an array of additional tasks. The most essential task that depends on cell division is probably in the context of development. Here, highly coordinated cell divisions that are often asymmetric give rise to an organism with diverse tissues and cell types. However, also in fully developed multicellular organisms, cell division is vital. When animals suffer injury, cells need to divide in the context of wound healing and when they are attacked by a pathogen, the immune cells need to multiply via cell division to provide adequate defense.

At each cell division, two basic tasks need to be performed: Duplication of the genetic information and segregation of the two genome copies to the daughter cells. Eukaryotic cells go through different discrete steps to achieve this goal, which are collectively called the cell cycle (Fig. 1.4). The cell cycle is traditionally divided into four phases, which are characterized by different events. The **G1-phase** - the first "gap" phase - is the time where cells grow and replenish their intracellular molecules after division. The name "gap" phase comes from the fact that cells do not perform any of the core functionalities (replication and segregation) during this step. Many cells that do not undergo cell division exit the cell cycle at this stage, which is then called **G0-phase**. When cells then enter the cell cycle, they first duplicate their DNA in the so-

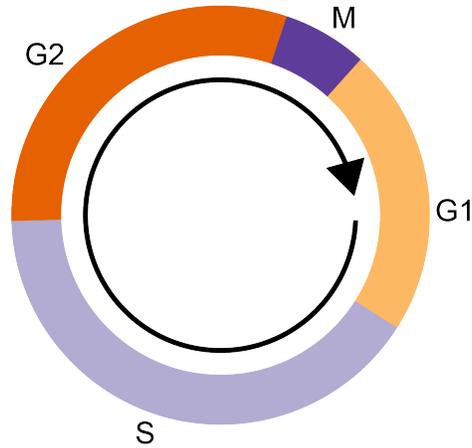


Figure 1.4: **Stages of the cell cycle.** The stages of the cell cycle are depicted with the ring segments denoting their approximate relative length. G1, Gap-phase 1; G2, Gap-phase 2; M, Mitosis; S, S-Phase

called **S-phase**, before reaching the second "gap" phase, **G2-phase**. Then, cells enter mitosis, or **M-phase**, where chromosomes are segregated to the developing daughter cells. The cell cycle is then completed in a process called **cytokinesis**, where the membrane connection between the two daughter cells is capped [67, Chapter 17].

Since cell division is so immensely important for life, it is not surprising that cells have evolved control mechanisms that govern its correct function. Most cell divisions - with the notable exceptions of cleavage divisions during embryonic development - are thus governed by three cell cycle checkpoints that constitute this control mechanism. These checkpoints are independent of the machinery performing the structural and biochemical changes that happen in each cell cycle stage and can be thought of as the cell cycle's "regulatory layer". These checkpoints check whether certain events that were supposed to happen at a given cell cycle stage did in fact happen. The nature of these checkpoints is "switch-like", meaning that once a requirement has been fulfilled and a checkpoint was passed, there is no turning back and the resulting events will ensue. The three cell cycle checkpoints thus constitute "switches" that are only flicked if certain events have happened [67, Chapter 17].

The first checkpoint is called "Start" and controls the transition between G1 and S-phase. Here, the cell checks whether it should enter the cell cycle, which can be governed by different considerations in different organisms. Simple unicellular organisms check whether they have grown enough since the last cell cycle and whether there are enough nutrients in the environment. Cells in

multicellular organisms - on the other hand - mostly sense growth-factor concentration, since the decision to divide is often not determined by individual cells in an organismic context. The next checkpoint that has to be passed is the DNA damage checkpoint. The goal here is to assess whether DNA has been replicated correctly and to determine if the cell is ready for entering mitosis (see 1.4.1). This checkpoint is active during S-Phase and G2 and needs to be satisfied before cells enter mitosis. The next and last checkpoint is the so-called Spindle assembly checkpoint (SAC) that determines whether the sister chromatids have been attached correctly to the mitotic spindle. If this checkpoint is passed, cells segregate their sister chromatids, undergo cytokinesis and thus complete cell division (reviewed in [151]).

On a biochemical level, regulation of cell cycle progression is primarily done via post-translational modifications and adjustment of protein amount. In fact, *X. laevis* egg extract is capable of making several cell-cycle transitions without gene expression, suggesting that - on a very basal level - post-translational events are sufficient for cell cycle progression [152]. Here, one of the most important post-translational modifications is phosphorylation. Phosphorylation is involved in all cell-cycle checkpoints and is responsible for activating numerous effector proteins that perform the structural changes needed. The main kinases that control cell-cycle dependent phosphorylation are called **Cyclin dependent kinases (Cdks)**. These proteins are inactive by themselves and require co-factors - called cyclins - for their activity. But cyclins not only activate Cdks, they also define their substrate specificity and thus the downstream effects of Cdk-phosphorylation. To ensure differential regulation of the different cell-cycle transitions, cells have evolved several cyclins that are expressed in a highly cell-cycle dependent fashion. In fact, *S. cerevisiae* only contains one Cdk - *Cdc28* - that is expressed throughout the cell cycle and all cell-cycle specific phosphorylation events are controlled by the expression of the associated cyclins. In mammals, this system is more complicated with several Cdks being expressed at different cell-cycle stages, but the underlying principle remains: Cdks together with tightly controlled cyclin expression direct cell-cycle stage specific phosphorylation (reviewed in [153]).

The second main biochemical driver of cell-cycle progression is protein degradation (reviewed in [154]). Here, two multisubunit protein complexes direct the ubiquitylation of key proteins at specific time points to ensure orderly cell-cycle progression. The first of these is the anaphase promoting complex, also called cyclosome (APC/C). This E3-ligase is activated at the metaphase-to-anaphase transition if the SAC has been satisfied and directs the ubiquitylation of securin and mitotic cyclins to ensure chromosome segregation and mitotic exit (see 1.4.2 for more details). The second main E3-ligase that controls cell-cycle progression is called SCF. SCF ubiquitylates a multitude of different proteins mainly during S-phase and G2 to ensure orderly DNA replication and mitotic entry. However, while post-translational modifications are required for cell-cycle progression, in higher organisms this toolbox is complemented using gene expression regulation on multiple levels to ensure a tightly regulated and robust cell cycle.

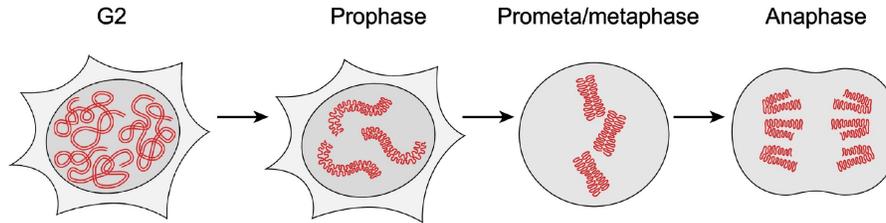


Figure 1.5: **DNA structural changes during Mitosis.** As cells enter mitosis, chromosomes start to condense in prophase, before congressing at the metaphase plate in metphase to finally segregate to the developing daughter cells in anaphase. Adapted from Fig. 1a of [155].

Given that the two main events that need to happen during every cell cycle - DNA replication and segregation of the genome - are tightly linked to DNA, it is not surprising that DNA undergoes major structural changes during the cell cycle, with the most drastic ones associates with mitosis.

1.4.1 Mitosis

Mitosis is one of the central events in the cell division cycle. It starts when DNA is fully replicated and culminates with the segregation of the two genomes to the daughter cells followed by abscission during cytokinesis [67, Chapter 17]. One of the most striking features of this process is the transformation of the extended genome to the highly compact, cylindrical mitotic chromosomes, a process that can be readily observed using even rudimentary light microscopes, having lead to a large body of research accumulating over the past 100 years. Traditionally, mitosis has been divided into several stages, characterized mostly by the visual changes that happen in the cell (Fig. 1.5).

In **prophase**, cells partially detach from the solid support they are growing on and round up to the characteristic spherical shape of mitotic cells. Additionally, chromosomes start to condense and individualize, being now visible as separated rods in light microscopy. Then, in **prometaphase** the nuclear envelope disassembles in a process called nuclear envelope breakdown (NEB) and chromosomes start to congress their chromosomes, before they are aligned at the equator of the cell in **metaphase**. The chromosomes then remain at the so-called "metaphase-plate" until the spindle assembly checkpoint (SAC) is satisfied and each replicated chromosome is attached to spindle fibers emanating from opposing spindle poles. This is called "bipolar" attachment, and when this is achieved, cells enter **anaphase**. Here, each replicated copy of each chromosome is moved to opposing cell poles by the mitotic spindle, before the two genomes reach their destination in **telophase** and the abscission of the cell

membrane during cytokinesis marks the completion of cell division (reviewed in [156]).

With the structural changes of DNA being so readily visible, the question of their purpose naturally arises. The most intuitive explanation comes from considering the constraint that cells need to move the replicated copies of the genome "from within". This means that the mitotic spindle needs to separate the replicated copies by pulling them to one side of the cell in a space that is bounded from above by the cell's radius. Since a typical cell measures $100\ \mu\text{m}$ in diameter and even the smallest human chromosome is over one centimeter long, it follows that DNA needs to be highly compacted for the separation of the two genomes to be feasible. Furthermore, chromosome condensation also leads to the individualization of chromosomes, separating from each other macroscopically. This is vital since entangled chromosomes would be very difficult to segregate and could lead to segregation defects. But chromosome compaction does not only change the volume of chromosomes and individualizes them, but it dramatically increases their rigidity. This further enables orderly segregation since the spindle pulls on mitotic chromosomes at a single attachment point, the kinetochore. If chromosomes lacked rigidity, pulling wouldn't cause a net movement of chromosomes, but mainly deformation. Thus, chromosome compaction solves several issues connected to the segregation of the replicated genome (reviewed in [155]). However, another issue arises during DNA replication.

Specifically, DNA-replication induces knots and intertwines between sister chromatids, resulting in highly entangled, replicated chromosomes. Therefore, to move each replicated chromatid to different cell poles during anaphase, sister chromatids need to be disentangled. This process is called sister chromatid resolution and is generally believed to happen at the same time as chromosome condensation during prophase and prometaphase (reviewed in [157]). A central prerequisite for sister chromatid resolution is the ability to perform strand passages, since chromosomes are much too long to disentangle them by threading. This process involves first cutting dsDNA, then moving another different piece of dsDNA through this cut-site and then finally ligating the cut again. This critical process is catalyzed by topoisomerase II and plays a central role in sister chromatid resolution [158]. However, to disentangle sister chromatids, an activity that catalyzes strand passages alone is not sufficient since this process also needs to be directional: Strand passages need to lead to disentangling, not entangling. Here, modeling studies suggest that loop extrusion (see 1.2.3) - a key activity attributed to condensin - is sufficient to impart directionality to topoisomerase II action, leading to sister chromatid resolution [30]. Experimental support for this notion comes from studies done in *D. melanogaster* embryos, suggesting that condensins provide the directionality for topoisomerase II action also *in vivo* [159]. Thus, condensin is the main driver of sister chromatid resolution. However, sister chromatid resolution is not the only chromatin structure change that is dependent on condensin. Indeed, condensin is thought to be responsible for the chromosome compaction and individualization that is so characteristic of mitosis.

Condensins are believed to be activated at the onset of mitosis by phos-

phorylation through CDK1-cyclinB, the main initiator of mitotic events [160]. They are thought to structure chromosomes through building huge DNA loops via a mechanism called loop extrusion (see 1.2.3). This is achieved by the concerted action of condensin I and condensin II, although they contribute to this process in different ways. Condensin II is present during early mitosis and initiates the bundling of mitotic chromosomes into large loops of 30-40 kb. Then, upon nuclear envelope breakdown, condensin I gains access to mitotic chromosomes and builds smaller loops within the large loops made by condensin II, which results in a nested loop configuration with outer loops growing to a size of 400 kb in prometaphase, harboring inner loops of roughly 80 kb [28]. Interestingly, experiments that perturbed condensin I and II suggest that chromatin compacts upon entry into mitosis even in the absence of both condensin complexes. This compaction leads to reduced chromatin volume, but not chromosome individualization and is hypothesized to be mediated by biochemical characteristics of chromatin itself [28]. Thus, the action of condensin I and II as well as chromatin compaction are responsible for mediating the necessary changes to chromatin structure during mitosis.

While resolving sister chromatids into individual cylindrical bodies is a prerequisite for bipolar attachment of chromosomes to the mitotic spindle, premature separation of the replicated chromatids would lead to massive chromosome missegregation. The cell avoids this problem by tightly controlling the initiation of chromosome segregation through proteinaceous bridges between sister chromatids. These bridges are destroyed only after bipolar attachment of all chromosomes to the mitotic spindle, when the spindle assembly checkpoint is satisfied. This destruction then initiates the transition from metaphase to anaphase and the segregation of chromosomes to the opposing spindle poles.

The protein that constitutes this linkage is the SMC protein cohesin (see 1.2.3) that not only controls major aspects of interphase 3D-genome structure, but is also responsible for sister chromatid cohesion.

1.4.2 S-Phase and sister chromatid cohesion

The central event of the cell cycle - next to chromosome segregation - is the duplication of the genome. While this duplication is necessary for cell division, it exacerbates a basic problem that cells face: Already in G1, human cells need to fit almost 2 m of DNA into roughly 100 μm of space, but after replication this amount doubles, complicating this daunting task even further. But the cell is not only tasked with packaging more DNA in a tight space, it also needs to ensure that the replicated sister chromatids stay in close proximity throughout S-Phase and early mitosis to allow equal segregation of the genome to the developing daughter cells. Interestingly, the cohesin complex is involved in tackling both these problems.

The cohesin complex is widely accepted to mediate compaction of the interphase chromosomes by structuring the genome into loops and topologically associating domains (see 1.2.3 for more details). However, cohesin was originally discovered in a screen for factors that are necessary for sister chromatid

cohesion [89] and was later found to be absolutely essential for this process [161], suggesting that cohesin performs at least two, seemingly unrelated tasks connected to shaping the 3D-genome. This dichotomy of cohesin is further reflected by the discovery that DNA replication leads to the emergence of two pools of cohesin with vastly different chromatin residence times: One pool with a residence time of roughly 10 min and another with a residence time of multiple hours [162]. These two pools have been named "dynamic" cohesin and "stable" cohesin respectively and are thought to reflect the separation of function between structuring the genome and mediating sister chromatid cohesion. On a biochemical level, a large body of research has elucidated the biochemical details that lead to this drastically different behavior.

In the G1 cell cycle stage, dynamic cohesin is thought to undergo loading to and unloading from chromatin through what is called the "cohesin cycle". Here, the cohesin-loader NIPBL(SCC2)/Mau2(SCC4) is thought to promote loading of cohesin onto chromatin as shown both *in vivo* for human cells and yeast as well as *in vitro* through biochemical assays [96, 163]. This loading was initially thought to involve topological engagement between cohesin and DNA, where the cohesin ring is opened and "clamped" around DNA, but is now rather believed to be pseudotopological meaning that a loop of DNA is threaded through the central cavity of cohesin (reviewed in [164]). Interestingly, NIPBL(SCC2)/Mau2(SCC4) has been recently shown to be an essential cofactor for cohesin's loop extrusion activity [97], suggesting that the processes of loop extrusion and cohesin loading might be mechanistically connected, potentially by extruding loops through cohesin's central cavity. The "cohesin cycle" is then completed by the unloading activity of Wapl/Pds5, which is thought to promote the removal of cohesin from chromatin [165]. The short residence time of dynamic cohesin is thus brought about by the kinetics of cohesin loading and unloading.

However, after DNA replication, a subset of cohesin changes its properties and is now stably bound to chromatin for multiple hours [162]. This change is thought to establish sister chromatid cohesion, with "stable" cohesin being the proteinaceous link that holds sister chromatids together until the metaphase-to-anaphase transition. This change in residence time is tightly connected to DNA replication, as this event is required to establish sister chromatid cohesion genome-wide. On a biochemical level, this change in residence time is associated with acetylation of the SMC3 subunit of cohesin by the acetyltransferase ESCO1 and ESCO2 (reviewed in [166]). It is thought that the acetyltransferase ESCO2 is associated with the DNA replication machinery and acetylates cohesin in passing, thus in part mediating the dependence of cohesion establishment on DNA replication [167]. Whether the replication machinery traverses through cohesin rings, or whether cohesin is temporally unloaded and reloaded again is an open question. Recent studies, however, show that DNA replication does not cause dissociation of a marked cohesin pool from chromatin, suggesting that cohesin is not unloaded completely by the passing replication machinery [168].

While cohesin acetylation is necessary in all eukaryotes for cohesion establishment, in mammals this is not enough. Here, the action of a protein called

sororin is needed to mediate establishment of a "stable" cohesin pool both in cell free extracts and *in vivo* [169, 170]. Interestingly, while "stable" cohesin has a residence time on chromatin of multiple hours, sororin is much more transiently bound, unbinding after 1 min on average [171]. Moreover, sororin is constantly required for sister chromatid cohesion, as degradation in G2 - after cohesin establishment - results in loss of cohesion in the subsequent mitosis [171]. On a mechanistic level, experiments showing that sororin is not required for sister chromatid cohesion in cultured human cells in the absence of Wapl, suggested that sororin counteracts the cohesin releasing activity of Wapl/Pds5. These findings were then corroborated by biochemical analyses showing that sororin is capable of displacing Wapl from Pds5 *in vitro* [169]. On a genomics level, ChIP-seq analyses revealed that DNA replication causes genome-wide binding of sororin, overlapping cohesin ChIP-seq peaks to a high degree. This suggests that cohesive cohesin binds the entire genome after completion of S-Phase and that cohesion is established genome-wide [171].

However, after entry into mitosis, the bulk of cohesive cohesin is unloaded again from chromatin through the action of the prophase pathway, with only cohesin close to the centromere remaining [98]. During this process, the cohesin subunit SA2 is phosphorylated through the action of mitotic kinases, leading to the dissociation of sororin and subsequent unloading of cohesin through Wapl/Pds5 [172]. Cohesin at the centromere is protected from this unloading by the protein Shugoshin, most likely via its interaction with the phosphatase PP2A [173]. It is thought that this centromeric cohesin is responsible for sister chromatid cohesion in mitosis, providing a mechanical connection between sister chromatids. This connection is thought to provide a counter-force to the pulling and pushing action of the mitotic spindle, contributing to the congression of mitotic chromosomes at the metaphase plate and preventing premature segregation of sister chromatids to the respective spindle poles. This connection is then removed in a controlled way at the onset of the metaphase-to-anaphase transition: If bipolar attachment of sister chromatids has been achieved and the SAC was satisfied, APC/C binds to Cdc20 and ubiquitylates securin, which is degraded and releases separase. Separase then proteolytically cleaves centromeric cohesin and initiates chromosome segregation and mitotic exit. A failure of this process can lead to massive chromosome missegregation, which in turn potentially causes cell death and malignant transformation of the resulting daughter cells (reviewed in [174]).

1.5 Open questions in sister chromatid biology

While sister chromatid biology is of central importance to all areas of life, we still don't understand many aspects of it. First, we do not know how a single protein complex - cohesin - is able to balance two seemingly unrelated functions, namely holding sister chromatids together, while also shaping the complex 3D-structure of the interphase genome. Here, it is likely that the two populations of cohesin that emerge during DNA replication - dynamic and stable cohesin [162] (see 1.2.3 and 1.4.2 for more details) - represent a hint of

this separation of function. However, how this happens exactly and how these two types of complexes coordinate on the replicated genome to perform these functions simultaneously is completely unknown.

Next, on a larger level, the cell balances the two opposing requirements of sister chromatid resolution and cohesion: Sister chromatids need to resolve to be able to attach to opposing spindle poles, but also need to stay cohesed to prevent premature sister chromatid separation during mitosis. We know from prior studies that sister chromatid cohesion is established during S-phase (reviewed in [166]) and persists into mitosis, when sister chromatid resolution is thought to happen simultaneously with chromosome condensation [158]. However, these studies have only looked at this process at a stage when sister chromatids were visible as clearly separated rods during prometaphase. It is thus not known whether resolution is coordinated with cohesion at an earlier stage, potentially starting individualization prior to mitotic entry.

Additionally, sister chromatid cohesion is established genome-wide during S-phase and then removed again in mitosis, only leaving a small fraction of it intact [98]. This practice constitutes a huge energy investment of the cell since cohesion establishment is connected to extensive acetylation (reviewed in [166]), making it unlikely that genome-wide cohesion does not serve a cellular function. However, so far, no study has addressed this issue and we still don't know why cohesion is established on chromosomal arms in interphase nor why the prophase pathway exists.

On an epigenetics level, as soon as the genome is replicated, all the intricate mechanisms described in 1.2 that shape the 3D-genome to effect the host of different biological functions mentioned in 1.3 need to adapt to accommodate the fact that a second sister chromatid is present. Here, it is particularly daunting how a duplication of all the regulatory elements specifying the exact gene dosage of key regulators can be compensated for. This likely entails complex interactions of chromatin state on the two sister chromatids to ensure correct gene dosage. However, how this exactly happens is completely unknown.

Finally, as mentioned in 1.3.2, one of the key functions that involves interactions between sister chromatids is in the context of HDR upon DNA damage. Here, homologous regions need to find each other within the vast extent of the human genome to template resynthesis of damaged DNA on the second sister chromatid [175]. While many factors have been found that are involved in this process and it is known that sister chromatid proximity is required [146, 147], we still don't know how DNA structure is changed to permit effective homology search and how the epigenetic state influences this process.

1.5.1 Methodological short-comings

Many of these questions could be tackled, if we had a suitable method to look at sister chromatid specific structure with high resolution. If such a method existed, we could examine how cohesin structures both single chromatids and sister chromatid interactions simultaneously, potentially revealing hidden connections between these two functions. Furthermore, such a method would allow

us to observe sister chromatid resolution during mitosis at unprecedented detail, allowing us to pinpoint the interplay between this process and sister chromatid resolution. Additionally, sister chromatid specific structural information would allow us to look at cohesion in S-phase in a genome-wide fashion, most likely revealing patterns and correlations with the epigenome that could shed light on the function of cohesion during S-phase. Finally, detailed knowledge of sister chromatid conformation in G2 would be invaluable to the study of DNA repair: Heterogenous sister chromatid interactions could help explain different frequencies of DNA repair pathway choice at different genomic locations [176, 177] and allow visualization of homology search during HDR. Thus, a method to look at sister chromatid specific chromatin structure could allow us to tackle many of the posed open questions. However, currently all methods to examine chromatin structure genome-wide are blind to sister chromatid specific information.

1.6 Methods to describe DNA structure

1.6.1 Electron microscopy

Electron microscopy (EM) has been a valuable tool for cell biologists and biochemists for many years and is capable of delivering pictures with unmatched resolution, very recently reaching atomic resolution in a Cryo-EM setting (reviewed in [178]). This is possible by using an electron beam instead of light to image samples, thus overcoming the resolution limit of light microscopy. These capabilities have been exploited in the context of chromosome structure examination to both elucidate the looped configuration of mitotic chromosomes in pioneer studies as well as more recently to examine the interphase chromatin topology in great detail [104, 179].

While the resolution of EM is tremendously high, a downside in the context of DNA structure examination is that EM does not readily give sequence information. It is thus very challenging to derive conclusions about the specific structure of certain genomic regions and EM is therefore better suited to judge overall changes in DNA structure genome-wide.

1.6.2 FISH

As mentioned in 1.6.1, the resolution limit of light microscopy is not good enough to resolve DNA such that we can assign sequence information. Here, fluorescence in situ hybridization (FISH) provides one possible solution. The method works by hybridizing small DNA or RNA probes to target loci in the genome and then visualizing them directly in the case of labelled probes, or hybridizing a secondary probe carrying a fluorophore followed by imaging. This technique is very powerful and allows the specific visualization of target loci at remarkable resolution when combined with localization based super-resolution approaches. Here, entire genomic neighborhoods have been traced, revealing the exact path of the chromatin fiber [37, 127].

However, while much progress has been made in automating visualization of many loci through the use of microfluidics, FISH is still a very laborious

method and it will take considerable technological advances to employ it to visualize genome-wide chromatin conformation at high resolution.

1.6.3 Chromosome conformation capture techniques

The advent of high-throughput sequencing based techniques to look at chromosome structure marked a revolution in genome 3D-chromatin structure examination [21]. The principle underlying these approaches is remarkably simple. First, chromatin is fixed to preserve its 3D-conformation. Then, chromatin is digested using a nuclease, producing fragments that range from 150 bp in the case of MNase to several kb if a six-cutter restriction enzyme is used. This process is thought to produce "chunks" of chromatin that represent local neighborhoods, containing several of the cut DNA-pieces each. Then, these fragments are religated using a DNA-ligase, either *in cellulo* or in a dilute solution to ensure that most re-ligations happen within one of the chromatin "chunks". Then, the DNA is purified and - depending on the variant of chromosome conformation capture used - either examined via PCR (in the case of 3C) or subjected to high-throughput sequencing (in the case of Hi-C and Micro-C) (reviewed in [180]).

It may seem counter intuitive to first cut chromatin into small parts and then ligate them back together. However, if a DNA piece does not ligate with its previous part and instead combines with another piece of DNA present in the chromatin "chunk", this suggests that these two pieces were in close proximity and allows to deduce information about chromatin structure. Additionally, this information is then encoded in the DNA sequence itself, allowing researchers to transfer established techniques for analyzing DNA sequences and promoting scaling of this technology with revolutions in sequencing approaches.

Chromosome conformation capture techniques and especially the high throughput variant Hi-C have allowed significant insight into the structure of DNA, revealing loops [22] and TADs (see 1.1.2; [24, 134]) as well as elucidating the nested loop structure of mitotic chromosomes (see 1.4.1; [27, 28]). However, these techniques only measure contact frequency, often in a population of cells to ensure an adequate amount of sequencing reads. This means that direct deduction of chromatin structure in individual cells is challenging and may significantly deviate from the population picture. Nevertheless, Hi-C (and more recently Micro-C) is currently the state-of-the-art in 3D-structural analysis of the genome, both considering the resolution and multiplexing capabilities of this technique.

1.7 Aim of this thesis

While current methods to look at chromatin structure have allowed significant insight, they all suffer from a central problem when it comes to analyzing sister chromatids: They cannot distinguish identical DNA molecules. FISH relies on hybridizing probes to specific DNA sequences that are identical across sister chromatids. In replicated cells, FISH will thus produce two dots per targeted

locus, precluding sister assignment and resolution beyond the diffraction limit [181]. The same restriction holds for Hi-C: Here, ligation junctions within a sister chromatid cannot be distinguished from junctions between sister chromatids, suggesting that Hi-C experiments performed on cells with replicated genomes are always a mixture of these contact types.

The first aim of my thesis thus was to establish a technique that is capable of analyzing sister chromatid specific chromatin structure at high resolution. Specifically, I will present a Hi-C variant that employs sister chromatid specific labeling using artificial nucleotides to elucidate the detailed structure of replicated chromatids.

I then used this technique to generate the first genome-wide sister chromatid specific chromosome conformation maps of G2 and prometaphase cells. These maps both serve as a resource for future researchers aiming to tackle the intricacies of sister chromatid biology, but also allow correlative deduction of folding principles of the replicated human genome, which I will present.

Finally, I employ targeted degradation of two key-regulators of the cohesin complex to find that two different types of cohesin shape sister chromatid topology in the G2 cell cycle stage: Cohesive cohesin is necessary to maintain sister chromatid alignment, whereas loop extruding cohesin mediates local separation within the confines of TADs.

Taken together, this thesis establishes a novel way to examine the structure of replicated chromosomes and lays the foundation to understand key open questions in sister chromatid biology (see 1.4.2).

2 Methods

All entries in this section except 2.21.16, 2.21.17, 2.21.18, 2.21.19 and 2.14 are adapted from the methods section of [182].

2.1 Cell culture

All cell lines used in this thesis have been regularly tested negatively for mycoplasma contamination. The parental HeLa cell line ('Kyoto strain') was obtained from S. Narumiya (Kyoto University, Japan) and validated by a Multiplex human Cell line Authentication test (MCA). Cells were cultured in WT medium (DMEM high-glucose [Sigma], buffered with HEPES [Applichem] and Sodium bicarbonate [Sigma], adjusted to pH 7.1-7.3 and supplemented with 10 % [v/v] FCS [Gibco], 1 % [v/v] Penicillin/Streptomycin [Gibco] and 1 % [v/v] GlutaMAX [Gibco]) in a humidified incubator at 37 °C and 5 % CO₂. For culturing HeLa Sororin-AID and HeLa NIPBL-AID cells, medium was supplemented with 0.5 µg/ml Puromycin (Calbiochem). Cells were passaged every 48 h by dissociation using Trypsin/EDTA-Solution (Gibco).

2.2 Generation of cell lines

All cell lines used in this study are listed in Table 5 and the plasmids used in their generation are listed in Table 6. The HeLa Kyoto N-terminally-tagged Sororin auxin-inducible degron (AID) cell line was created by CRISPR/Cas9-mediated genome editing as described previously [41]. The gRNA sequences that were used for generating EGFP-AID-Sororin were CACCGCGCTCACCGGAGCGCTGAG, and CACCGACGTGAGGTCGAGCCGTTT together with the repair template 'EGFP-AID-Sororin-HR'. The primers used for genotyping were CTGCGGGGACAATACCAAT and CCGATCTCAGATTCCTGCCC. Subsequently, Tir1 expression was introduced by transducing a homozygous cell clone with lentiviruses using pRRL containing the constitutive promoter from spleen focus forming virus (SFFV) followed by *Oryza sativa* Tir1-3xMyc-T2A-Puro. Cells expressing Tir1 were selected by culturing in medium containing 2.5 µg/ml Puromycin (Sigma-Aldrich).

HeLa Kyoto N-terminally tagged AID-GFP-NIPBL cells were generated using CRISPR/Cas9-mediated genome editing based on a double nickase strategy [183]. The template for homologous recombination introduced sequences coding for monomeric EGFP (L221K) and the *Arabidopsis thaliana* IAA17 71-114 (AID*) mini-degron [184]. The gRNAs used for CRISPR/Cas9-mediated genome editing were CACCGCCATTCATCCTGAATTC and CACCGC-CCCATTACTACTCTTGCG together with the repair template 'AID-GFP-NIPBL-HR'. Single clones were obtained by sorting into 96-well plates on a BD FACS Aria III machine (BD Biosciences) and homozygous tagging was confirmed by PCR using the forward primer ATCGTGGGAACGTGCTTTGGA and reverse primer GCTCAGCCTCAATAGGTACCAACA. Subsequently, Tir1 expression was introduced as for HeLa Sororin-AID described above.

HeLa Kyoto RIEP H2B-mCherry cells were derived from HeLa Kyoto RIEP cells [185] using a lentiviral delivery system [185] to stably integrate a plasmid carrying H2B-mCherry (Lenti-H2B-mCherry). Cells were sorted into 96-wells to derive single clones using a BD Aria III instrument (BD Biosciences).

2.3 FACS analysis of cell cycle stage

For S-Phase release experiment, cells were trypsinized, washed with phosphate buffered saline (PBS; made in-house) and fixed using 70 % MeOH (Sigma) for at least 1 h at -20 °C. Cells were spun down (1100 x g; 1 min) and stained using 50 $\mu\text{g}/\text{ml}$ Propidium Iodide (Sigma) in a solution with 10 mM TRIS-HCl (Sigma; adjusted to pH 7.5), 5 mM MgCl_2 (Sigma) and 200 $\mu\text{g}/\text{ml}$ RNaseA (Qiagen) at 37 °C for 20 min. Samples were then measured on a FACSCanto II instrument (BD Biosciences). Analysis was performed using FlowJo(v10.6.0) as follows: Gate for cells in FSC-A/SSC-A, for single cells in FSC-A/SSC-H, histogram of PI-intensity. For all other FACS experiments, cells were trypsinized, washed with phosphate buffered saline (PBS; made in-house) and fixed using 70 % EtOH (Sigma) for at least 1 h at 4 °C. Cells were spun down (1100 x g; 1 min) and permeabilized using 0.25 % Triton-X100 (Sigma) in PBS for 15 min on ice. Cells were spun down again and stained using 0.25 μg α -H3S10p (Merck Millipore 04-817) in 1 % Bovine serum albumin (BSA; Sigma) for 1 h at room temperature (RT). Cells were washed once with 1 % BSA and then stained using 1:300 α -mouse-AF488 (Molecular Probes A11001) in 1 % BSA for 30 min at RT in the dark. Cells were washed once with 1 % BSA and incubated with a solution containing 200 $\mu\text{g}/\text{ml}$ RNase A (Qiagen) and 50 $\mu\text{g}/\text{ml}$ Propidium Iodide (Sigma) in PBS for 30 min at RT in the dark. Samples were then measured on a FACSCanto II instrument (BD Biosciences). Analysis was performed using FlowJo(v10.6.0) as follows: Gate for cells in FSC-A/SSC-A, for single cells in PE-A/PE-H, scatterplot of FITC and PI intensity.

2.4 DNA damage assay

WT HeLa Kyoto cells were seeded into an 8-well Lab-Tek (Thermo Scientific) and grown for 16 h. Then, different concentrations of 4sT (2mM-10mM; Carbo-synth) and 50 μM etoposide (Sigma) were added and cells were incubated for 24 h. For immunofluorescence (IF), cells were washed two times with PBS and fixed using 4 % formaldehyde (Sigma) in PBS for 5 min. Formaldehyde was quenched using 20 mM TRIS-HCl (Sigma; adjusted to pH 7.5) in PBS for 3 min and washed with PBS. Cells were permeabilized using 0.5 % Triton-X100 (Sigma) in PBS for 10 min. Then, cells were blocked using 2 % BSA in PBS for 30 min at RT, followed by incubation with 1:500 α -phospho- γ -H2A.X (ABCAM ab2893) in 2 % BSA [PBS] for 1.5 h at RT. Then, cells were washed 3x for 5 min using PBS, followed by incubation with 1:1000 α -mouse-AF488 (Molecular Probes A11001) in 2 % BSA [PBS] for 30 min at RT in the dark. Then, cells were washed one time using PBS for 5 min, followed by staining using 1 $\mu\text{g}/\text{ml}$ 4,6-diamidino-2-phenylindole (DAPI; Thermo Scientific) for 5 min. Then, cells

were washed again for 5 min in PBS. Samples were imaged on a customized Zeiss LSM780 microscope using a 20x, 0.8 NA, Oil DIC Plan-Apochromat objective (Zeiss). Images were analyzed using CellCognitionExplorer 1.0.2 [186] for segmentation and intensity extraction and Python scripts to visualize the data.

2.5 Viability assay

Cells carrying a stable H2B-mCherry integration were seeded into a 96 well imaging plate (Greiner) in imaging medium (custom; DMEM High-glucose [Gibco] without Riboflavin and Phenolred containing 10% [w/w] FCS [Gibco], 1% [w/w] P/S [Gibco] and 1 % [w/w] Glutamax [Gibco]) supplemented with 1 μ M TO-PRO[®]-3 (Thermo Fisher Scientific). After 16 h, compounds to be tested were added and imaging was started on a Molecular Devices ImageXpressMicro XL screening microscope with a reflection-based laser auto focus and a 10x, 0.75 NA, S Fluor dry objective (Nikon). Cells were maintained for 24 h in a microscopic stage incubator at 37 °C in a humidified atmosphere at 5 % CO₂ and images in the mCherry and TO-PRO[®]-3 channel were recorded every 2 h. Images were analyzed using CellCognition [187] for segmentation and intensity extraction and Python scripts to visualize the data.

2.6 Western Blot

Cell suspension (1 million cells/ml) was mixed with 6x SDS loading buffer and 10 mM DTT (Roche) and incubated at 95 °C for 10 min. Samples were separated on a NuPage 4-12% Bis-Tris Gel (Invitrogen) and transferred onto a Hybond P 0.45 polyvinylidene difluoride (PVDF) membrane (GE life sciences) using wet blotting. Sororin was probed using a custom antibody kindly provided by Jan-Michael Peters (1:500). GAPDH was probed using a polyclonal antibody (1: 2000; Abcam ab9485) and NIPBL was probed using a monoclonal antibody (1:1000; Absea 010702F01). Horseradish peroxidase-conjugated secondary antibodies were used (anti-rat Amersham NA935 at 1:2000; anti-rabbit Biorad 170-6515 at 1:5000), and blots were visualized using ECL Plus Western Blotting Substrate (Thermo Fisher Scientific). Western blots were imaged on a Biorad Imager operated by ImageLab.

2.7 Metaphase congression assay

Cells were synchronized to G2 as explained in the cell synchronization section. Then, 1 h before RO3306 wash-out, cells were supplemented with 250 ng/ml SIR-DNA (Spirochrome). Then, cells were washed 2x with imaging medium containing 250 ng/ml SIR-DNA, followed by imaging every 3 min for 120 min on a customized Zeiss LSM780 microscope at 37 °C and 5 % CO₂ using a 20x, 0.8 NA, Oil DIC Plan-Apochromat objective (Zeiss). Congression time was measured by visual inspection using Fiji and ImageJ [188] for all cells that entered mitosis in the indicated time frame. Results were visualized using Python.

2.8 Cell synchronization for scsHi-C

Cells were seeded into 25 cm² flasks and grown for 3 h, then supplemented with 2 mM Thymidine (Sigma). Cells were released 16 h later by washing 2 times with prewarmed WT medium. 8 h later, cells were supplemented with 3 μ g/ml aphidicolin (Sigma) and 2 mM 4sT (Carbosynth). Cells were released 16 h later by washing 2 times with PBS and addition of medium containing 2 mM 4sT (Carbosynth). For Sororin-AID experiments, 500 μ M Indole-3-acetic acid (Sigma) was added 1 h prior to S-phase release. For S-phase release experiments, samples were taken at the indicated time-points. For synchronization to G2 and prometaphase, after 4 h release, 9 μ g RO-3306 (Sigma) or 200 ng/ml nocodazole (Sigma) were added respectively. For NIPBL-AID experiments, 500 μ M Indole-3-acetic acid (Sigma) was added 8 h after released. Samples were processed 16 h later. Cells were harvested by washing with PBS, followed by trypsinisation and resuspension in WT medium. Cells were then spun down, washed again with PBS, followed by fixation for 4 min in 1 % formaldehyde (Sigma). Cell pellets were stored at -20 °C or processed immediately.

2.9 Hi-C sample preparation

Fixed cells were permeabilized using ice-cold Hi-C lysis buffer (10 mM TRIS-HCl pH 8 [Sigma], 10 mM NaCl [Sigma], 0.2 % Nonidet P-40 substitute [Sigma], 1x Complete EDTA-free Protease inhibitor [Roche]) for 30 min at 4 °C. Then, cells were spun down (2500 x g for 5 min), supernatant was discarded, and digestion mix was added (375 U DpnII [NEB] in 1x DpnII buffer [NEB]) and cells incubated for 16 h at 37 °C under rotation. Then, cells were spun down, supernatant was discarded and fill-in mix was added (38 μ M Biotin-14-dATP [Thermo Fisher Scientific], 38 μ M dCTP, dGTP and dTTP [Thermo Fisher Scientific], 50 U Klenow Polymerase [NEB], 1x NEB 2 buffer) and cells incubated for 1 h at 37 °C under rotation. Then, cells were spun down again, and ligation mix was added (1x T4 DNA ligase buffer [Thermo Fisher Scientific], 0.1 % Triton X-100 [Sigma], 100 μ g/ml BSA [Sigma], 50 U T4 DNA ligase [Thermo Fisher Scientific]) and incubated at RT for 4 h. Then, cells were spun down, resuspended in 200 μ l PBS and gDNA was purified using the DNeasy Blood and Tissue kit (Quiagen). DNA was transferred to a Covaris microTUBE (Covaris) and sheared on a Covaris S2 instrument (Duty cycle 10 %, Intensity 5.0, Cycles/burst 200) for 25 s. Double size selection was performed by employing AMPure XP beads (Beckman Coulter) first at 0.8-fold sample volume according to the standard protocol, followed by transfer of the supernatant and bead application at 0.12-fold sample volume. The resulting DNA was bound to Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific) in Biotin binding buffer (5 mM Tris-HCl pH 7.5 [Sigma], 0.5 mM EDTA [AppliChem], 1 M NaCl [Merck]) for 1 h at RT. Beads were then washed 2x in Tween wash buffer (5 mM Tris-HCl [Sigma], 0.5 mM EDTA [AppliChem], 1 M NaCl [Merck], 0.05 % Tween-20 [Sigma]) and 1x in H₂O. Beads were resuspended in H₂O and library preparation was performed using the NEBNext Ultra II DNA library

prep kit for Illumina [NEB] according to the standard protocol. After this, beads were washed 4x using Tween wash buffer and DNA was eluted using 95 % formamide (Sigma), 10 mM EDTA (AppliChem) at 65 °C for 2 min. DNA was then precipitated using 80 % EtOH (Sigma), washed with 75 % EtOH and resuspended in H₂O. Then, 4sT was converted to methyl-cytosine using OsO₄ / NH₄Cl (see below), followed by qPCR according the NEBUltra Ultra II DNA library prep kit for Illumina [NEB]. The finished libraries were purified using AMPure XP beads (Beckman Coulter) at 0.9x sample volume following the standard protocol.

2.10 Sequencing

Sequencing of all samples was performed either on an Illumina NovaSeq 6000 instrument using patterned SP flowcells using read-mode PE250 or on an Illumina MiSeq instrument using a Nanoflowcell using read-mode PE250 (v2).

2.11 Quantification of 4sT incorporation into gDNA

HeLa Kyoto cells were cultured for 24 h after splitting and supplemented with either H₂O or 2 mM 4sT (Carbosynth) and then cultured for 5 days. DNA was extracted using the DNeasy Blood and Tissue kit (Quiagen). Extracted DNA was digested using 1 U DNA Degradase Plus (Zymo Research) in 1x DNA Degradase buffer (Zymo Research) for 2 h at 37 °C. Deoxyribonucleosides were quantified by injecting 1 μ l of the acidified digest on a RSLC ultimate 3000 (Thermo Fisher Scientific) directly coupled to a TSQ Vantage mass spectrometer (Thermo Fisher Scientific) via electrospray ionization. A Kinetex C18 column was used (100 Å, 150 x 2.1 mm), employing a flow rate of 100 μ l/min. An 8-minute-long linear gradient was used from 0% A (1 % acetonitrile, 0.1 % formic acid in water) to 60% B (0.1 % formic acid in acetonitrile). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed by employing the selected reaction monitoring (SRM) mode of the instrument. Thymidine and 4-thio-thymidine were quantified by analyzing the in-source fragments of the respective nucleotides at an elevated declustering potential. For thymidine the transition 127.1 m/z \rightarrow 54.1 m/z (CE 23 V) and for 4-thio-thymidine the transition 143.1 m/z \rightarrow 126.1 m/z (CE 25 V) were used. A calibration curve of synthetic standard nucleosides was used to quantify the relative percentage of 4-thio-thymidine in total thymidine in the biological samples. Each sample was measured in duplicate.

2.12 Conversion analysis of 4sT on synthetic oligos

Conversion was done similarly as described in [189]: A 4sT-containing oligonucleotide was synthesized as described below. The molecular weight of the oligonucleotide with and without OsO₄ / NH₄Cl treatment (see below) was analyzed on a Finnigan LCQ Advantage MAX ion trap instrument connected

to an Amersham Ettan micro LC system in the negative-ion mode with a potential of -4 kV applied to the spray needle. LC: Sample (200 pmol RNA dissolved in 30 μ l of 20 mM EDTA solution; average injection volume: 30 μ l), column (Waters XTerra[®] MS, C18 2.5 m; 2.1 x 50 mm) at 21 °C; flow rate: 30 μ l min⁻¹; eluent A: Et₃N (8.6 mM), 1,1,1,3,3,3-hexafluoroisopropanol (100 mM) in H₂O (pH 8.0); eluent B: MeOH; gradient: 0–100% B in A within 30 min; UV detection at 254 nm.

2.13 Conventional sequencing library preparation to estimate 4sT mutation rates

Cells were harvested by trypsinization, spun down at 1100 x g for 1 min and the supernatant was discarded. Then, cells were spun down, resuspended in 200 μ l PBS and gDNA was purified using the DNeasy Blood and Tissue kit (Quiagen). DNA was transferred to Covaris microTUBE (Covaris) and sheared on a Covaris S2 instrument (Duty cycle 10 %, Intensity 5.0, Cycles/burst 200) for 25 s. Double size selection was performed by employing AMPure XP beads (Beckman Coulter) first at 0.8-fold sample volume according to the standard protocol, followed by transfer of the supernatant and bead application at 0.12-fold sample volume. DNA library preparation was performed with the resulting DNA using the NEBNext Ultra II DNA library prep kit for Illumina [NEB] according to the standard protocol. The unamplified libraries were then treated using OsO₄ (see below) and amplified according the NEBNext Ultra II DNA library prep kit for Illumina [NEB]. The finished libraries were purified using AMPure XP beads (Beckman Coulter) at 0.9x sample volume following the standard protocol.

2.14 Iodoacetamide mediated conversion of 4sT

Iodoacetamide conversion of 4sT containing oligos and genomic DNA was done as described in [190]. Briefly, DNA was incubated with conversion solution (50 % DMSO [Sigma], 10 mM iodoacetamide [Sigma], 50 mM sodiumphosphate pH 8 [Sigma]) for 15 min at 50 °C and quenched using 100 mM DTT [Sigma]. Material was then either PCR amplified with different polymerases (see 3.2.3) and subjected to sanger sequencing or used for conventional sequencing library preparation (see 2.13).

2.15 OsO₄ / NH₄Cl-mediated conversion of 4sT

For synthetic oligos, lyophilized DNA (1 nmol) was dissolved in water (10 μ l) and denatured for 2 min at 90 °C. Then, the solution was heated to 60 °C and NH₄Cl buffer (2 μ l, 2 M, pH 8.88) and OsO₄ solution (10 μ l, 1 mM) were added to yield final concentrations of 0.45 mM OsO₄ and 180 mM NH₄Cl in a total volume of 22 μ l. The reaction mixture was incubated for three hours at 60 °C. The DNA was precipitated by adding 90 μ l of precipitation solution (made of water (650 μ l), aqueous NaOAc solution (150 μ l; 1 M, pH 5.2), and

glycogen (10 μ l; 20 mg/ml) and 250 μ l of cold ethanol. The mixture was kept at -20 °C for 30 minutes, followed by centrifugation (13,000 rpm, 4 °C, 30 min). The supernatant was discarded, and the precipitated DNA analyzed by anion exchange HPLC and mass spectrometry (see above). Genomic DNA was incubated with 0.45 mM OsO₄ (Sigma) and 200 mM NH₄Cl (Sigma) adjusted with NH₃ (Honeywell Fluka) to pH 8.88 first for 5 min at 95 °C followed by 60 °C for 3 h on a T100 Thermal Cycler (Bio-Rad) with the heated lid set to 105 °C. DNA was then precipitated using 80 % EtOH, washed with 75 % EtOH and resuspended in H₂O.

2.16 Melting curve analysis of hairpin oligo

The absorbance versus temperature profiles were recorded at 260 nm on a Varian Cary 100 spectrophotometer equipped with a multiple cell holder and a Peltier temperature-control device. The 4sT containing DNA hairpins and their reference oligonucleotides were measured at a concentration of 2 μ M in melting buffer (150 mM NaCl, 10 mM Na₂HPO₄ at pH 7.0). Three cycles of cooling and heating between 30 °C to 95 °C and a rate of 0.7 °C min⁻¹ were recorded. Sample preparation: An aliquot of oligonucleotide stock solution was lyophilized, dissolved in 1 ml of melting buffer to give the desired final concentration. The solution was transferred into a quartz cuvette and degassed. A layer of silicon oil was placed on the surface of the solution to minimize evaporation during the measurements.

2.17 Synthesis of a 4-thiothymidine (4sT) phosphoramidite building block

General information: Chemical reagents and solvents were purchased from commercial suppliers and used without further purification. Thymidine phosphoramidite was purchased from ChemGenes. Organic solvents for reactions were dried overnight over freshly activated molecular sieves (3 Å). All reactions were carried out under argon atmosphere. Analytical thin-layer chromatography (TLC) was carried out on Marchery-Nagel (Polygram SIL G/UV254, 0.2 mm silica gel) plates. Flash column chromatography was carried out on silica gel 60 (70-230 mesh). ¹H and ³¹P NMR spectra were recorded on Bruker 400 and 700 MHz spectrometers. The chemical shifts are referenced to the residual proton signal of the deuterated solvents: CDCl₃ (7.26 ppm), d₆-DMSO (2.50 ppm) for ¹H NMR spectra; ³¹P-shifts are relative to external 85% phosphoric acid. ¹H assignments were based on COSY experiments. Mass spectrometric analysis of low molecular weight compounds was performed on a Thermo Scientific Q Exactive Orbitrap mass spectrometer in the positive ion mode. Procedure: Thymidine phosphoramidite (229 mg, 0.307 mmol) was dissolved in dry dichloromethane (3 ml). Then, triethylamine (37 mg, 51 μ l, 0.366 mmol), 4-dimethylaminopyridine (1 mg) and 2-mesitylsulfonyl chloride (56 mg, 0.256 mmol) were added. The solution was stirred at room temperature for 2 hours.

In the meantime, 3,3'-dithiobis(propionitrile) (200 mg, 1.16 mmol) was suspended in aqueous 2 M HCl solution, followed by slow addition of zinc powder (220 mg, 3.36 mmol). The mixture was stirred at room temperature for one hour, extracted three times with dichloromethane, dried over Na₂SO₄ and evaporated. The 3-mercaptopropionitrile was obtained as slightly yellow oil. Then, *N*-methylpyrrolidine (261 mg, 319 μ l, 3.07 mmol) and the freshly prepared 3-mercaptopropionitrile (133 mg, 1.54 mmol) were mixed in dry dichloromethane (1 ml) and added to the reaction mixture containing the activated nucleoside. Stirring was continued at 0 °C (ice bath) for one hour. Finally, the solution was diluted with dichloromethane, washed with saturated NaHCO₃, dried over Na₂SO₄ and evaporated. The crude product was purified by column chromatography on silica gel (ethyl acetate/cyclohexane, 15:100 – 75:25). Yield: 172 mg (69%) white foam. TLC (ethyl acetate/cyclohexane, 1:1): R_f = 0.1. HR-ESI-MS (m/z): [M+Na]⁺ calculated: [836.3217]; found: [836.3110]. ¹H-NMR (400 MHz, CDCl₃): δ 1.15 – 1.18 (m, 12 H, 2x (H₃C)₂CHN); 1.43 (s, 3H, H₃C(5)); 2.22 – 2.28 (1H, HaC(2')); 2.33 – 2.35 (1H, HC(N)); 2.53 – 2.56 (1H, HC(N)); 2.60 – 2.66 (1H, HbC(2')); 2.78 – 2.88 (2H, H₂CCN); 3.26 – 3.34 (m, 4H, H₂CS, H₂C(5')); 3.44 – 3.54 (m, 4H, H₂CCN, H₂CO); 3.73 (s, 6H, 2x H₃CO(DMT)); 4.13 (m, 1H, HC(4')); 4.58 (m, 1H, HC(3')); 6.21 (m, 1H, HC(1')); 6.73 – 6.78 (m, 4H, HC(DMT)); 7.27 – 7.41 (m, 9H, HC(DMT)); 7.87 (s, 1H, HC(6)) ppm. ³¹P-NMR (282 MHz, CDCl₃): δ 148.64; 149.33 ppm (Extended Data Fig. 9b-c).

2.18 Solid-phase synthesis of 4sT containing DNA

CCGGAAGGTATGAACC(4sT)TCCG was synthesized by automated solid-phase synthesis (ABI 392 Nucleic Acids Synthesizer) using standard DNA nucleoside phosphoramidites (ChemGenes), the 4-thiothymidine phosphoramidite (as described above), and polystyrene support (GE Healthcare, Primer Support 80s, 80 μ mol per g; PS 200). The following set-up was applied: detritylation (80 s) with dichloroacetic acid/1,2-dichloroethane (4/96); coupling (2.0 min) with phosphoramidites/acetonitrile (0.1 M, 130 μ l) and 5-(benzylthio)-1*H*-tetrazole/acetonitrile (0.3 M, 360 μ l); capping (0.4 min, three cycles) with Cap A: 4-(dimethylamino)pyridine in acetonitrile (0.5 M) and Cap B: Ac₂O/symcollidine/acetonitrile (2/3/5); oxidation (1.0 min) with I₂ (20 mM) in THF/pyridine/H₂O (35/10/5). Acetonitrile (DNA synthesis grade) was purchased from Anteris Systems GmbH. Acetonitrile, acetonitrile solutions of amidites, and acetonitrile solution of 5-(benzylthio)-1*H*-tetrazole were dried over activated molecular sieves (3 Å) overnight.

2.19 Deprotection of 4sT containing DNA

After DNA strand assembly, the beads were treated with 1,8-diazabicyclo[5.4.0]undec-7-en (DBU) in anhydrous acetonitrile (5 ml, 1 M) for three hours. Subsequently, the beads were incubated with *tert.*-butyl amine/ethanol/water (1/1/2, v/v/v) and dithiothreitol (50 mM) for five hours at 55 °C. Then, the supernatant was removed and the beads were washed three times with 1 ml eth-

anol/water (1/1). The combined phases were evaporated to dryness. The crude DNA was dissolved in water (1 ml).

2.20 Analysis and purification of 4sT containing DNA

After the deprotection, the crude DNA was analyzed by anion-exchange chromatography on a Dionex DNAPac PA-100 column (4 mm x 250 mm) at 80 °C. Flow rate: 1 ml min⁻¹, eluant A: 25 mM Tris-HCl (pH 8.0), 6 M urea; eluant B: 25 mM Tris-HCl (pH 8.0), 0.5 M NaClO₄, 6 M urea; gradient: 0 – 60% B in A within 50 min, UV detection at 260 nm. The DNA was purified on a semipreparative Dionex DNAPac PA-100 column (9 mm x 250 mm) at 80 °C with flow rate 2 ml min⁻¹, using the same eluents A and B as for analytical analysis, but with flat gradients that were optimized according to the length of the oligonucleotide. DNA containing fractions were loaded on a C18 SepPak Plus cartridge (Waters/Millipore), washed with 0.1 - 0.15 M (Et₃NH)⁺HCO₃⁻, H₂O and eluted with H₂O/CH₃CN (1/1). DNA containing fractions were evaporated to dryness and then dissolved in 1 ml water (stock solutions for storage at -20 °C). The quality of purified DNA was again analyzed by anion-exchange chromatography. The molecular weight of the DNA was analyzed by LC-ESI MS. Yields were determined by UV photometrical analysis of oligonucleotide solutions.

2.21 Sequencing data analysis

2.21.1 Calling HeLa single-nucleotide polymorphisms

In order to discriminate between 4sT-introduced mutations and single-nucleotide polymorphisms (SNPs), HeLa Kyoto SNPs were called on DNA-seq data from WT HeLa cells and a new consensus genome based on hg19 was constructed using bcftools (<https://github.com/samtools/bcftools>).

2.21.2 Mutation rate analysis

First, sequencing data was aligned to the hg19 genome containing HeLa SNPs using *bowtie2*. Then, relative point mutation rates for all possible point mutations were calculated using a custom Python script as follows: Absolute point mutation rates were counted and normalized to the total covered amount of the source base (e.g. for T-to-C normalization was done to T in the reference genome). These values were then normalized to an unlabelled control sample that had undergone OsO₄ / NH₄Cl treatment. The incorporated amount of 4sT was determined by sequencing analysis, based on calculating the ratio of the sum of T-to-C and the A-to-G absolute mutation rates to the sum of all Ts and all As, respectively. The fraction of read pairs being labelled in samples derived from DNA-seq libraries was calculated as follows: Only high-quality read-pairs (alignment score > 20, longer than 240 bp, no more than one non-signature mutation [other than T-to-C or A-to-G]) were counted. Only point mutations that had a Phred-score higher than 30 were counted. The number of T-to-C

and A-to-G mutations were counted on both read-pairs of a paired-end read. The read halves were then assigned to be labelled if they contained 2 or more signature point mutations. A read-pair was classified as double labelled if both halves were labelled. To calculate a histogram of signature mutations per read, the number of T-to-C and A-to-G mutations per read was counted and plotted for control samples (not treated with 4sT) and samples treated with 4sT for 5 days.

2.21.3 Correlation analysis of loci splitting frequency and FISH with scsHi-C

The average number of trans-sister contacts (balanced as described in 2.21.4) was extracted at target sites of the 16 gRNAs from Stanyte et al. [181] within a 600 kb window and $1 - \text{average}(\text{trans} - \text{sister contacts})$ was calculated and converted to a Z-score by subtracting the mean and dividing by the standard deviation. The resulting value correlated with the average frequency of split loci 1.2 h before G2 phase [181] for 11 WT G2 replicates. A similar analysis was performed for the 5 loci for which both gRNA splitting data and FISH data was available.

2.21.4 Hi-C data preprocessing

Hi-C samples were preprocessed using a custom nextflow pipeline (https://github.com/gerlichlab/scshic_pipeline). Briefly, bcl2 files were first demultiplexed using bcl2tofastq. Then, fastq files were aligned to hg19 with HeLa SNPs using bwa, aligning read pairs independently. Then, pairsam files were constructed using pairtools (<https://github.com/mirnylab/pairtools>), followed by sorting and deduplicating. Then, reads were split into cis-sister and trans-sister contacts based on the presence of signature mutations using pairtools select: A read was assigned to the Watson strand if it contained two or more A-to-G mutations and no T-to-C mutations. Similarly, if a read contained two or more T-to-C mutations, but no A-to-G mutations it was assigned to the Crick strand. Then, contacts were classified as cis sister contacts if (after correcting for the opposite read-strandedness of Illumina sequencing of the two mates) both mates mapped to the same strand. Conversely, contacts were classified as trans-sister contacts if the two mates mapped to opposing strands. Then, cooler (<https://github.com/mirnylab/cooler>; [191]) was used to construct .cool files, and binned at multiple resolutions. After completion of the nextflow pipeline, cis-sister and trans-sister contacts were merged, and the resulting file balanced using cooler [192]. Balancing was done as described in [192], excluding the 0th-diagonal to avoid Hi-C artefacts. Bins that had marginal read-count with a median absolute deviation (MAD) > 5 based on the genome-wide distribution were excluded from balancing and further analysis. Then, the resulting weights were transferred to the individual cooler files containing the cis-sister and trans-sister contacts. Hi-C matrices containing all contacts (not stratified into cis-sister and trans-sister contacts) were balanced similarly.

2.21.5 Hi-C genome scaling plots

Scaling plots were calculated separately for cis-sister and trans-sister contacts using pairlib (<https://github.com/mirnylab/pairlib>). Briefly, contacts were binned into geometrically spaced bins from 10 kb to 100 Mb with a total of 64 bins. Then, the number of contacts in each bin was divided by the number of covered base pairs. When multiple samples were compared on the same plot, they were down sampled to contain an equal number of combined cis-sister- and trans-sister contacts that are separated further than 1 kb using the NGS package (<https://github.com/gerlichlab/ngs>).

2.21.6 Observed-over-expected transformation of Hi-C matrices

The expected number of Hi-C contacts e at a given genomic separation k in Hi-C bin units was calculated using the cooltools (<https://github.com/mirnylab/cooltools>) package:

$$e(k) = \frac{1}{v} \sum_{0 \leq i \leq m, j=i+k} M_{i,j}$$

with $e(k)$ being the expected number of Hi-C contacts separated by k Hi-C bins, v being the number of valid bin-interactions with separation k (interactions between bins that were assigned valid balancing weights during the ICE-procedure) and M being the ICE-corrected Hi-C interaction matrix containing m bins. Note that the expected number of contacts is obtained from the upper-triangular part of the Hi-C matrix only since the matrix is symmetric. The observed-over-expected Hi-C matrix (OE) was then obtained as follows:

$$OE_{i,j} = \frac{1}{e(|j-i|)} M_{i,j}$$

2.21.7 Hi-C aggregate maps at TAD-centers

Aggregate maps of Hi-C submatrices around TAD-centers of genomic neighborhoods were calculated within a custom ipython notebook using the cooltools package (<https://github.com/mirnylab/cooltools>). First, 900 kb-sized submatrices centered around TAD-centers were extracted and the pixel-wise average of the ICE-corrected contacts over all the windows was calculated. In order to avoid Hi-C artefacts, the main diagonal as well as the neighboring diagonals were blanked out in the plot.

2.21.8 Extraction of sample regions

Sample regions of ICE-corrected Hi-C-matrices were extracted using the cooler Python API. For calculations of ratio maps, the observed/expected values were calculated for the respective ROI using the cooltools (<https://github.com/mirnylab/cooltools>) package as described above and then trans/cis ratios

were calculated. Before plotting, a pseudocount of 0.01 was added to avoid removal of 0-bins from the image during log-transformation.

2.21.9 TAD-calling

TAD-calling was done using OnTAD[35] on a G2 WT Hi-C matrix construct from all generated contacts merged over all replicates. OnTAD calls TADs based on the insulation score, which was shown previously to correlate with enrichment of known boundary factors such as CTCF and SMC3[33]. The bin size for TAD-calling was 50 kb and the only parameter that was changed from the standard set was the maximum TAD-size, which was restricted to 6 Mb. The TADs used for all analysis in this paper can be found here https://github.com/gerlichlab/scsHiCanalysis/blob/master/data/TADs_final.bedpe.

2.21.10 Pairing-score and contact-density calculation

I defined the contact density as the average contact frequency within a sliding window of half-length w in Hi-C bin units:

$$CD(i) = \frac{1}{v} \sum_{i-w \leq m, n \leq i+w} M_{m,n}$$

with $CD(i)$ denoting the contact density at bin i , M the Hi-C matrix (either ICE-corrected or observed-over-expected transformed), v being the number of valid Hi-C pixels within the window of summation (interactions between bins that were assigned valid balancing weights during the ICE-procedure; Note that the main diagonal does not contain valid pixels) within the sliding window. I then defined the pairing-score to be the contact density subtracted by the genome-wide average and converted to a Z-score by dividing by the genome-wide standard deviation:

$$PS(i) = \frac{CD(i) - median(CD)}{std(CD)}$$

with $PS(i)$ referring to the pairing score at genomic bin i , $median(CD)$ referring to the genome-wide median of (CD) and $std(CD)$ referring to the genome-wide standard deviation of (CD) .

2.21.11 Stack-up analysis of line profiles

Stack-ups of line profiles along a set of regions was calculated as follows:

1. The contact density within a sliding diamond of half-length w was calculated along each region within the set of regions as described above (either for ICE-corrected matrices or observed-expected-transformed matrices), resulting in a vector of size n for each region.

2. Then, these vectors were stacked into an $(m \times n)$ matrix with m denoting the number of regions and n denoting the length of the line profile along each region as described in 1.
3. Finally, the rows of the matrix were sorted based either on the size of TADs within the regions of interest or based on the average line profile signal within the center bins – bins with index in the interval $[\lfloor \text{median}(0, n) \rfloor - 5, \lfloor \text{median}(0, n) \rfloor + 5]$.

For display of observed-over-expected transformed values, a pseudocount of 0.01 was added before log-transformation. Moreover, line profiles that only contained invalid Hi-C bins were removed from the stack-up.

2.21.12 LOLA-analysis of highly paired and highly unpaired TADs

LOLA [193] probes whether genomic features like ChIP-seq peaks are significantly enriched in a ‘query’ set of regions compared to all possible regions. The probability of observing a given number of ChIP-seq peaks (or a greater number) by chance (p-value) is then calculated using Fisher’s exact test. LOLA enrichment analysis was done for TADs with high trans-sister contact density and low trans-sister contact density as follows: The average contact density (see above for details) for every annotated TAD (see above for details) was calculated for a window with size w that corresponded to the size of the respective TAD centered on the TAD-center $\lfloor \frac{TADstart+TADend}{2} \rfloor$ for a Hi-C contact matrix binned at 10 kb and containing ICE-corrected trans-sister contacts. The 90th and 10th percentile of trans-sister contact density within these TADs was calculated and the TADs that showed a trans-sister contact density larger than the 90th percentile were denoted “highly paired”, whereas TADs that had a trans-sister contact density smaller than the 10th percentile were denoted “highly unpaired”. LOLA was then run using the LOLA Extended data set [193] for the highly paired and highly unpaired TAD regions using all TADs as the region universe. Only chromatin data sets that were from HeLa cells are shown. A pseudocount of 0.05 was added to be able to display data sets that were very close to 0.

2.21.13 HiCRep analysis

HiCRep [194] was run using the python wrapper *hicreppy* (<https://github.com/cmdoret/hicreppy>) for all conditions tested using a Hi-C matrix with bin size 100 kb, a smoothing parameter $v = 10$, a maximum distance of $maxdist = 10^6$ bp without subsampling.

2.21.14 False-positive rate estimation of double labelled reads and trans-sister contacts

To estimate the false-positive rate of double labelled Hi-C contact, a Hi-C sample from cells that were grown in the absence of 4sT was analyzed. The reasoning was that all reads that were classified as double labelled in this condition would

be false-positives. A Hi-C contact was annotated as double labelled, if both half-reads exhibited more than a threshold amount of signature mutations (A to G or T to C). Then, the false-positive rate

$$FPR(\text{Double labelled}) = \frac{\text{Double labelled contacts}}{\text{All contacts}}$$

was calculated for different thresholds of signature mutations. To estimate the false-positive rate of trans-sister contact assignment, an approach developed in [195] was adapted. I assumed that all contacts of a G2 WT Hi-C sample that exhibited a genomic separation below 1 kb were Hi-C artefacts, namely uncut DNA. Such contacts should be exclusively classified as cis-sister contacts since a successful digestion and re-ligation is needed to generate a trans-sister contact. The false-positive rate of trans-sister contacts was therefore defined as

$$FPR(\text{trans} - \text{sister}) = \frac{\text{trans} - \text{sister below 1kb}}{\text{Double labelled below 1kb}}$$

To estimate the number of wrongly assigned trans-sister Hi-C contacts I assumed that the false-positive rate of trans-sister assignment is independent of genomic separation and that the amount of wrongly assigned cis-sister contacts is negligible. I further assumed that contacts with genomic separation larger than 1 kb constitute true Hi-C contacts. I therefore defined the fraction of wrong trans-sister Hi-C contacts *WTC* as

$$WTC = \frac{FPR(\text{trans} - \text{sister}) \times \text{cis} - \text{sister above 1kb}}{\text{trans} - \text{sister above 1kb}}$$

and calculated this value for different signature mutation thresholds.

2.21.15 H3K27me3 enrichment analysis

In order to calculate the enrichment of H3K27me3 ChIP-seq signal at highly paired and highly unpaired TADs, ChIP-seq data from [196] was downloaded and the average enrichment of H3K27me3 with respect to the control data set calculated for both region sets

2.21.16 Prediction of pairing score

ChIP-seq data sets for feature generation were collected from ENCODE [197] and from data sets shared from the lab of Jan-Michael Peters. For these data sets, both the absolute enrichment within a window of 50 kb as well as the number of peaks in this window was taken as a feature. Additionally, insulation scores (200 kb window size, 50 kb binsize on cis-sister contacts) and eigenvectors (all-contacts; 200 kb binsize) derived from Hi-C data presented within this thesis for G2 synchronized cells using the *cooltools* and *cooler* package (<https://github.com/mirnylab/cooltools>; <https://github.com/mirnylab/cooler>) were included. See Table 7 for all features.

As dependent variable, the pairing score (see 2.21.10) was calculated for trans-sister contacts at a window of 50 kb from matrix with binsize 10 kb. Then, outlier data points - defined as having a median absolute deviation of more than 3 - were removed and the resulting values standardized and missing values imputed using the mean of all data points.

Features were first standardized and missing values imputed using the per-feature mean before data was merged with the dependent variable. Then, data set was split into a training and test set (90 %: 10 % of all data points). A random forest regressor from the scikit-learn package (<https://scikit-learn.org/stable/>) was fit to the training set with the following parameters:

`n_estimators = 600; min_samples_split = 2; min_samples_leaf = 1; max_features = "sqrt"; max_depth = None; bootstrap = False.` Performance of regressor was then assessed on the test set using the R^2 -Score.

2.21.17 *Generation of average trans-sister classes*

First, observed/expected transformation (see 2.21.6) was performed on Hi-C matrices of trans-sister contacts of the merged G2 WT sample (see Table 1) binned at 20 kb. Then, all pixels separated by more than 200 bins were blanked, resulting in a Hi-C band-matrix. Then, the matrix was smoothed using a gaussian kernel ($\sigma = 2$ bins) and foreground elements detected using the adaptive threshold algorithm implemented in `skimage.filters.threshold_local` using the following parameters: `blocksize = 121; offset = -0.3`. Then, connected components were called as regions and thresholded at a minimum size of 41 bins². Regions of interest (ROI) were then defined to start at the upper right corner of a segmented region and extended down to the main diagonal. To avoid artefacts due to nested regions, all smaller, overlapping regions of a given ROI were filled with the mean value of the overall matrix. Then, the resulting ROIs were scaled such that they span 100 bins using `OpenCV` and flattened into a feature vector. K-means clustering was then performed on these feature vectors using `sklearn.cluster.KMeans` using standard parameters except `n = 15` clusters. Representations of each cluster were calculated by taking the pixel-wise average of all ROIs within a given cluster.

2.21.18 *ORI pileups*

Labeled reads (2 or more AG or TC mutations) from sequencing samples from cells released for different times from a G1-arrest into S-Phase were called using a custom python script. Then, bigwig files of labeled reads were constructed and pileups at ORIs [198] were generated using `deeptools` (<https://github.com/deeptools/deepTools>). The resulting profiles were plotted using a custom python script.

2.21.19 *Saddle plot analysis*

Compartment analysis was done using `cooltools` (<https://github.com/mirnylab/cooltools>). Briefly, the first eigenvector was calculated on HiC matrices

binned at 200 kb and phased using fractional gene coverage with bins containing a higher number of genes being assigned a positive sign ("A compartments") and bins containing a low number of genes a negative sign ("B compartments"). Then, eigenvector values were binned into 50 bins, discarding 2.5% of lowest and highest values. Then, each off-diagonal HiC bin was assigned to two eigenvector bins and the corresponding observed over expected value extracted. Finally, the average observed over expected value for each combination of eigenvector bins was displayed as a heatmap. The compartment score $Score_{comp}$ is then calculated using the following formula:

$$Score_{comp} = \log_2\left(\frac{A-A + B-B}{A-B + B-A}\right)$$

with $A-A$, $B-B$, $A-B$ and $B-A$ referring to the average observed/expected value in the top 5 % interactions (based on the value of the first eigenvector) between the indicated compartment types.

3 Results

3.1 Concept of a sister chromatid sensitive Hi-C approach

The study of sister chromatid biology has been severely held back by shortcomings of methodology. Specifically, we lack the appropriate tools to measure sister chromatid proximity and alignment on a genome-wide level at sufficient resolution. The underlying reason for this is that sister chromatids are - by most measures - chemically identical and it is thus extremely difficult to apply established genomics and molecular biological techniques since most of them were not designed with this challenge in mind. The crux of the problem is that sister chromatids have identical DNA sequence and due to the rise of sequencing technologies in the past decade, most techniques to look at chromatin structure rely on a sequencing based readout [21]. Moreover, the remaining techniques that can be used to look at chromatin structure are based on hybridization of DNA or RNA probes, which in turn also relies on distinct DNA sequences and therefore cannot distinguish sister chromatids [37, 127]. In order to gain insight into sister chromatid specific DNA structure, it is therefore necessary to artificially make sister chromatids distinct.

A way to achieve this goal is to exploit the semi-conservative nature of DNA-replication [199] to label sister chromatids differentially with artificial nucleotide analogs. The idea is to synchronize cells to G1 and then release them into S-phase in the presence of such a nucleotide analog. Upon completion of S-Phase, the cells will then harbor sister chromatids that each contain one strand that is labelled through the incorporation of the nucleotide analog. Prior work has adopted this idea to achieve differential labelling of sister chromatids with different nucleotide analogs [158]. This was done by first labelling cells with F-ara-EdU for several cell-cycles to ensure homogeneous labelling with this nucleotide. Cells were then switched to medium with BrdU for one S-Phase, resulting in cells that carry sister chromatids that contain one strand labelled with BrdU and one with F-ara-EdU. Upon completion of another S-Phase in the absence of any nucleotide analog, cells then harbored sister chromatids where one contained the BrdU label and the other the F-ara-EdU label (Fig. 3.1a). Since these nucleotides can be differentially detected using fluorescent probes, this synchronization scheme enables the distinction of sister chromatids in microscopy experiments. While this technique enabled significant insight into the global behavior of sister chromatids upon entry into mitosis [158], this approach is limited by the resolution of light microscopy and does not allow discrimination of different genomic loci. It can therefore not be used to gain a genome-wide view of sister chromatid specific chromatin structure.

In order to solve this problem, it would be necessary to enhance state-of-the-art chromatin structure probing methods [21–23] with the ability to distinguish sister chromatids. Since these techniques all rely on sequencing, this would require finding a way to differentially detect sister chromatids in such an experiment. A potential way to do that would be to adapt the synchronization scheme developed by Nagasake et al. [158] using a nucleotide analog

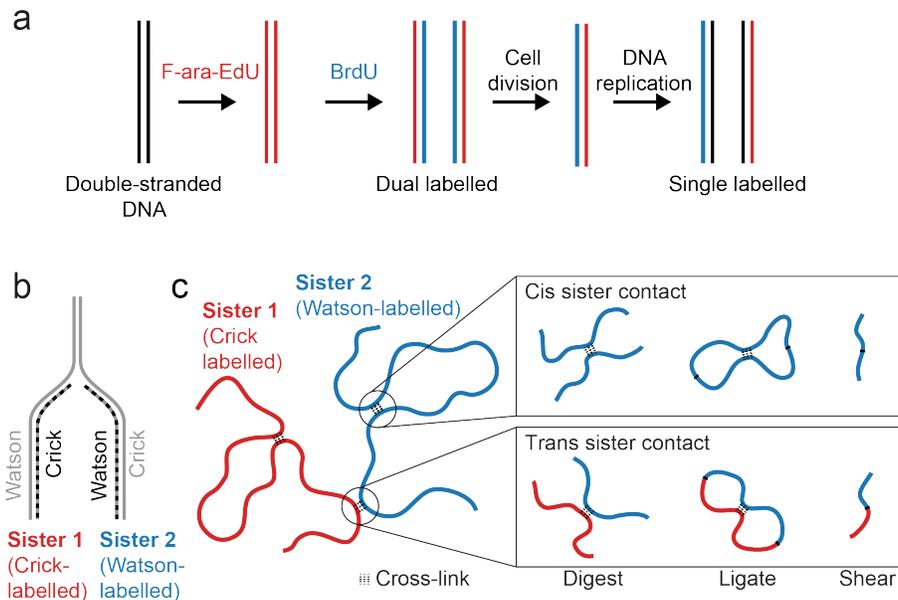


Figure 3.1: **Approaches for differential labelling of sister chromatids.**¹(a) Schematic of prior approach to label sister chromatids differentially with different nucleotides. Adapted from [158] (b) Schematic depiction of DNA replication in cells that were synchronized to the G1/S cell-cycle stage and released into S-Phase in the presence of a synthetic DNA analog. (c) Strategy to distinguish sister chromatids in a Hi-C experiment in cells that have been labeled as in (b).

that can be detected using sequencing. In such an approach, cells would be synchronized to the G1/S-boundary and released into S-Phase in the presence of a sequencing-detectable nucleotide analog. Upon completion of S-Phase, each sister chromatid would contain one strand of labelled DNA, but the key difference is that one sister chromatid would carry the label on the Watson strand and the other on the Crick strand (Fig. 3.1b). Since DNA sequencing readily gives information about the strand identity of the sequenced DNA fragment, this would allow the distinction of reads that come from one sister chromatid from reads that come from the other. When combined with chromatin conformation capture techniques, such an approach would therefore give information of whether a given 3D-contact happened within a sister chromatid (cis-sister contact) or between sister chromatids (trans-sister contact) (Fig. 3.1c). To make this idea reality, a nucleotide analog that is detectable using DNA sequencing is needed.

The idea of introducing artificial nucleotides into cells in order to obtain a

¹Panel (b) and (c) are adapted from Fig. 1a and Fig. 1b from [182]

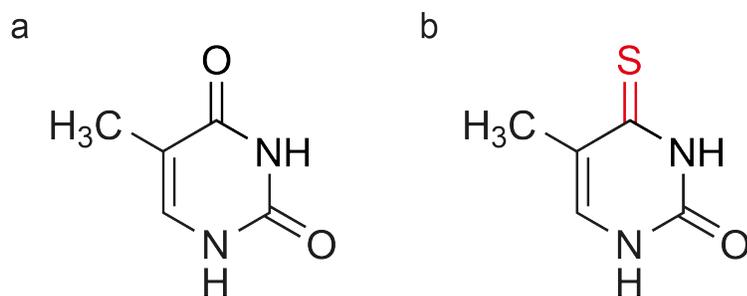


Figure 3.2: **Thymine and 4-thio-thymine.**²(a) Chemical structure of thymine. (b) Chemical structure of 4-thio-thymine. Differences to thymine are highlighted in red.

read-out via sequencing is not new. In fact, this has been done for RNA analogs very successfully in order to distinguish newly synthesized RNA from old RNA [190, 200]. Approaches that have done this employed 4-thiouracil, an RNA analog of uracil, for a short time-period to label nascent RNA, followed by a chase of uracil to purge the labelled nucleotide from cells. 4-thiouracil is not toxic and is readily taken up by a wide variety of cell types [190, 201]. Moreover, it was shown by a multitude of different measures that 4-thiouracil is treated equivalently to uracil by most cellular machineries including RNA polymerases [190]. In order to obtain a read-out, 4-thiouracil is either alkylated using iodoacetamide, or oxidized to cytosine using OsO_4 . Both conversion methods have been employed successfully to gain important biological insights [190, 200]. I therefore hypothesized that 4-thiothymidine (4sT) (see Fig. 3.2a, b) - the DNA analog of 4-thiouracil - could be a good candidate for a nucleotide that can be detected using DNA sequencing.

3.2 4sT is a novel nucleotide label that allows sequencing based detection

4-thio-thymidine (4sT) was chosen as a potential DNA nucleotide label that could be compatible with sequencing based detection. Not only because it is very similar to 4-thio-uracil, which has been employed successfully for RNA (see 3.1 for details), but also because early work showed that 4sT can be incorporated into genomic DNA of cultured human cells (although these experiments only allow a qualitative assessment of incorporation density) [202, 203]. Toxicity is a concern for sulfur containing nucleotides since a number of these compounds induce DNA damage and are used as cancer therapeutics due to their anti-proliferative properties [204]. Also here, prior studies suggest that 4sT does not exhibit pronounced toxicity - at least in some cell types - on its own, but

²Panel (a) and (b) are adapted from Fig. 1c from [182]

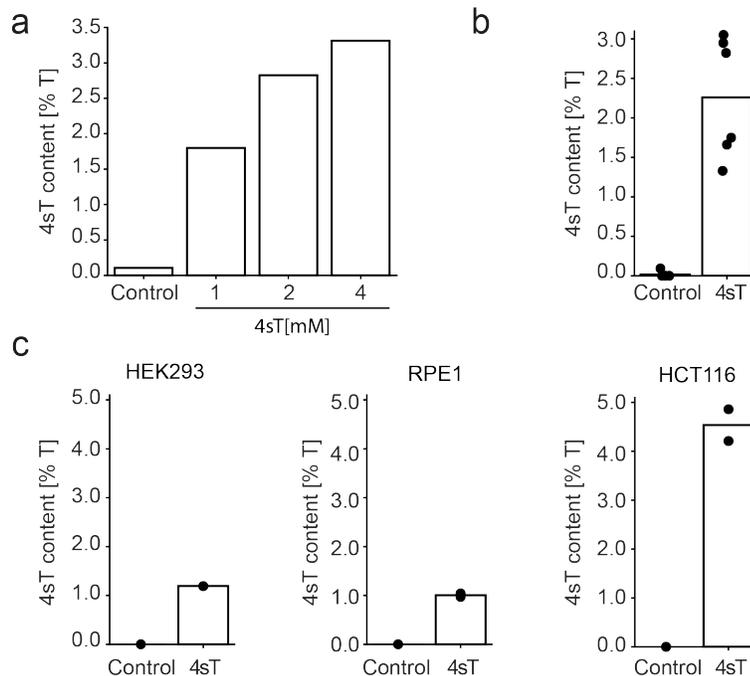


Figure 3.3: **Incorporation measurements of 4sT.** (a) 4sT incorporation into genomic DNA of HeLa cells after 5 days of growth in the indicated concentration of 4sT. Bars represent mean of $n=1$ biological replicates. (b) 4sT incorporation into genomic DNA of HeLa cells after 5 days of growth in 2 mM 4sT. Bars represent mean of $n=6$ biological replicates. (c) 4sT incorporation measurement of the indicated cell types grown for 5 days in the 2 mM 4sT. Bars represent mean of $n=2$ biological replicates.

only when combined with UV-treatment [202]. These considerations suggested that it might make sense to thoroughly test and characterize the properties of 4sT when employed as a DNA label. Specifically, I wanted to see whether 4sT can be incorporated in sufficient amounts into genomic DNA and whether 4sT perturbs cellular function.

3.2.1 4sT incorporation rate

First, I wanted to determine whether 4sT can be incorporated into DNA at a sufficient amount to allow distinction between labelled and unlabelled reads. Based on theoretical considerations, I hypothesized that an incorporation density of 2-5 % (4sT per thymidine) would be enough to obtain a usable amount of reads that can be assigned based on a read-length of 250 bp, which is the longest read mode of Illumina's newest NovaSeq platform. To test experimentally whether

such an incorporation density can be achieved, I wanted to establish conditions where almost all cellular DNA has undergone replication in the presence of 4sT. Since unlabelled DNA can only be diluted, but not eliminated from cells, 5 days of growth in 4sT was used as a compromise between label completeness and experimental time.

The first cell line that I analyzed were HeLa Kyoto cells, since prior reports suggested successful 4sT incorporation [202] and since the vast majority of modified cell lines in our lab are derived from HeLa Kyoto. Mass spectrometry was chosen as a readout since it allows quick and sensitive detection of small, intracellular metabolites. Specifically, cells were grown for 5 days in the presence of differing concentrations of 4sT, genomic DNA was extracted and digested down to nucleotides using a fungal enzyme to enable mass spectrometric quantification. This experiment showed that already at 1 mM, almost 2 % of thymidine was replaced by 4sT. This incorporation density increased with increasing 4sT concentration, but did not scale linearly with concentration increase above 2 mM (Fig. 3.3a). I therefore chose 2 mM as a concentration to further characterize in replicate experiment, where the incorporation density could be corroborated (Fig. 3.3b). To see whether 4sT can also be incorporated into the DNA of other cell lines, a panel of three different human cell lines including two non-cancer cell lines was probed with regard to their 4sT incorporation density. This showed that all tested cell lines incorporate 4sT, although the specific amount differed: RPE1 cells and HEK293 cells incorporated roughly half the amount of HeLa cells, whereas HCT116 cells exhibit a higher incorporation density (Fig. 3.3c). Taken together, this suggests that 4sT can be incorporated into the genomic DNA of a wide variety of cell lines at a density that is high enough to allow distinction between labelled and unlabelled DNA. While visual inspection suggested that 4sT does not perturb cellular function, I wanted to test this in more detail.

3.2.2 4sT toxicity

In order to investigate the toxicity of 4sT, I wanted to determine whether incorporation leads to increased cell death. To do this, we used live cell microscopy in combination with TOPRO-3, a non-cell-permeable dye that stains dead cells [205]. Specifically, cells were incubated for 24 h with differing concentrations of 4sT as well as etoposide, a topoisomerase II inhibitor that served as a positive control for dead cells [206]. 24 h was chosen as an incubation period since this is the maximum amount of time that cells would be exposed to 4sT in the proposed synchronization scheme (see 3.1). This experiment showed that 4sT did not affect the percentage of dead cells up until a concentration of 6 mM (Fig. 3.4a). These results suggest that 4sT is non-toxic at concentrations that allow sufficient amount of incorporation for sequencing based detection.

Next, I wanted to see whether the presence of 4sT influences cell proliferation. This experiment was done in parallel with the experiment to investigate the amount of dead cells using live cell imaging. This analysis showed that proliferation is decreased in the presence of 4sT already at 1 mM. This effect

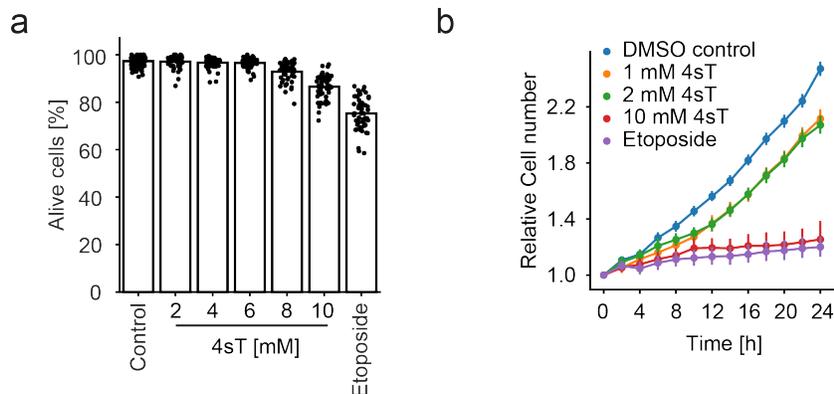


Figure 3.4: **Toxicity measurement of 4sT.**³(a) Percentage of live cells determined via Topro-3-Iodide staining of dead cells after 24 h incubation with the indicated compounds. Bars indicate mean of live cell percentages of individual wells from $n = 2$ biologically independent experiments. (b) Growth curve of HeLa Kyoto cells over 24 h in a medium containing the indicated amounts of 4sT or $50 \mu\text{M}$ etoposide. Points indicate mean and error indicates standard error from individual wells of $n = 2$ biologically independent experiments. These experiments were performed by Claudia Blaukopf.

increased with the amount of 4sT in the medium and cells did not replicate anymore at 10 mM (Fig. 3.4b). In combination with the lack of dead cells in the presence of 4sT shown before, this decrease in proliferation rate might potentially be due to slower DNA replication, likely via a similar mechanism to how thymidine inhibits DNA replication [207]. However, the decrease in proliferation is mild for 1 mM and 2mM - concentrations that already allow high incorporation densities - and can be compensated in the context of cell synchronization by adjusting the relevant timings.

Since sulfur containing nucleotides have been used in the context of cancer therapy due to their DNA damage induction [204], I wondered whether 4sT incorporation might lead to DNA damage. I thereof incubated HeLa Kyoto cells with different concentrations of 4sT and performed immunofluorescence detection of phospho- γ -H2A.X, a marker for DNA damage (Fig. 3.5a, b; [208]). This suggested that 4sT does not upregulate DNA damage to a detectable degree compared with the positive control etoposide, which led to a massive increase of the phospho- γ -H2A.X signal.

To sum up, 4sT can be incorporated into the genomic DNA of a wide variety of cell lines at sufficient amounts to allow confident detection. Furthermore, this incorporation can be obtained without impacting cell death or DNA damage pathways. This suggested that 4sT was a good candidate for optimizing

³Panel (a) caption is adapted from ED Fig. 1c from [182]

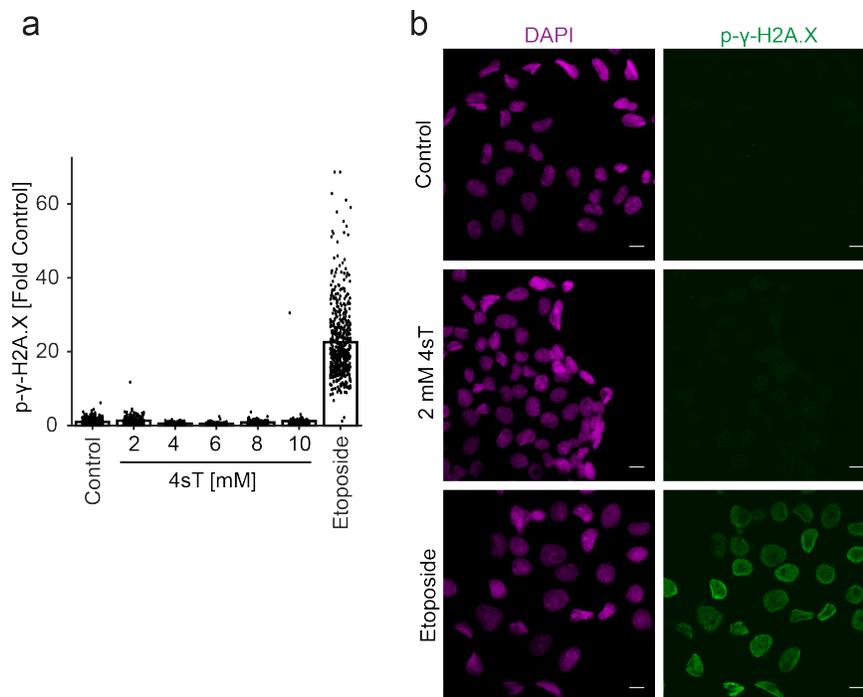


Figure 3.5: **DNA damage response to 4sT.**⁴(a) Quantification of mean fluorescence in cell nuclei stained by anti-p- γ -H2A.X antibody shown in (b). Bars indicate mean of individual cells from $n = 2$ biologically independent experiments.(b) DNA damage assay performed after 24 h incubation with the indicated compounds. Scale bars indicate 5 μ m.

detection using DNA sequencing.

3.2.3 Iodoacetamide mediated conversion of 4sT

While 4sT can be incorporated into genomic DNA at high density and low toxicity, many nucleotides exhibit these characteristics. What makes 4sT special is the introduction of a useful functional group: 4sT has a thioketo group that shows similar hydrogen bonding characteristics as the keto group naturally found in thymidine, but allows differential modification due to properties of the containing sulfur atom [209]. The general idea to exploit this fact to achieve sequencing based detection was to modify 4sT in such a way that its base-pairing characteristics are changed to be similar to cytosine. This would then lead to a point mutation when sequencing the modified material that marks the position

⁴Panel (b) is taken from ED Fig. 1d of [182]. Panel(a) and (b) caption are adapted from ED Fig. 1d of [182]

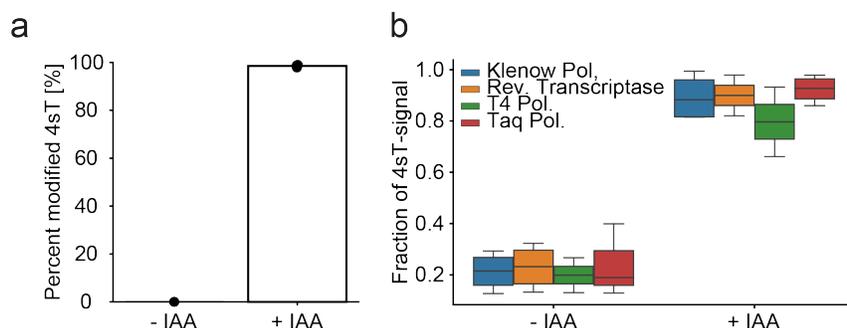


Figure 3.6: **Iodoacetamide mediated conversion of 4sT.** (a) Mass spectrometry analysis of alkylation of 4sT using iodoacetamide. Percentage of signal attributable to alkylated 4sT with and without iodoacetamide modification of genomic DNA containing 4sT. (b) Sanger sequencing based analysis of sequencing errors opposite of 4sT with and without iodoacetamide modification. An oligo containing 4sT was treated with iodoacetamide and with DMSO only and amplified with the indicated DNA polymerases, followed by Sanger sequencing. Values refer to the fraction of 4sT positions that show a point mutation from A to G.

of the artificial nucleotide.

The first approach to achieve this was inspired by SLAM-seq [190]. Specifically, the high nucleophilic potential of the sulfur atom of 4sT is used to attack iodoacetamide, which leads to alkylation of 4sT. This alkylation introduces an amine group, which mimics the base pairing characteristics of thymidine [190]. To determine whether this modification that was shown for 4-thio-uracil also works for 4sT, 4sT was treated with iodoacetamide and mass spectrometric detection of the reaction products was performed. This analysis suggested that 4sT can be alkylated completely using the same reaction conditions that work for 4-thio-uracil (Fig. 3.6a).

4-thio-uracil detection in SLAM-seq relies on the ability of reverse transcriptase to fit the bulky alkyl-adduct into its reaction center and read it as cytosine. However, for 4sT, this step has to be performed by a DNA polymerase, which exhibit much more stringent steric constraints, especially if they contain proof reading activity (reviewed in [210]). I therefore wanted to test whether DNA polymerases can extend over DNA that contains alkylated 4sT and whether it is read as cytosine. To this end, I employed a synthetic oligo that contains 4sT at known positions. I modified it using iodoacetamide, amplified it using different DNA polymerases and performed sanger sequencing. This analysis suggested that DNA polymerases can extend over the bulky adduct and indeed read alkylated 4sT mostly as cytosine (Fig. 3.6b).

The DNA polymerases that were employed for this test were all very per-

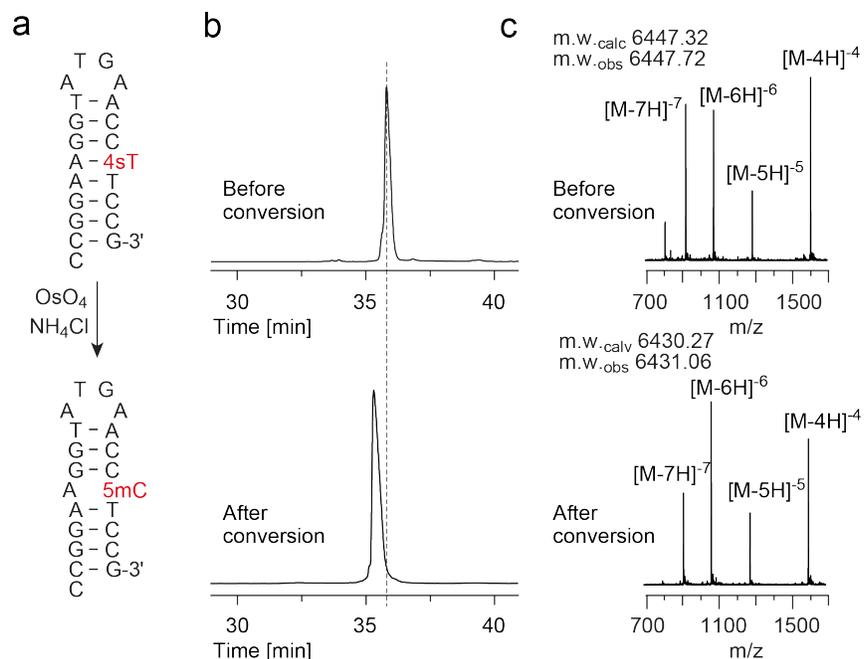


Figure 3.7: **OsO₄ mediated conversion of 4sT.**⁵(a) Synthetic hairpin-oligonucleotide used to probe 4sT conversion by OsO₄. The theorized reaction educts and products are highlighted in red. (b) High performance liquid chromatography (HPLC) trace at 260 nm of the oligos depicted in (a) before and after the conversion by OsO₄. The peak position of the oligo before conversion is indicated by a dashed line. These experiments were performed by the Micura group at the University of Innsbruck.

missive and lacked proof-reading activity to allow efficient amplification of DNA containing the bulky alkyl adduct. But for detection using next generation sequencing, a stringent proof-reading polymerase would be preferable since the much lower error rate of these polymerases would allow a higher signal-to-noise ratio when detecting a mutation based readout. I therefore tested whether it is possible to amplify a iodoacetamide modified oligo containing 4sT with proofreading polymerases. This showed that proofreading polymerases were not able to amplify the alkylated oligo, while Taq-polymerase was able to do so. This suggests that iodoacetamide based detection of 4sT is possible, but will likely suffer from a low signal-to-noise ratio.

⁵Panel (a), (b) and (c) and their captions are adapted from Fig. 1d, e and ED Fig. 1c of [182]

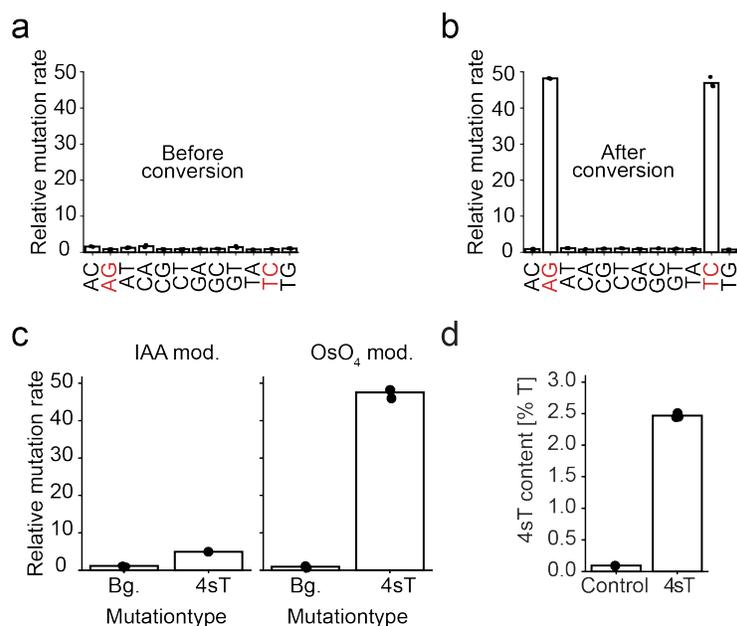


Figure 3.8: **Comparison of OsO₄ and iodoacetamide mediated conversion of 4sT.** ⁶(a) Point-mutation rates of genomic DNA from HeLa cells grown in medium containing 4sT relative to control DNA from cells grown in the absence of 4sT, before OsO₄-mediated conversion. Bar graphs indicate the mean of n = 3 biologically independent experiments. (b) Point-mutation rates of genomic DNA from HeLa cells grown in medium containing 4sT relative to control DNA from cells grown in the absence of 4sT, after OsO₄-mediated conversion. Bar graphs indicate the mean of n = 3 biologically independent experiments. (c) Comparison of relative signature mutation rate (A-to-G and T-to-C; 4sT mutationtype) upon iodoacetamide (IAA) and OsO₄ mediated conversion of 4sT containing genomic DNA. "Bg." refers to background mutations (all mutations that are not A-to-G or T-to-C). Bars indicate the mean of n = 3 biologically independent experiments. (d) Quantification of 4sT incorporation using DNA sequencing. Cells were 4sT labelled and purified genomic DNA chemically converted. Indicated values are the sum of the A-to-G-mutation rate and the T-to-C mutation rate, normalized to the total amount of adenosine and thymidine measured respectively. Bars indicate mean of n = 3 biologically independent experiments.

3.2.4 OsO₄-conversion allows quantitative detection of 4sT

Since iodoacetamide based conversion of 4sT likely suffers from low signal-to-noise ratio (see 3.2.3), I looked for other possibilities to make 4sT detectable using sequencing. A promising approach was TUC-seq: In this methodology, 4sU is converted directly to cytosine using OsO₄ based oxidation [200]. I reasoned that the lack of a bulky alkyl adduct and the direct conversion to a natural nucleotide could allow the use of proof-reading polymerases in 4sT detection. In a collaboration with the lab of Ronald Micura from the University of Innsbruck, we were able to show that OsO₄ based conversion of 4sT to 5-methyl-cytosine works for small oligos containing 4sT (Fig. 3.7a-c).

To test whether this detection is also possible for genomic DNA extracted from cultured cells, I grew WT HeLa cells for 5 days in the presence of 2 mM 4sT, extracted DNA and modified it using OsO₄. When I compared the point mutation rates with cells that were not treated with 4sT, I saw an almost 50-fold increase of the T-to-C and A-to-G mutation, while the other point mutation types were not affected by the treatment (Fig. 3.8a, b). Interestingly, when I sequenced the same samples without OsO₄ treatment, the point mutation rates were not different from control cells, suggesting that 4sT does not introduce mutations inside cells (Fig. 3.8a). When I compared the relative point mutation rates with samples that were obtained similarly, but treated with iodoacetamide, I saw that the OsO₄ sample showed a vastly increased signal-to-noise ratio (Fig. 3.8c). In order to assess the ratio of incorporated 4sT to detected 4sT, I compared the mass spectrometry based incorporation tests to the sequencing results (Fig. 3.3b and Fig. 3.8d). This comparison showed an almost identical incorporation rate when measured with mass spectrometry or with sequencing. These results suggest that OsO₄ mediated conversion allows quantitative detection of 4sT in genomic DNA.

To further corroborate the ability to detect 4sT in genomic DNA and ascertain that the signal that I see is truly due to incorporation of the synthetic nucleotide, I wanted to correlate 4sT incorporation with replication timing. To this end, I synchronized HeLa Kyoto cells to the G1/S boundary, released them into S-Phase in the presence of 4sT and collected samples at different time points until the cells reached the G2 cell cycle stage. I hypothesized that if the detected mutational signature was due to 4sT incorporation, the fraction of "labelled" reads should increase with S-Phase progression. To this end, I defined a read as labelled if it contained more than one "signature" mutation (see 3.4.2 for more details). When I calculated the fraction of labelled reads, it was indeed evident that this fraction started at almost zero and increased to 37 % in cells synchronized to the G2 cell cycle stage (Fig. 3.9a). This not only showed that the mutational signature increases with S-Phase progression, but that I am able to assign the state of labelling of a large fraction of all reads. To further corroborate the correlation of mutational signature with replication timing, I performed read density pileups of labelled reads at origins of replic-

⁶Panel (a) and Panel (d) caption are adapted from Fig. 1f and ED Fig. 1b of [182]

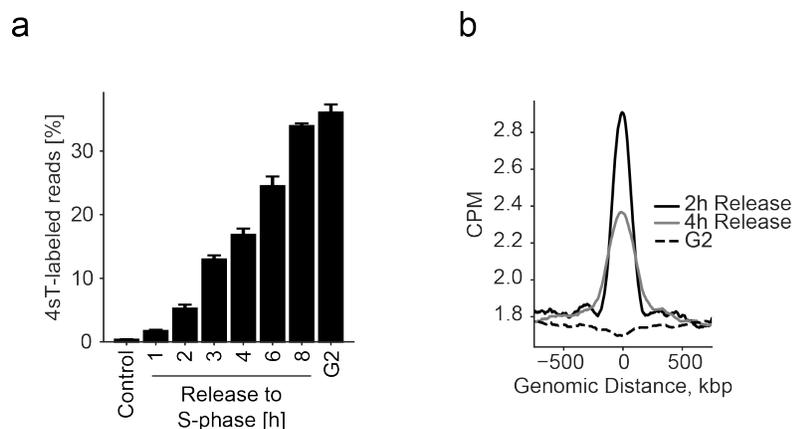


Figure 3.9: **Genomics based validation of 4sT incorporation.** (a) Percentage of labelled reads based on at least two signature point mutations from cells released for the indicated time into S-phase in the presence of 2 mM 4sT. Control refers to cells that were not released and G2 refers to cells that were arrested using RO3306 after labelling. (b) Pileup of labelled reads at origins of replications of cells released for the indicated time into S-Phase in the presence of 4sT or were arrested in G2 using RO3306. Shown are counts per million (CPM).

ation [198]. The expectation was that in early S-Phase, labelled reads should be clustered around origins of replication, whereas in G2, the labelling profile should be smooth along these sites. And indeed, this is what I observed (Fig. 3.9b), suggesting not only that it is possible to confidently detect 4sT using this approach, but also that homogeneous labelling is obtained in a G2 synchronized population.

3.2.5 Accounting for HeLa SNPs increases signal-to-noise ratio of 4sT detection

HeLa Kyoto cells are derived from a human tumor and therefore exhibit a high degree of genomic aberrations, including duplications, deletions and point mutations [211]. Since the exact genome of each HeLa clone changes with time as mutations accumulate, I reasoned that a sizeable fraction of reads assigned to originate from a 4sT labelled population based on their mutational signature might be false-positives. To account for this, I sequenced the genome of HeLa cells without 4sT incorporation and called SNPs using bcftools. Following this, I constructed a new version of the *hg19* genome assembly, with all HeLa specific SNPs blanked out. The genome was constructed in such a way that all duplications or deletions were not accounted for. This was done since these aberrations are unlikely to affect 4sT detection since they do not affect point mutation rates and - more importantly - allows the use of all available annotations of

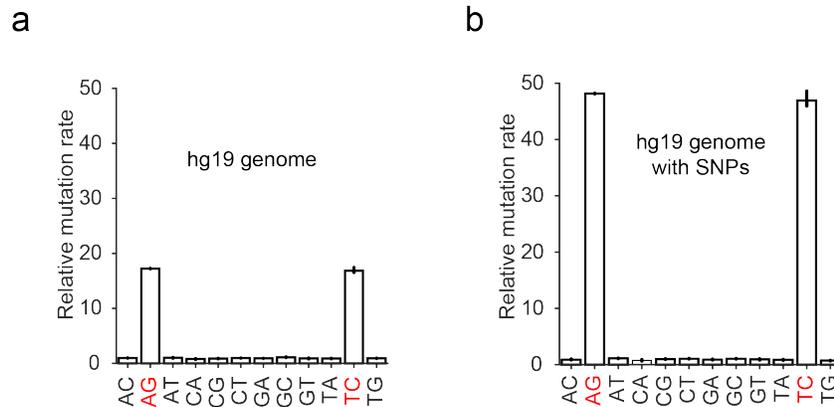


Figure 3.10: **Impact of HeLa SNPs on 4sT detection.** (a) Relative mutation rate of 4sT treated samples to control samples aligned to the *hg19* genome. Bars indicated mean and error bars indicate standard error of $n = 3$ biologically independent replicates. (b) Same samples as in (a) but aligned to an *hg19* variant where HeLa SNPs are ignored. Bars indicated mean and error bars indicate standard error of $n = 3$ biologically independent replicates.

hg19. When I compared the relative point mutation rates of fully labelled 4sT samples (see 3.2.4 for details) aligned to the standard *hg19* genome and the SNP-corrected genome, I saw an almost 3-fold increase of signal-to-noise ratio (Fig. 3.10a,b). Therefore, all data sets shown in this thesis were aligned to the SNP-corrected genome.

3.3 A synchronization scheme for sister chromatid sensitive Hi-C

After having developed an incorporation concept that would allow sister chromatid distinction in a Hi-C experiment (see 3.1), I established that 4sT fulfills all the requirements to be used as an artificial nucleotide in this synchronization scheme. I now had all the prerequisites to put the theory to the test and develop a specific synchronization scheme that brings these two components together. For this, I needed a way to reliably synchronize cells to the G1/S boundary, let cells progress through S-Phase in the presence of 4sT and then arrest them at specific cell cycle stages of interest to perform experiments.

3.3.1 Thymidine/aphidicoline block for G1/S synchronization

I first tackled the problem of synchronizing cells to the G1/S-boundary. Here, a natural first idea was to use a doubly thymidine block [212]. This approach exploits the fact that high concentrations of thymidine in the growth medium inhibit the synthesis of DNA [207]. The strategy to synchronize cells using this

approach starts by treating asynchronous cells with thymidine for 16 h. This will stall all cells that are in S-Phase and allow enough time for cells that are in other cell cycle stages to reach the beginning of S-Phase. Cells are then released from the thymidine block for 8 h, so that both cells that were stalled in S-Phase and arrested at its beginning have enough time to complete it. Then, thymidine is added again for 16 h to finally collect all cells at the beginning of the G1/S boundary. While HeLa cells readily respond to synchronization by double thymidine block, it is most likely not compatible with 4sT incorporation from the very start of S-Phase: Thymidine and 4sT will compete for the same enzymes to build up an intracellular pool of usable nucleotide and thus dilute the active 4sT concentration available during early S-Phase.

In order to solve this problem, I hypothesized that aphidicoline could be used to replace the second thymidine block. Aphidicoline is a fungal toxin that inhibits DNA polymerase α and has thus been used successfully in the past to synchronize cells to the G1/S-boundary [213]. Aphidicoline could not only allow to overcome competition between thymidine and 4sT, but will also allow tighter synchronization, since it inhibits S-Phase at an earlier stage than thymidine [207, 213]. Having cells synchronized to the G1/S boundary will allow to derive synchronous populations of cells that have replicated their genome in the presence of 4sT at other cell cycle stages.

3.3.2 4sT can be used in the proposed synchronization scheme

I showed earlier (see 3.2.2) that 4sT does not exhibit toxicity at concentrations that allow sufficient incorporation and that it does not induce DNA damage. However, 4sT did reduce the proliferation rate of HeLa Kyoto cells. I hypothesized that this could be due to a prolonged S-phase, which could impact timing optimizations of the proposed synchronization scheme and I therefor wanted to determine to what extent 4sT elicits such a delay in S-Phase.

To this end, I synchronized HeLa cells to the G1/S boundary using the mentioned thymidine/aphidicoline strategy (see 3.3.1) and released them into S-Phase in the presence of different concentrations (2 mM, 4 mM and 6 mM) of 4sT. To determine the timing of DNA replication, I collected samples at different time points after the release and performed FACS analysis of DNA content. This suggested that cells progress through S-Phase in the presence of 4sT, but slightly slower than control cells (Fig. 3.11). Moreover, samples collected with cells synchronized to the next G1 phase by addition of thymidine 8 h after release showed that all cells pass through mitosis and thus are able to successfully complete the cell cycle in the presence of 4sT. While 2 mM 4sT in the growth medium barely slowed progression through S-phase, 4 mM and 6 mM slowed the progression severely and at 6 mM 4sT, a population of cells did not enter S-Phase at all. This suggests that 4sT can indeed act as retardant of S-Phase progression, likely through a similar mechanism as thymidine [207]. For this reason, all subsequent experiments shown here were done with 2 mM 4sT as it balances impact on cellular function and incorporation density.

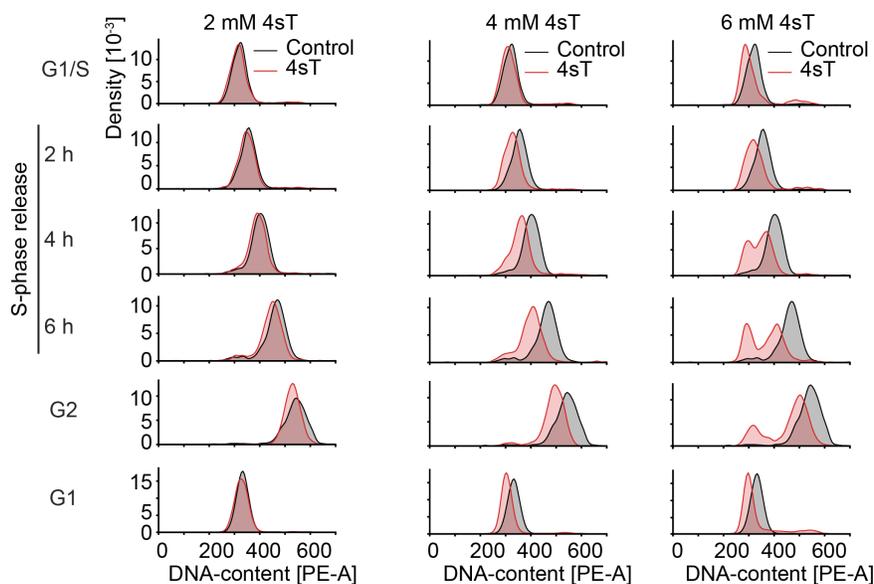


Figure 3.11: **S-phase progression in the presence of 4sT.**⁷Flow cytometry (FACS) analysis of cells progressing through S-phase in the presence or absence of 2 mM, 4 mM and 6 mM 4sT. DNA was stained using propidium iodide and kernel density estimation of signal in the PE-channel is shown. Cells were pre-synchronized to G1/S by thymidine and aphidicolin and released into S-phase by removal of thymidine. The G2 sample was arrested by RO3306 and the G1 sample was arrested after progression through mitosis using aphidicolin.

3.3.3 Optimization of synchronization schemes for different cell cycle stages

In order to synchronize HeLa cells to the G2 cell cycle stage while labelling DNA with 4sT, cells were first synchronized to the G1/S cell cycle stage using the thymidine/aphidicolin strategy (see 3.3.1). Then, cells were released in the presence of 4sT and 4 h after release the CDK1 inhibitor RO3306 was added. RO3306 was not added immediately at the time of release to both limit the time of exposure of cells to the drug and to avoid interfering with CDK1 function during origin firing [214]. After completion of S-Phase, cells were then incubated further with RO3306 for 16 h to ensure complete replication also of late replicating regions. When I measured both DNA content and fraction of mitotic cells using FACS, I noticed that a sizeable number of cells already started entering mitosis under these conditions. In subsequent optimizations, I found that this is due to the 20 min time window between harvesting cells and fixation, during which RO3306 is not present. When I added RO3306 to all

⁷Figure caption is adapted from ED Fig. 1e of [182]

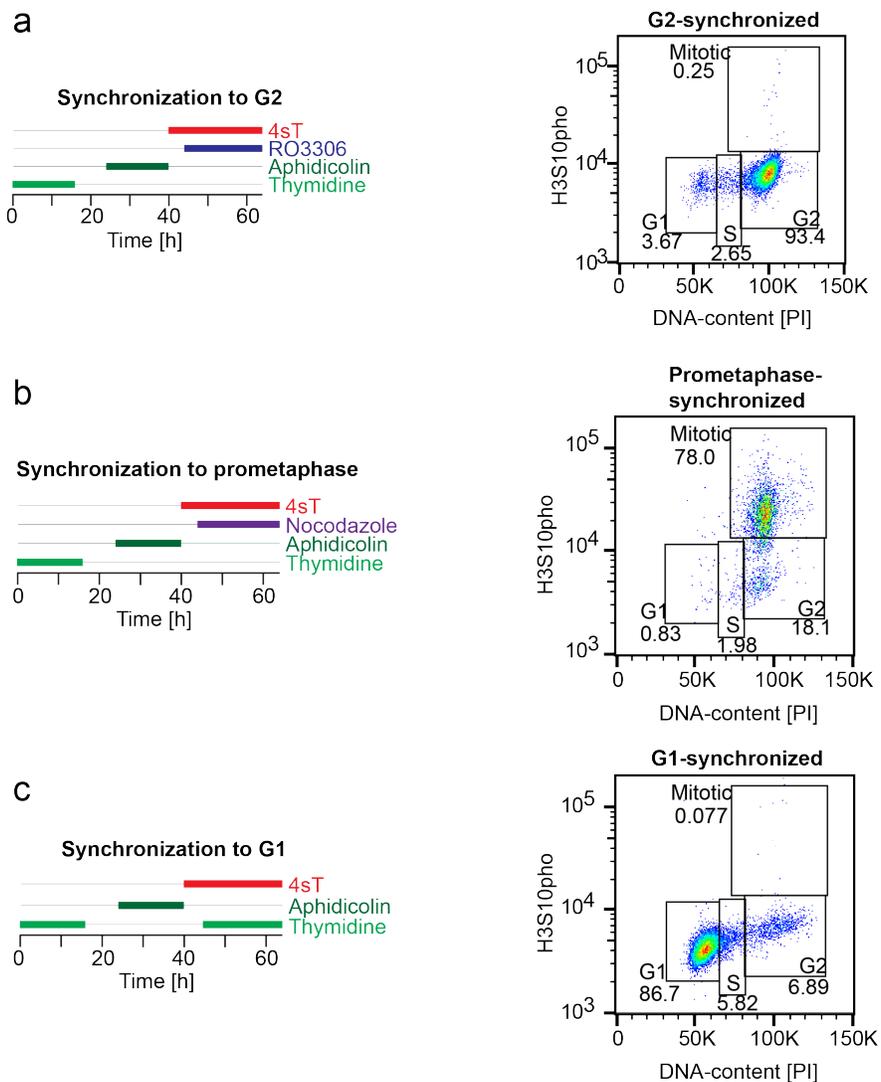


Figure 3.12: **Synchronization for scsHi-C.**⁸(a) Procedure of cell synchronization for scsHi-C sample preparation of cells synchronized to G2 as well as FACS analysis of HeLa Kyoto cells that underwent this procedure. (b) Procedure of cell synchronization for scsHi-C sample preparation of cells synchronized to prometaphase as well as FACS analysis of HeLa Kyoto cells that underwent this procedure. (c) Procedure of cell synchronization for scsHi-C sample preparation of cells synchronized to G1 as well as FACS analysis of HeLa Kyoto cells that underwent this procedure.

buffers involved in these steps, I was able to obtain a population that contained more than 90 % cells in the G2 cell cycle stage (Fig. 3.12a).

In order to synchronize cells to mitosis, cells were presynchronized to the G1/S boundary as described above, released into S-Phase in the presence of 4sT and treated with nocodazole 4 h after S-Phase release. Cells were then collected similarly to the G2 sample on the next day (Fig. 3.12b). For synchronization to the subsequent G1 cell cycle stage, cells were presynchronized to the G1/S boundary, let complete S-Phase in the presence of 4sT and treated again with thymidine 8 h after release. These cells were then harvested 16 h after this thymidine block (Fig. 3.12c).

3.4 Quality control of scsHi-C

After having established a synchronization protocol for HeLa cells that is compatible with 4sT labeling, I generated sister chromatid resolved Hi-C (scsHi-C) data sets for cells synchronized to different cell cycle stages. Specifically, cells were synchronized to different S-phase stages, to the G2 cell cycle stage, to prometaphase and to the following G1 stage (see 3.3.3 for details). The obtained material was then processed using standard Hi-C procedures (adapted from [215]) and libraries were prepared using a commercial DNaseq kit. Before the final amplification and sequencing, the obtained libraries were treated with OsO₄ to convert 4sT to 5-methylcytosine (see 3.2.4) in order to convert 4sT to a detectable point mutation.

3.4.1 Feasibility of sister chromatid-sensitive Hi-C

Hi-C is a method that relies on a vast amount of sequencing reads in order to deduce detailed information about the architecture of the 3D-genome. So in order for scsHi-C to be of any practical value, the amount of contacts that can be assigned a sister chromatid identity needs to be substantial. Otherwise, experiments would be impractical both due to inadequate library complexity and sequencing cost.

Based on the proposed concept to distinguish cis-sister from trans-sister contacts (see 3.1), both contacts involving only the labelled strands of both sister chromatids and contacts involving only the unlabelled strands should be informative. Contacts involving a labelled and an unlabelled piece of DNA are not informative since they can both arise through trans-sister and cis-sister DNA ligation (Fig. 3.13a). These considerations are based on complete labelling, assuming that labelled reads can be distinguished from unlabelled reads with perfect accuracy. For this to be possible, the absence of signature point mutations would need to be as informative as their presence. While reads with more than one signature point mutation are very likely to derive from the labelled population (Fig. 3.13b), the absence of signature mutations also happens frequently for reads from a labelled sample. This is due to the low incorporation density of 4sT and precludes using Hi-C contacts between unlabelled DNA pieces for

⁸Panel (a), (b) and (c) are adapted from ED Fig. 2c and d of [182]

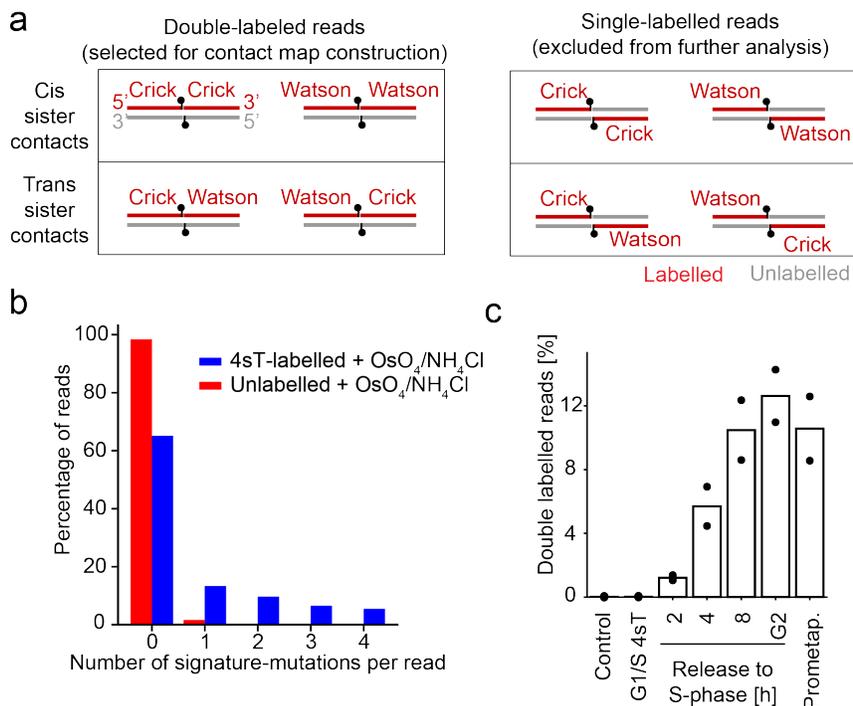


Figure 3.13: **Usable reads for scsHi-C.**⁹(a) Depiction of all possible Hi-C ligation products of a sample where one strand of each sister chromatid has been labelled with a synthetic nucleotide. (b) Histograms of signature point mutations per read (AG or TC) of conventional sequencing libraries constructed from cells that were grown for 5 days in the presence (“4sT-labelled”) or absence (“unlabelled”) of 4sT and treated with OsO₄. (c) Quantification of Hi-C reads that are labelled on both sides for sister-specific contact classification, as a percentage of all reads. Cells were synchronized to the G1/S boundary and released into S-Phase in the presence of 4sT for the indicated times. The G2 sample was arrested using RO3306; the prometaphase sample was arrested using nocodazole; the control sample refers to unlabelled DNA and the G1/S 4sT sample refers to a sample that was treated with 4sT, but not released into S-Phase.

scsHi-C since this would substantially increase wrong contact assignment. I can therefore only use Hi-C contacts formed between reads that contain above a threshold number of signature mutations, which I coin "double labelled" contacts.

Based on the histogram of signature mutations per read from conventional single-end read data of fully labelled samples (Fig. 3.13b), a threshold of at least two signature mutations was judged to be a good trade-off between false-positive rate and usable read amount. To determine the fraction of "double labelled" reads that I can obtain in a Hi-C experiment, I synchronized cells to the G1/S boundary, released them into S-Phase in the presence of 4sT and collected samples at different time points after S-phase release and generated scsHi-C data sets. This experiment showed that the fraction of "double labelled" reads increases as cells progress through S-phase and reaches almost 12 % in G2 synchronized cells (Fig. 3.13c). This suggests that under these conditions, 12 % of reads can be assigned a sister chromatid identity, which allows detailed investigation of sister chromatid specific DNA structure with moderate monetary investment.

3.4.2 Confidence of sister chromatid assignment

Since the usability of the obtained reads strongly relies on the signal-to-noise ratio of 4sT specific signal, I wanted to know what the false-positive rate of double labelled read assignment is. To this end, I took an scsHi-C sample of cells synchronized to the G1/S-boundary, before 4sT incorporation. This sample should contain no double labelled reads and thus all such reads are false-positives. I calculated false-positives at different signature mutation thresholds and found that the false-positive rate starts off high at 2.5 % for 1 mutation and goes down to below 0.1 % for 2 mutations, to drop even further for more stringent thresholds (Fig. 3.14a). This suggests that the chosen mutation threshold of at least 2 signature point mutations exhibits a very low false-positive rate of double labelled read assignment.

While a low false-positive rate of double labelled reads is a requirement for confident sister chromatid contact assignment, it does not assess the validity and error rate of the proposed classification scheme. I therefore looked for a method to assess the confidence of cis-sister and trans-sister contact assignment. To this end, I adapted an approach that was initially developed for assessing the accuracy of homolog chromosome contact assignment in *D. melanogaster* [195]. The idea is to exploit the fact that each Hi-C library contains artefacts that arise from chromatin that has not been cut by the applied restriction enzyme. These contacts are "very" short-range in nature as the distance between the two ends is the insert size of the library, which for the experiments here was between 500 and 800 bp. Since a chromatin piece that has not been cut and re-ligated during the Hi-C procedure cannot be a trans-sister contact, such artefacts should be classified as cis-sister contacts.

⁹Panel (a) and (b) and captions are taken from ED Fig. 3c and a of [182] respectively. Panel (c) caption is taken from Fig. 1h of the same paper.

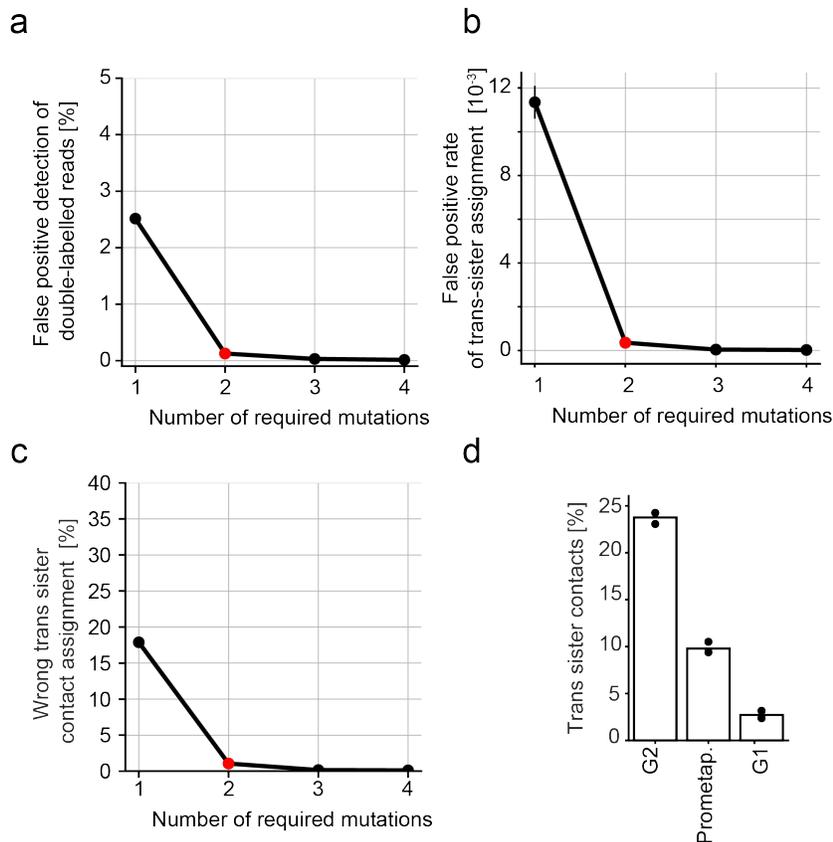


Figure 3.14: **False-positive rate of scsHi-C.**¹⁰(a) False-positive rate of double labelled read detection in a sample that does not contain 4sT at different mutation thresholds. (b) false-positive rate of trans-sister contact assignment based on short range contacts in a G2 synchronized sample at different mutation thresholds. (c) Fraction of wrong trans-sister contacts in a G2 sample at different mutation thresholds. (e) Percentage of trans-sister contacts based on all double-labelled reads that exhibit a genomic separation larger than 10 kb. Cells were released from G1/S block into medium containing 4sT and then arrested in G2 using RO3306, in prometaphase using nocodazole, or the following G1 using thymidine.

So assuming that all contacts below 1 kb are artefacts, I can derive a false-positive rate of trans-sister assignment by calculating the fraction of trans-sister contacts with separation smaller than 1 kb. To perform these measurements, I used scsHi-C samples from cells synchronized to G2. I reasoned that these

¹⁰Panel (a) and (c) and their captions are adapted from ED Fig. 3 (b) and (e) of [182] respectively. Panel (d) caption is adapted from Fig. 1i of the same paper.

samples should give the strongest trans-sister signal since the genome is entirely replicated, but chromosomes are not yet macroscopically resolved. I calculated the false-positive rate of trans-sister assignment for different signature mutation thresholds to get information about the expected signal-to-noise ratio at these cut-offs. The measured false-positive rate dropped off sharply from one to two minimally required mutations and decreases further when the number of required mutations was increased (Fig. 3.14b). However, the false-positive rate was already extremely low if at least two mutations were required, a threshold that yields a large fraction of usable reads.

Next, I wanted to estimate the fraction of wrongly assigned trans-sister contacts to further determine the optimal signature mutation threshold. In order to do this, I assumed that the false-positive rate of trans-sister assignment is independent of contact separation and that all cis-sister contacts are correctly assigned. I then calculated the fraction of wrong trans-sister contacts using the following formula:

$$False_{trans} = \frac{FPR_{trans} \cdot Contacts_{cis}}{Contacts_{doubly}}$$

with $False_{trans}$ referring to the fraction of wrongly assigned trans-sister contacts, FPR_{trans} being the measured false-positive rate of trans-sister contact assignment, $Contacts_{cis}$ being the amount of cis-sister contacts and $Contacts_{doubly}$ denoting the amount of double labelled contacts. For this measurement, only Hi-C contacts with a separation of more than 10 kb were used, since those are likely free of Hi-C artefacts [195]. This analysis showed a rather substantial amount of wrongly assigned trans-sister contacts with the signature mutation threshold set to one (Fig. 3.14c). However, the fraction decreases steeply as the threshold is increased to at least two signature mutations, staying below 2%. As the threshold is increased further, the confidence increases, but these increases are only marginal and don't justify the substantial amount of reads that are lost (Fig. 3.13b). I therefore decided that a threshold of at least 2 signature point mutations is the optimal balance between confidence and yield and will be used for all further experiments.

3.4.3 Biological controls for scsHi-C

To further corroborate the validity of the scsHi-C approach, I wanted to complement the above mentioned technical controls (see 3.4.2) with biological ones. In order to do this, I sought to replicate known features of sister chromatid organization using scsHi-C, specifically the separation of sister chromatids in mitosis and the segregation of sister chromatids to different daughter cells [67, 158]. So I first generated scsHi-C data sets for cells synchronized to prometaphase using nocodazole (see 3.3.3). Here, sister chromatids should be largely separated based on prior microscopy based experiments. And indeed, when I compared the fraction of trans-sister contacts in samples synchronized to G2 with samples in prometaphase, I saw a substantial drop of trans-sister contacts (Fig. 3.14d).

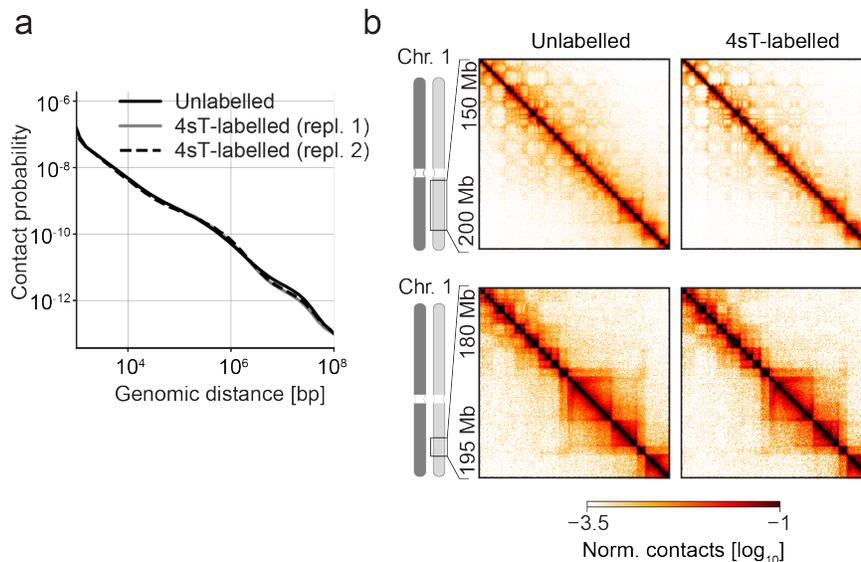


Figure 3.15: **Impact of 4sT on genome folding.**¹¹(a) Average contact probability over different genomic distances of HeLa cells synchronized to G2 that were either labelled with 4sT or unlabelled. (b) Hi-C interaction matrices at example regions of HeLa cells synchronized to G2 that were either labelled with 4sT or unlabelled.

I next wanted to test whether the segregation of labelled sister chromatids to different daughter cells changes scsHi-C signal accordingly. Specifically, I would expect that since the labelled sister chromatids now reside in different cells that there should be no trans-sister contacts remaining. And indeed, when I compared the fraction of trans-sister contacts in cells that have segregated their labelled chromosomes with cells synchronized to G2, I saw a substantial drop of trans-sister fraction from 23.4 % to 2.7 %, suggesting that most trans-sister contacts disappear when cells segregate their sister chromatids (Fig. 3.14d). The remaining contacts are likely due to synchronization problems since the purity of the population in the G1 samples is 89 % (Fig. 3.12c). Assuming the other 11 % of cells are in the G2 or S-phase, I would expect roughly 2 % of trans-sister contacts to remain. Taken together, this suggests that the overall behavior of trans-sister chromatids is in line with known sister chromatid biology.

In order to test whether the introduction of 4sT into genomic DNA alters the folding of the 3D-genome, I generated scsHi-C data sets for cells synchronized to G2, both with 4sT labelling and without. As a first metric, I generated scaling plots for both types of samples and compared the decay of contact probability

¹¹Panel (a) and (b) as well as their captions are adapted from ED Fig. 2e and f of [182] respectively.

with genomic distance. This analysis suggested that there is no substantial difference on the scaling behavior of data sets with and without 4sT (Fig. 3.15a). To further corroborate these findings, I generated heatmaps for an example region on chromosome 1 with a large binsize to assess the global similarity. In both data sets, compartments are present and the heatmaps look almost identical. Finally, I generated heatmaps that show a more zoomed-in view of chromosome 1 to assess the organization of TADs. Also here, there were no obvious differences between samples labelled with 4sT and without (Fig. 3.15b). To conclude, the introduction of 4sT does not alter 3D-genome folding based on the assays that were performed.

3.5 Large scale organization of sister chromatids

3.5.1 Large scale organization in the G2 cell cycle stage

After having established that scsHi-C can confidently distinguish between trans-sister and cis-sister contacts, I went ahead and generated deep Hi-C data sets of WT HeLa Kyoto cells synchronized to the G2 cell cycle stage (see 3.3.3). In order to gain insight into the detailed fine structure of sister chromatid topology, I generated 11 biological replicates with a total of $1.7 \cdot 10^9$ reads, $1.65 \cdot 10^8$ of which could be assigned a sister chromatid identity (see Table 1). To assess the reproducibility of these replicates, I employed HiCRep [194]. HiCRep uses a stratified correlation coefficient to solve severe problems associated with calculating bin-wise correlation coefficients of Hi-C matrices. The problem with ordinary correlation is that most of the correlation coefficient between two Hi-C samples is dominated by contacts close to the diagonal since the contact probability decays exponentially with increasing genomic separation. HiCRep computes a correlation coefficient per diagonal and takes the average of all used diagonals to avoid inflation of high frequency bins. Applying HiCRep to these replicates shows that they are highly reproducible, both for cis-sister contacts as well as for trans-sister contacts (Fig. 3.16a, b).

To assess the large scale organization of G2 sister chromatids, I generated Hi-C matrix heatmap representations of the long arm of chromosome 1, binned to 500 kb of all pooled replicates. When I looked at cis-sister contacts, they looked very much like all contacts and exhibited familiar patterns commonly seen in Hi-C data (Fig. 3.17a; [22]): A strongly pronounced main diagonal, indicative of self-interaction of genomic regions within 500 kb bins, as well as the established exponential decay of contact probability with increasing genomic separation. Moreover, the plaid-pattern indicative of compartmentalization is readily visible especially at regions closer to the telomere of the long arm of chromosome 1. When I looked at trans-sister contacts, I saw that also these contacts cluster around the main diagonal, suggesting overall alignment of sister chromatids. Upon closer inspection, however, I saw that trans-sister contacts were substantially reduced at the main diagonal as compared to cis-sister contacts. This suggests that sister chromatids are globally aligned at the G2 cell cycle

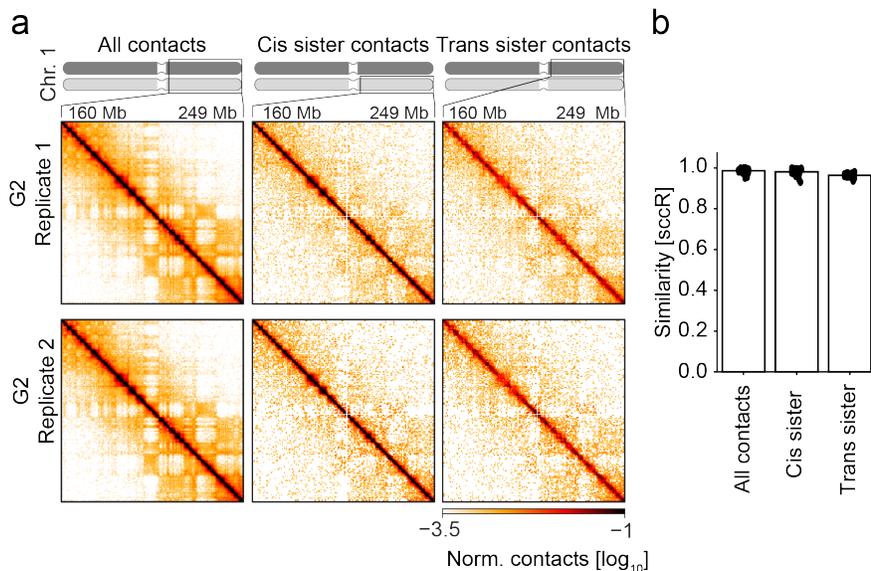


Figure 3.16: **Reproducibility of G2 scsHi-C.**¹²(a) Hi-C interaction matrices of the long arm of chromosome 1 of all contacts, cis-sister, and trans-sister contacts shown for two of the 11 G2 WT replicates. The all-contacts matrix was normalized to the total number of corrected contacts in the region of interest (ROI), whereas cis-sister and trans-sister contacts were normalized to the total amount of cis-sister contacts and trans-sister contacts in the ROI. Bin size of the matrix is 500 kb. (b) HiCrep analysis of all, cis-sister and trans-sister contacts of all $n = 11$ biologically independent G2 replicates. Bars show the mean of all comparisons.

stage, but resolve on a local scale.

To quantify this qualitative observations of global alignment and local separation, I calculated genome-wide scaling plots for the G2 samples for cis-sister and trans-sister contacts. In these plots, contacts are binned into bins of exponential spacing based on their contact distance [21]. Based on this assignment, genome-wide average contacts per bin are calculated. This number is then further normalized to the total number of possible genomic bins at a given separation to calculate a contact frequency per distance bin. Since prior studies have established that contact frequency decays exponentially with increasing genomic separation, scaling plots are typically displayed with log-transformed axes [21]. When such plots were prepared for cis-sister and trans-sister contact, the familiar exponential decay (visible as linear decay on a log-transformed plot), was readily detectable for cis-sister contacts (Fig. 3.17b). It is thought that the

¹²Panel(a) and its caption is adapted from ED Fig. 4a of [182]. Panel (b) caption is adapted from ED Fig. 4b of [182]

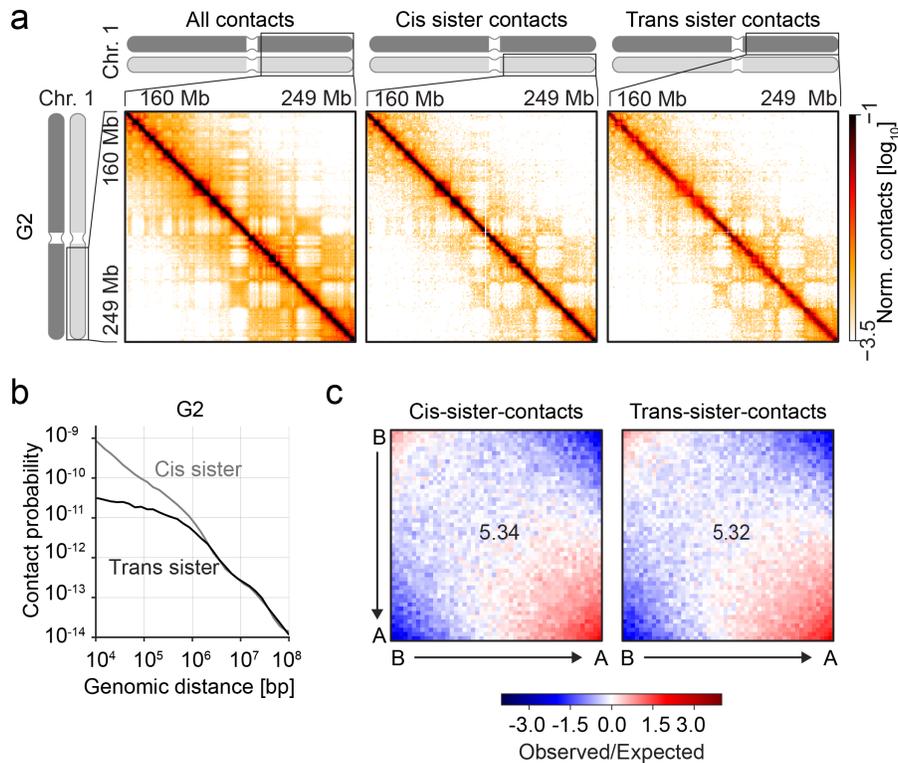


Figure 3.17: **Large scale organization of G2 sister chromatids.**¹³(a) Hi-C interaction matrices of the long arm of chromosome 1 of all contacts, cis-sister-, and trans-sister contacts merged G2 samples from $n = 11$ biologically independent experiments. The all-contacts matrix was normalized to the total number of corrected contacts in the region of interest (ROI), whereas cis-sister and trans-sister contacts were normalized to the total amount of cis-sister and trans-sister contacts in the ROI. Bin size of the matrix is 500 kb. (b) Average contact probability over different genomic distances for cis-sister and trans-sister contacts of the G2 sample shown in (a). (c) Saddle plots assessing compartment strength for cis-sister and trans-sister contacts of samples shown in (a). The inlaid numbers refer to the compartment score (see 2.21.19).

"hump", characterized by an initial lower decay from 10 kb to 500 kb, followed by a steeper one from 500 kb onward, is a signature of topologically associating domains (TADs), suggesting that these structures are present in this data set, in-line with previous reports of data derived from cells in the G2 cell cycle phase [74, 216].

In contrast to cis-sister contacts, trans-sister contacts showed a much shall-

¹³Panel (a) and (b) and captions are adapted from Fig. 2a and b of [182]

lower decay at short genomic distances (smaller than 3 Mb) and started off at a much lower level than cis-sister contacts (Fig. 3.17b). At long genomic distances, however, cis-sister and trans-sister contacts intermix and their contact probability was indistinguishable. The point at which this happens is roughly at 3 Mb, corroborating the qualitative observation that sister chromatids are locally resolved and globally aligned.

Since I qualitatively observed compartments both in cis-sister and trans-sister contact maps, I wanted to quantify to what degree compartmentalization is able to cross sister chromatid boundaries. A common way of displaying and quantifying compartment strength is called a saddle plot [41]. Here, contacts are binned together based on the "eigenvector value" (see 2.21.19 for details) of their respective genomic positions. This leads to two bin-values for each Hi-C pixel: one for the pixel's "x-coordinate" and one for its "y-coordinate" in the Hi-C matrix. Then, these bins are put into a table and the amount of contacts that span from one bin to another is entered. This representation carries information about compartment strength since it is immediately evident whether similar compartments tend to interact with each other more than expected and whether interactions between different compartments are depleted of contacts. This fact can be used to derive a compartment score (see 2.21.19 for details), which is a single number that describes the degree of compartmentalization in a Hi-C data set. When I generated saddle plots for cis- and trans-sister contacts, I noticed that the plots looked very similar and that both types of contacts exhibited a similar compartment score (Fig. 3.17c). This suggests that compartmentalization is similarly pronounced both within and between sister chromatids and that the forces responsible for compartmentalization are not sensitive to sister chromatid identity.

3.5.2 Large scale organization in prometaphase

After having established that sister chromatids are globally aligned, but locally resolved in G2 data sets, I wanted to see how this organization changes upon entry into mitosis. We know from prior work that sister chromatids resolve in mitosis, a process that starts at early prophase and culminates when sister chromatids are aligned at the metaphase plate [158]. While many studies were dedicated to sister chromatid resolution, so far this process has only been studied using bulk methods that can determine the overall degree of resolution, but are unable to resolve the fine structure and genomic neighborhood of sister chromatid organization [158]. To close this gap, I synchronized cells to prometaphase using nocodazole (see 3.3.3) and generated scsHi-C data sets (see Table 2). In order to determine the reproducibility of these data sets, I performed HiCrep analysis and concluded that all contacts, cis-sister contacts as well as trans-sister contacts were highly reproducible (Fig. 3.18a, b). When I looked at Hi-C heatmaps of the long arm of chromosome 1, I noticed that cis-sister contacts looked again very similar to all contacts and that the overall pattern is in line with published data sets of mitotic cells [27, 28]: A locus independent decay of contacts probability, loss of fine structure, decreased short-range con-

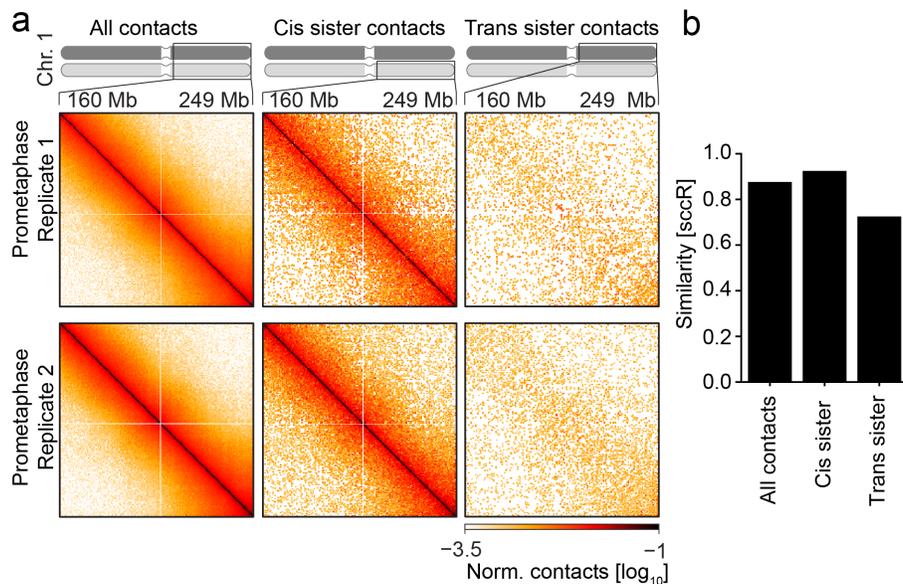


Figure 3.18: **Reproducibility of prometaphase scsHi-C.**¹⁴(a) Hi-C interaction matrices of the long arm of chromosome 1 of all contacts, cis-sister, and trans-sister contacts shown for two of the two prometaphase WT replicates. The all-contacts matrix was normalized to the total number of corrected contacts in the region of interest (ROI), whereas cis-sister and trans-sister contacts were normalized to the total amount of cis-sister contacts and trans-sister contacts in the ROI. Bin size of the matrix is 500 kb. (b) HiCrep analysis of all, cis-sister and trans-sister contacts of all $n = 2$ biologically independent prometaphase replicates. Bars show the mean of all comparisons.

tacts when compared to the G2 sample and increased long-range contacts (Fig. 3.19a). When I looked at trans-sister contacts, however, the behavior was very different. I observed an almost complete lack of trans-sister contacts, suggesting global resolution.

To quantify this notion, I again generated scaling plots to assess behavior of contact frequency as a function of genomic distance. Also here, cis-sister contacts conform to the published behavior of scaling in mitosis [27, 28]: Compared to G2, I observed a shallower decrease of contact probability for short-range contacts, a longer "plateau" of high contact probability for long-range contacts as well as a steep decay at around 10 Mb (Fig. 3.19b). Trans-sister contacts - on the other hand - assumed much lower values than cis-sister contacts over the entire range analyzed and never reach the level of cis-sister contacts. I can thus conclude that sister chromatids are resolved in prometaphase and move con-

¹⁴Panel (a) and (b) and their captions are adapted from ED Fig. 4c and d of [182]

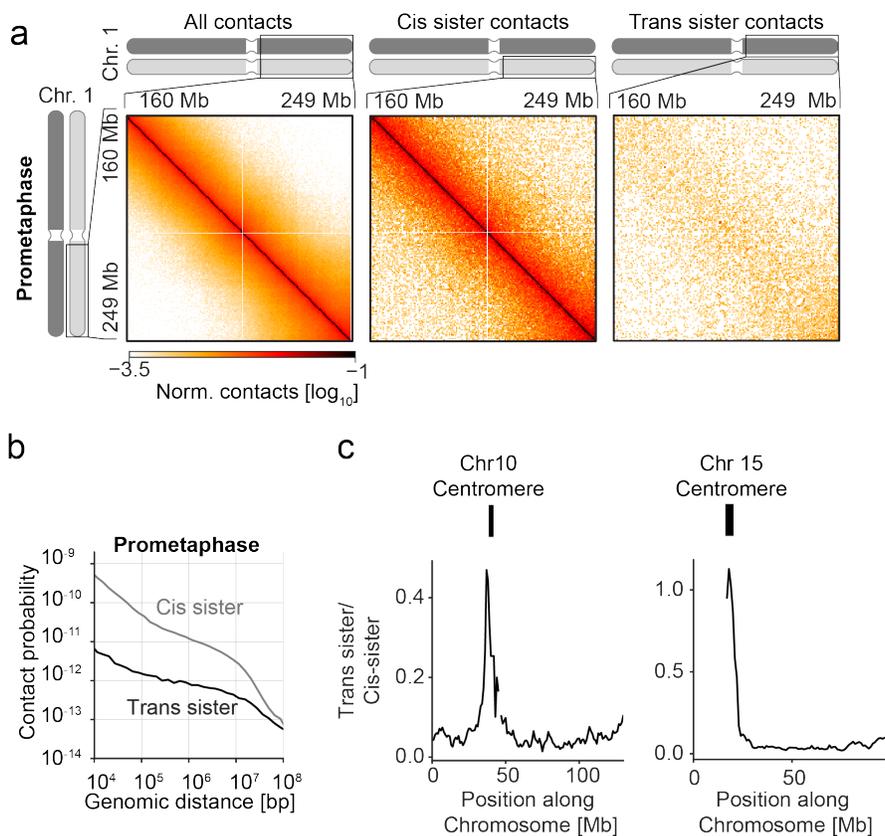


Figure 3.19: Large scale organization of Prometaphase sister chromatids.¹⁵(a) Hi-C interaction matrices of the long arm of chromosome 1 of all contacts, cis-sister-, and trans-sister contacts merged prometaphase samples from $n = 2$ biologically independent experiments. The all-contacts matrix was normalized to the total number of corrected contacts in the region of interest (ROI), whereas cis-sister and trans-sister contacts were normalized to the total amount of cis-sister and trans-sister contacts in the ROI. Bin size of the matrix is 500 kb. (b) Average contact probability over different genomic distances for cis-sister and trans-sister contacts of the prometaphase sample shown in (a). (c) Trans-sister over cis-sister contact ratio calculated along two representative chromosomes of the pooled prometaphase samples shown in (a). Values are the average of a sliding window of side-length 10 Mb.

siderably further apart as compared to G2. The scaling plots further suggest a strong signature of chromosome resolution in prometaphase. Having established that scsHi-C is able to detect sister chromatid resolution, I wanted to see

¹⁵Panel (a) and (b) and their caption are adapted from Fig. 2b and d of [182] respectively

whether I can pick up a signal connected to sister chromatid cohesion, the other major force acting on sister chromatids during mitosis.

It has been described before that in human cells, the prophase pathway removes most cohesin from chromosomal arms, leaving the centromeric population as the sole provider of sister chromatid cohesion, suggesting that cohesion should be strongest there [98]. The extent of centromere cohesion is typically measured in microscopy experiments using centromeric markers [98]. This however only measures the distance indirectly and does not yield insight into the precise chromatin structure surrounding the cohesion sites. Having scsHi-C data of prometaphase synchronized WT HeLa cells at hand, I wanted to see whether I can detect signals connected to centromeric cohesion. To this end, I looked at coarsely binned, whole chromosome Hi-C heatmaps and indeed saw an extreme accumulation of trans-sister contacts at centromeres, suggesting close proximity of sister chromatids. I then quantified the degree of sister chromatid proximity by calculating the average amount of trans-sister contacts within a sliding diamond of side length 10 Mb that was moved along the main diagonal. In order to avoid artificial inflation of this metric by increased mapping of reads to pericentromeric regions, I normalized the values to the average amount of cis-sister contacts within a similar diamond. This quantification showed a dramatic increase of trans-sister contacts near the centromere for all chromosome analyzed (Fig. 3.19c). However, the exact values of the maxima measured for the different chromosomes was different. One possibility is that the difference arises through different mappability of human centromeres: For some centromeres, genomic regions are resolved much closer to the α -satellites than for others [1]. This might then result in apparently differing degrees of cohesion, simply because closer regions are masked for some chromosomes. The other - more intriguing - possibility is that different chromosomes might exhibit different degrees of cohesion as measured by scsHi-C. In total, the striking signal at centromeres further corroborates the validity of scsHi-C and will also provide a powerful tool to quantify the extent of centromere cohesion in different contexts.

To sum up, the large scale structure of G2 and prometaphase data sets yield important insights into the relative organization of sister chromatids and provide credible validation for the scsHi-C method. Motivated by these findings as well as by the richness of structures already at this resolution for the G2 data, I went ahead and dove more deeply into this comprehensive data set.

3.6 TADs as units of trans-sister organization in G2

When I zoomed into the G2 data set and created maps with smaller binsizes, I noticed that trans-sister contacts diverged from cis-sister contacts substantially (Fig. 3.20a-c). Specifically, trans-sister contacts at bins around the main diagonal showed a much lower contact frequency than cis-sister bins, in line with the local separation that was already hinted at in the coarser representation. Additionally, I noticed that trans-sister contacts were not distributed randomly, but exhibited pronounced patterns. Specifically, areas of high trans-sister contact density alternated with areas that showed almost complete depletion of

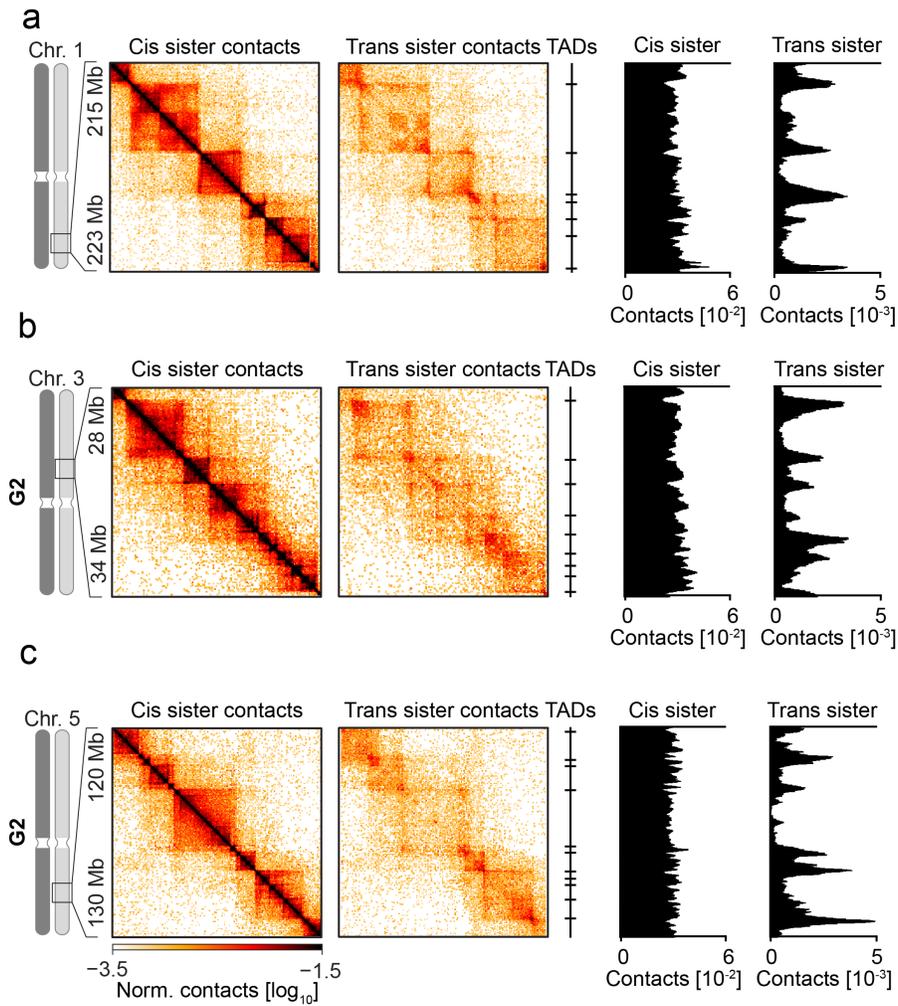


Figure 3.20: **Heterogeneity of trans-sister contacts of G2 synchronized cells.**¹⁶(a-c) Cis-sister- and trans-sister contacts of $n = 11$ biologically independent, merged G2 samples at representative regions on chromosome 1, 3 and 5 are displayed alongside the location of TAD boundaries (see 2.21.9 for details) and average trans-sister and cis-sister contact amount within a sliding window of 200 kb (see 2.21.10 for details). Bin size of matrix is 40 kb.

trans-sister contacts. Qualitatively, it also seemed as though trans-sister contact behavior was delineated by topologically associating domains (TADs) that were readily visible in cis-sister contacts. In addition, trans-sister contacts were often enriched at the boundaries of such domains. Due to the heterogeneous nature of trans-sister contacts and the striking divergence from cis-sister contacts, it was clear that established approaches to analyze Hi-C data had to be revisited.

3.6.1 New metrics to analyze scsHi-C data

So far there had been no analysis concept in traditional Hi-C analysis that would capture the local differences in contact-amount that was visible for trans-sister contact maps. The closest concept in recent literature that could be applied comes from analysis of homolog pairing in *D. melanogaster* Hi-C data [195]. Here, a metric was defined to capture the local amount of trans-homolog contacts by averaging contacts within a sliding diamond that was centered at the main Hi-C diagonal. I reasoned that this concept could be well suited to be applied to trans-sister contact maps and defined the contact density as the average contact frequency within a sliding window of half-length w in Hi-C bin units:

$$CD(i) = \frac{1}{v} \sum_{i-w \leq m, n \leq i+w} M_{m,n}$$

with $CD(i)$ denoting the contact density at bin i , M the Hi-C matrix and v being the number of valid Hi-C pixels within the window of summation. The contact density is thought to capture absolute variations in local contacts and is therefore expected to be useful for small scale comparisons of a limited number of regions. Indeed, when I calculated the contact densities for the regions shown in Fig. 3.20a-c, the enrichment is readily captured by this metric and hints further at localization of trans-sister enrichment at TAD-boundaries. Additionally, I defined the pairing score to be the contact density normalized to the genome-wide median:

$$PS(i) = \frac{CD(i) - \text{median}(CD)}{\text{std}(CD)}$$

with $PS(i)$ referring to the pairing score at genomic bin i , $\text{median}(CD)$ referring to the genome-wide median of CD and $\text{std}(CD)$ referring to the genome-wide standard deviation of CD . The pairing score - as defined here - allows to assess whether contacts are locally depleted or enriched compared to the genome-wide behavior. With these tools in hand, I went ahead to look for interesting correlations at regions that are enriched for trans-sister contacts.

¹⁶Panels and captions are adapted from ED Fig. 6 of [182]

3.6.2 CTCF is enriched at trans-sister contact sites

In order to find genomic locations that showed trans-sister contact enrichment in an unbiased fashion, the pairing score was calculated for a window of 50 kb genome-wide and all bins that were assigned a value above 0 were selected for enrichment analysis. Since the pairing score is normalized to the genome-wide behavior of the contact density, bins with a value above 0 correspond to the upper 50th percentile. To find unbiased associations of epigenetic characteristics and trans-sister contact enrichment, I used **Locus Overlap Analysis (LOLA)** [193].

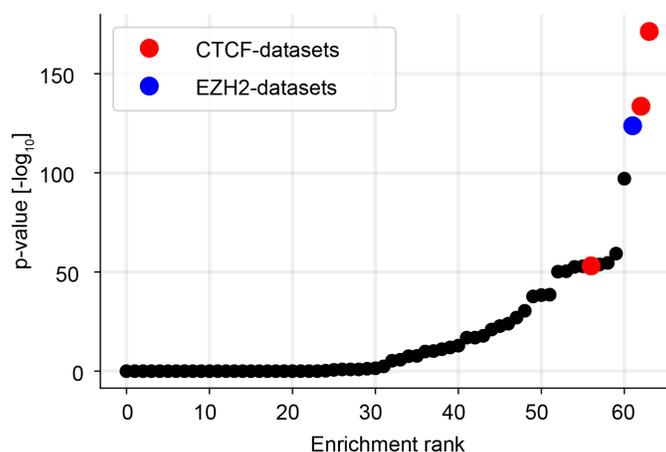


Figure 3.21: **LOLA analysis of regions with high trans-sister contact density.** LOLA results of the top 50th percentile of trans-sister contact density of G2 synchronized HeLa cells, sorted in descending order by p-Value. Data sets probing CTCF or EZH2 are highlighted.

LOLA is a tool that was developed as a parallel to established gene-set enrichment analysis [217], but with a focus on genomic intervals rather than specific genes. The idea behind LOLA is simple: A target set of regions serves as the input data set and the goal is to find out which genomic characteristics are significantly enriched amongst these as compared to all possible regions. Here, the choice of all possible regions - or the universe - is critical since it affects the enrichment results and can shift the focus of the entire analysis. For this particular data set, the upper 50th percentile of pairing score bins served as the input data set and all Hi-C bins with non-zero weights were chosen to be the universe. The query set of genomic characteristics consisted of all ChIP-seq data sets deposited in the ENCODE database [197] derived from HeLa cells.

When I looked at the result of the LOLA enrichment analysis, I saw that there are several factors that were highly enriched at regions with high trans-

sister pairing score. Amongst the top ten hits, there were three data sets that represent ChIP-seq experiments probing CTCF (Fig. 3.21). This is highly interesting since CTCF has been studied for years for its impact on the 3D-architecture of the genome and is thought to be one of the major determinants of mammalian genome folding [22, 23]. This finding also relates to the notion mentioned earlier that TADs might be units of trans-sister contact organization since CTCF is known to enrich at the boundaries of TADs [22]. With these hints that trans-sister contacts might be structured around TADs, I went ahead and investigated this association.

3.6.3 TAD-boundaries are contact sites between sister chromatids

In order to investigate TAD-specific organization of trans-sister contacts, I first needed to call TADs for our data set. In general, TADs have been defined as regions that interact more with themselves than with neighboring regions and many TAD callers are available that exploit slightly different characteristics of this initial definition [24, 134]. It has been suggested that the correlation of called TADs with known biological markers is a good indicator of the performance of a TAD caller [33]. Two such markers are the enrichment of CTCF and the cohesion subunit SMC3 at the boundaries of the called domains. Using these metrics, insulation score based callers have consistently outperformed more sophisticated approaches [33]. These programs call boundaries between TADs based on local minima of contact enrichment close to the diagonal. The idea is that such minima correspond to points that do not permit crossing contacts and thus constitute boundaries. However, these callers do not solve the "linking" problem, which refers to the decision of which boundaries should be grouped together to form TADs. OnTAD is an insulation score based caller that links domains by building local linking trees, taking into account that TADs can be nested [35]. OnTAD yields impressive results both when compared to other TAD callers and through visual inspection and I therefore went ahead and called TADs using OnTAD for the G2 WT data set.

To gain an understanding of sister chromatid organization at TADs, I first generated pileups centered on TAD-boundaries (Fig. 3.22a, b). Pileups are a good tool to understand the average behavior of a set of locations as they represent the pixel-wise average of all these regions. These pileups were performed both on the contact frequency directly (ICCF plots) and on contact frequency that was normalized to the expected number of contacts at a given genomic separation (see 2.21.6 for details). Such "observed over expected" plots allow a per-bin comparison of local enrichment or depletion as compared to the genome-wide average. When I looked at cis-sister contacts at TAD-boundaries, I noticed pronounced local depletion both in ICCF plots as well as when observed/expected values were plotted. This behavior is in line with the definition of TAD-boundaries as insulating loci. Moreover, "stripes" of enriched contact frequency were visible emanating from close to the TAD-boundary and extending towards the edge of the pileup region. These stripes are most likely a signature of aggregating differently sized focal enrichments - also called loops - that

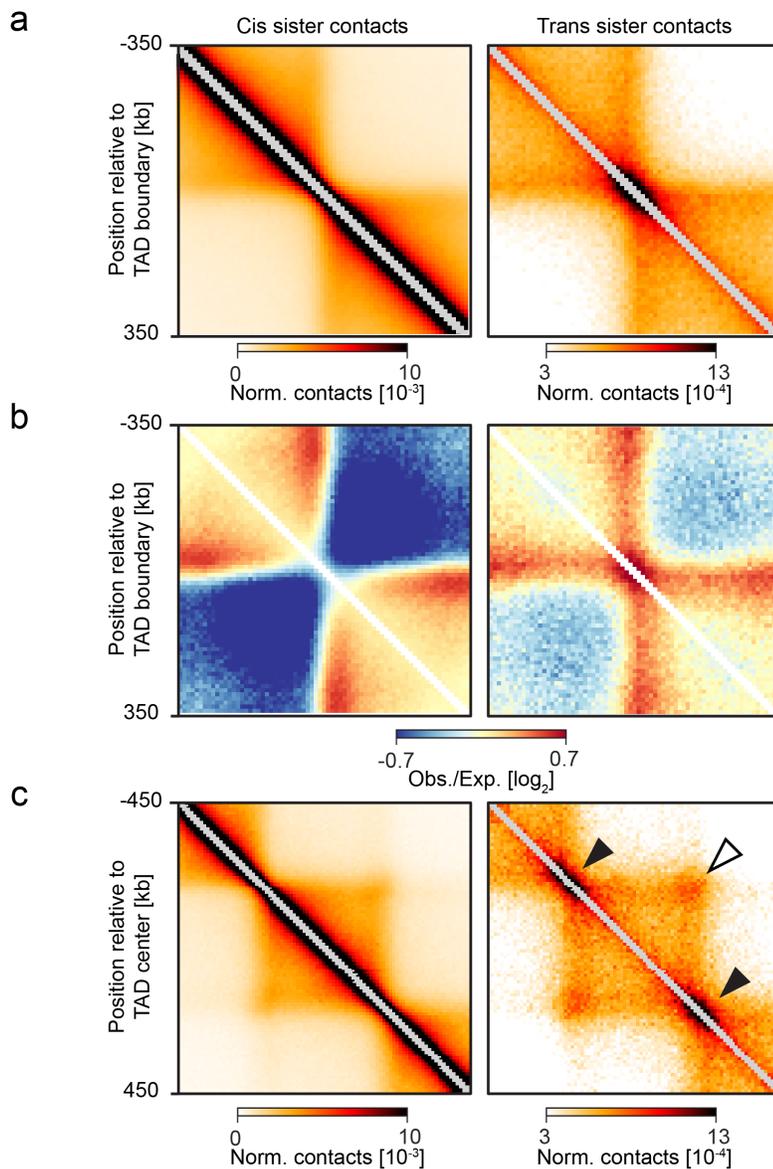


Figure 3.22: **TAD boundaries are sites of contact enrichment in replicated chromosomes.**¹⁷(a) Average cis-sister and trans-sister contact environment around TAD boundaries. ICE-normalized contacts are shown (see 2.21.4 for details). (b) Same data as in (a) but represented as observed/expected values. (c) Average cis-sister and trans-sister contact environment around TAD centers of TADs between 300 and 500 kb. ICE-normalized contacts are shown. Filled arrows indicate positions where the same TAD boundaries are connected across sister chromatids, whereas the hollow arrow indicates the connection of neighboring TAD boundaries across sister chromatids.

frequently occur at the edges of TADs [22].

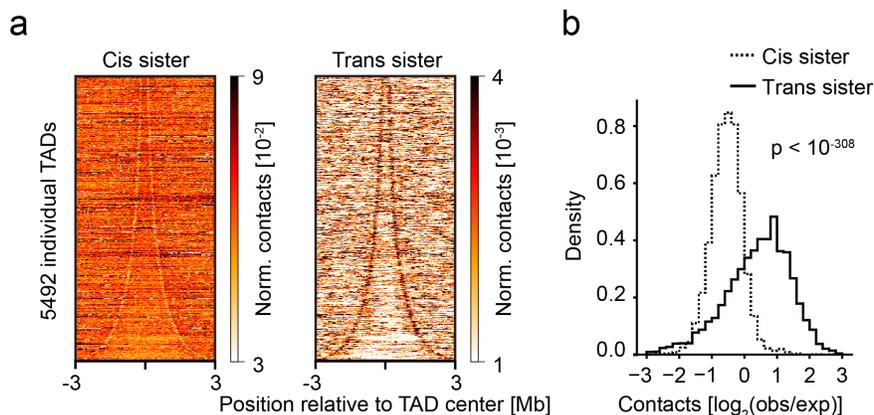


Figure 3.23: **Quantification of trans-sister enrichment at TAD boundaries.**¹⁸(a) Stack-up of average trans-sister and cis-sister contacts within sliding windows of 100 kb along TADs sorted by size. The panel shows windows of 6 Mb around the center of the respective TADs. (f) Quantification of trans-contact enrichment at TAD boundaries. The average observed/expected values for cis-sister and trans-sister contacts within a 80 kb window surrounding all ($n = 5801$) annotated TAD-boundaries (see 2.21.9 for details) are displayed as a histogram. P-value was calculated using a two-sided Mann-Whitney U test.

When I turned to trans-sister contacts, the behavior was strikingly different. Here, contacts were strongly enriched directly at the TAD-boundary, corroborating the qualitative observations made earlier. Interestingly, also here prominent stripes emanated from the boundary, although they appeared broader than the stripes seen for cis-sister contacts. Again, these stripes most likely arise through differently sized loops that were averaged together. To corroborate the average behavior of trans-sister contacts around TADs, I generated pileups of TAD centers, filtering for TADs that ranged in size between 300 and 500 kb. Also in these plots, the enrichment at TAD boundaries is clearly visible (Fig. 3.22c). Moreover, it seems that the interior of the average TAD is depleted of contacts, suggesting that TADs may form individually on each sister chromatid. Interestingly, in this plot, a loop signal is visible for trans-sister contacts, suggesting that not only the same TAD boundary is in contact with its copy across sister chromatids, but that also neighboring TADs exhibit this behavior.

The visualizations shown so far were qualitative in nature and do not give information about the distribution of trans-sister contact enrichment at TAD-boundaries. To look at this distribution, I calculated line-profiles of contact density along TADs and sorted them by size. This visualization clearly shows

¹⁷Panel (c) and captions is adapted from Fig. 3c of [182]

¹⁸Panel(a) and (b) and their caption are adapted from Fig. 3e and f of [182]

that most of the TADs that I called exhibit enrichment at their boundaries (Fig. 3.23a). To quantify this further, I calculated the average observed/expected contact value for a diamond of size 80 kb centered around all TAD-boundaries. This analysis suggests that most TADs show trans-sister enrichment and that this enrichment is roughly two-fold (Fig. 3.23b).

Thus, trans-sister contact distribution is very heterogeneous and seems to be dominated by the signature of cis-sister TADs. Moreover, trans-sister contacts enrich at almost all TAD-boundaries, whereas TAD-centers are depleted of trans-sister interactions. However, when I looked more closely at the contact maps, not all TADs showed depletion of trans-sister contacts within them. On the contrary, there was a population that was strikingly enriched for trans-interactions.

3.7 H3K27me3 determines trans-sister TAD contact state

Having established that TADs are the organizing principle of trans-sister contacts, I started to look closer at their internal organization and noticed that some TADs are filled completely with contacts, with no strong apparent decay of contact probability towards TAD edges. These regions rather constituted discrete "blocks" of uniformly increased trans-sister interactions. Interestingly, other TADs were completely devoid of trans-sister contacts, while still exhibiting a strong TAD-boundary signal (Fig. 3.24a, b). Moreover, qualitative assessment of cis-sister contacts did not show a similar behavior for those contacts, with there being no difference for cis-sister contacts between "filled" and "hollow" trans-domains. This suggests that trans-sister organization is very heterogeneous and not at all uniform along the genome.

Given this striking heterogeneity, I wanted to test whether it is possible to reproduce these findings using other methods. The strategy to do this was to calculate the contact density (see 3.6.1) of trans-sister contacts for scsHi-C data of G2 synchronized WT HeLa cells and correlate these numbers with results obtained via different means [181]. Specifically, for the calculation of the contact density, a diamond side length of 600 kb was chosen as it lies close to the median size of our called TADs. The first data set that I correlated with the scsHi-C data was generated using HeLa cell lines that harbor dCas9-GFP targeting different genomic regions [181]. Here, HeLa cells were synchronized to the G2 cell cycle stage and the fraction of cells that exhibit two dCas9-GFP spots was taken as a metric for sister chromatid distance. Reassuringly, scsHi-C contact density correlated very well with this independent sister chromatid distance metric (Fig. 3.25a). As a further control, I also correlated scsHi-C data with FISH data sets generated using probes that target the binding sites of the dCas9-GFP cell lines used before [181]. Interestingly, also here I found a very good correlation between scsHi-C data and FISH probe distance in G2 synchronized HeLa cells (Fig. 3.25b). Taken together, this suggests that independent methods of sister chromatid distance determination correlate well

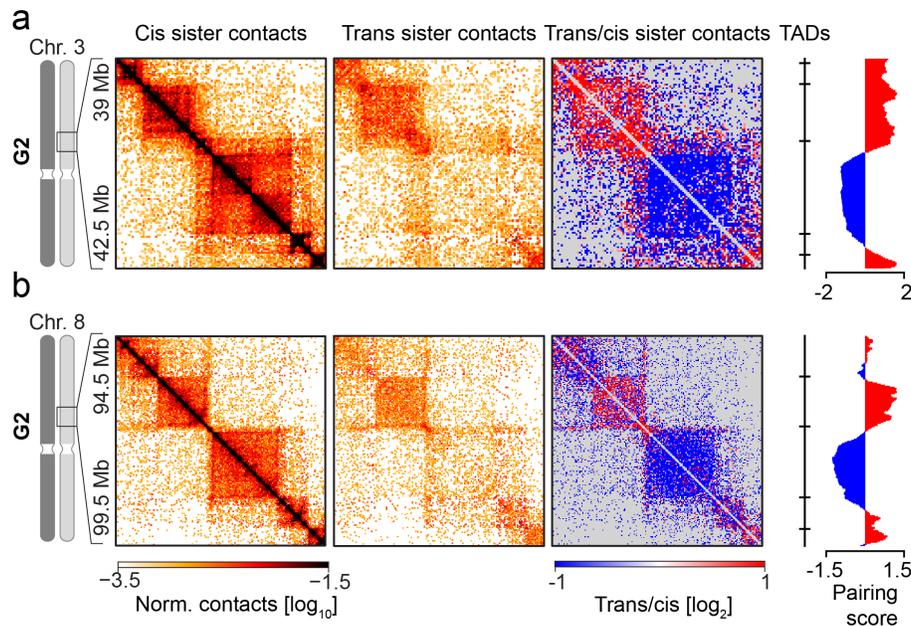


Figure 3.24: **Dichotomy of trans-sister TAD behavior.**¹⁹(a-b) All-contacts, cis-sister and trans-sister contacts, as well as the ratio of trans-sister observed/expected to cis-sister observed/expected of merged G2 samples from $n = 11$ biologically independent experiments at a representative region on chromosome 3 and 8 is displayed alongside the location of TAD boundaries and the trans-sister pairing score within a sliding diamond of 400 kb (see 2.21.10 for details). Bin size of the matrix is 30 kb and 20 kb respectively.

with scsHi-C in a size regime where trans-sister heterogeneity is observed.

Having gained confidence in the heterogeneity found for different TADs, I sought a way to quantify it. The idea was to take the top and bottom 10 % of TADs in terms of trans-sister contacts to obtain a strong signal for factors that are enriched for one type or the other. Specifically, I took the average amount of trans-sister contacts or contact density (see 3.6.1) for each TAD and coined the top 10 % "highly paired" and the bottom 10 % "highly unpaired" (Fig. 3.26a). When I computed stack ups of line profiles along these TADs sorted by size, I noticed that highly paired domains were smaller than highly unpaired domains (Fig. 3.26b). While this behavior is interesting, I also wanted to know whether an enrichment of trans-sister contacts is associated with a particular epigenetic state.

In order to gain insight into the epigenetic signature of highly paired domains I again turned to LOLA (see 3.6.2; [193]). For this particular analysis,

¹⁹Panel (a) and caption are adapted from Fig. 2e of [182] and Panel (b) and caption is adapted from ED Fig. 5a of [182]

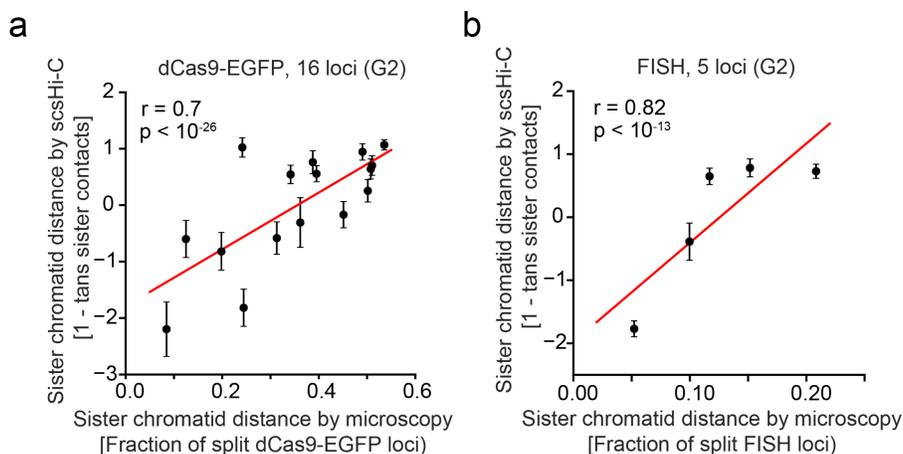


Figure 3.25: **Validation of scsHi-C using microscopy.**²⁰(a) Comparison of sister chromatid separation at 16 genomic loci measured by live cell microscopy and scsHi-C. Microscopy analysis was by live-cell imaging of 16 HeLa cell lines expressing dCas9-EGFP with different locus-specific gRNAs, using automated detection of merged or split sister loci in G2 cells, as reported in [181]. scsHi-C quantification of sister locus distance was done by calculating $(1 - \text{average trans-sister contacts})$ in a region spanning 600 kb around each FISH target site and standardizing the resulting value (see 2.21.3 for details). Two-sided p-value for a Wald-test with t-distribution of the test statistic is shown with the null hypothesis being a zero slope. (b) Comparison of sister chromatid separation at 5 genomic loci measured by fluorescence in situ hybridization (FISH) and scsHi-C. scsHi-C quantification of was done as in (a). Each dot indicates one target locus, measured in $n = 11$ biologically independent HeLa WT G2 samples by scsHi-C. Two-sided p-value for a Wald-test with t-distribution of the test statistic is shown with the null hypothesis being a zero slope

I used the highly paired TADs as input data set and defined all TADs as the region universe. Initially, I used all HeLa specific data sets deposited with the Roadmap epigenomics project as a query database [193], comprising 12 different - mostly histone modification - data sets. Strikingly, the only modification that was significantly enriched within highly paired domains was trimethylation of H3K27 (Fig. 3.26c). When I then quantified the amount of H3K27me3 at highly paired and highly unpaired domains, I indeed measured a highly significant enrichment of H3K27me3 at highly paired domains (Fig. 3.26d). In addition, when I look back at the analysis done for highly paired Hi-C bins, EZH2 - the methyltransferase that installs H3K27me3 - is the third most enriched factor (Fig. 3.21). Taken together, this suggests that H3K27me3 might

²⁰Panel (a) and (b) and caption are adapted from ED Fig. 5b and c of [182]

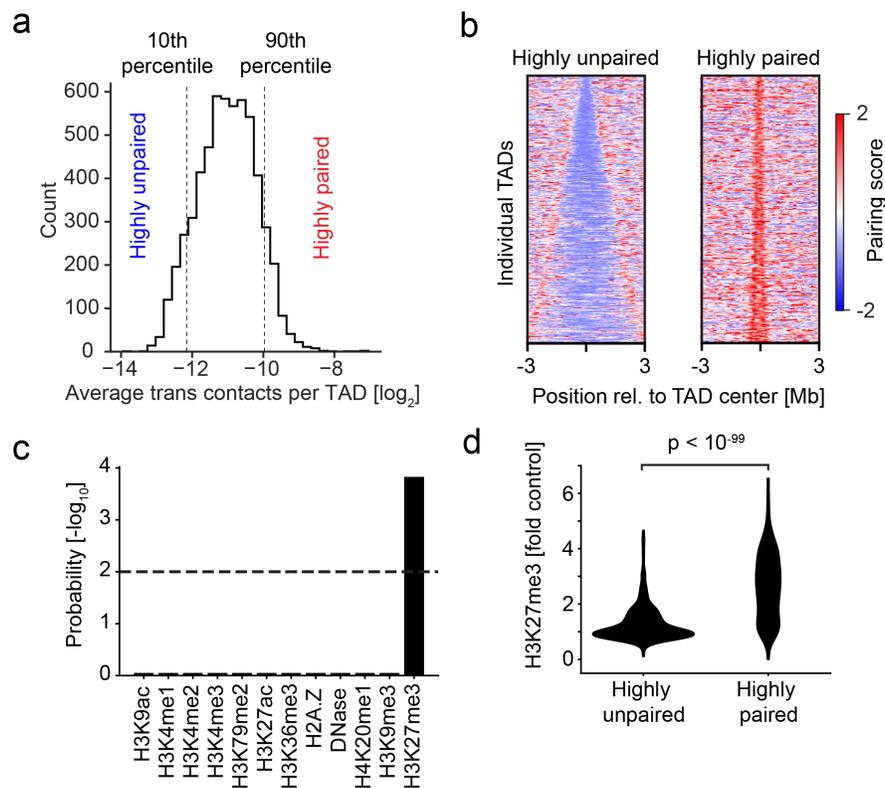


Figure 3.26: **Correlation of H3K27me3 with trans-sister contact enrichment.**²¹ (a) Histogram of average trans-sister contact frequency for annotated TADs. Vertical lines indicate the cut-offs for “highly paired” and “highly unpaired” TADs (b) Stack-up of trans-sister pairing score along TADs that are highly paired or highly unpaired, sorted by the size of TADs. Shown are windows of 6 Mb around the center of the respective TADs. Pairing scores were calculated within a sliding window of 200 kb on a Hi-C matrix with 20 kb bin size. (c) Visualization of enrichment analysis that was done on TADs that exhibit high pairing using LOLA. The panel shows all chromatin modification data sets in the extended LOLA database for HeLa cells with their respective p-value. P-value cut-off ($p < 0.01$) is displayed as a dashed line. (d) Quantification of H3K27me3 enrichment at highly paired ($n = 490$) and highly unpaired ($n = 570$) TADs, displaying the average fold-enrichment of H3K27me3 within the respective intervals.

be an important determinant of trans-sister contact density.

To quantify to what degree H3K27me3 determines trans-sister contact density and how this compares to other candidate determinants, I turned to a predictive model. To this end, I collected a broad array of different data sets

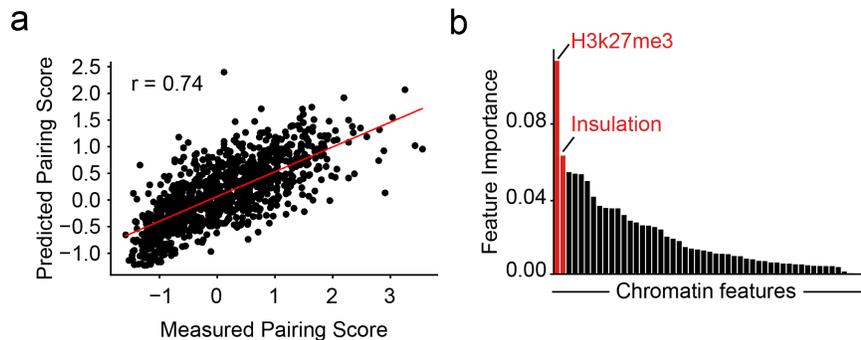


Figure 3.27: **Prediction of trans-sister contacts from chromatin features.** (a) Correlation analysis between predicted pairing score using a random forest regressor on an array of chromatin features (see 2.21.16 and Table 7 for details) and actual pairing score of trans-sister contacts within a sliding window of 50 kb. (b) Feature importance of features used (see Table 7) in the regression analysis shown in (a).

comprising chromatin associated factors, histone modifications, replication timing as well as Hi-C derived metrics such as insulation score and compartment identity (see Table 7 for details). To represent both the intensity of a signal as well as the inferred binding sites, I included both read pileup data as well as ChIP-seq peak calls in the used feature set. In order to clarify to what degree the amount of trans-sister contact density is determined by these factors, I used a random forest regressor to predict the pairing score genome-wide [218]. Interestingly, this collection of features contained enough information to fit a model that achieved an R^2 -score of 0.53, suggesting that a large fraction of the variation in the data can be explained (Fig. 3.27a). Importantly, random forest regressors allow the assessment of feature importance [219], yielding information about how important a certain feature was in the predictive outcome. Interestingly, the most important feature to predict trans-sister contact density was H3K27me3, followed by the insulation score (Fig. 3.27b). This suggests that an interplay of loop extrusion and chromatin state determines the sister chromatid interaction landscape.

3.8 Influence of Sororin on chromatin structure

Sororin is a protein that has been known for years to be essential for sister chromatid cohesion [170, 220]. It has been shown that sororin is constantly required to perform this function since degradation of sororin in G2 leads to cohesion loss in the subsequent mitosis [171]. Sororin is thought to act by preventing release of cohesin from chromatin by the cohesin release factor Wapl

²¹Panel (a) and caption are taken from ED Fig. 5d of [182]. Panel (b), (c) and (d) and their captions are taken from Fig. 2f, g and h of [182]

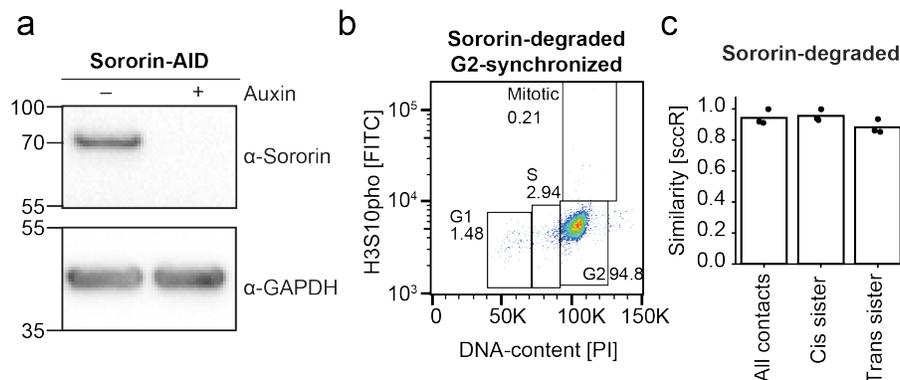


Figure 3.28: **Characterization of Sororin-AID HeLa cell line.**²² (a) Western blot for Sororin and GAPDH of HeLa Sororin-AID cells synchronized to G2 and either treated with auxin (+) or H₂O (-) (b) Cell cycle analysis of HeLa Sororin-AID cells synchronized to G2 and treated with auxin. Panel shows a FACS plot of cells stained for pH3S10 to mark mitotic cells and propidium iodide to measure DNA content. Gates for different cell cycle stages are shown and the indicated numbers reflect percentage of cells that were measured. (c) HiCrep analysis of all, cis-sister and trans-sister contacts of all replicates of Sororin-AID cells treated with auxin. Bars show the mean of all comparisons.

and is - together with cohesin acetylation - required to form stable cohesion after DNA replication [169]. In this regard, sororin can be viewed as a marker for the stable population of cohesin that exhibits residence times of multiple hours.

Prior work that looked at the function of sororin in connection with sister chromatid cohesion has only looked at indirect read-outs of this process by examining the morphology of sister chromatids on chromosome spreads or was limited by throughput as in live-cell labeling or FISH experiments [170, 171]. It is therefore not known whether sororin is required to hold sister chromatids in close proximity prior to mitosis genome-wide. Furthermore, whether sororin has a role in cohesin's function of shaping the 3D genome is also not known.

3.8.1 Sororin degradation leaves cis-sister structures intact

In order to elucidate these questions, I wanted to probe sister chromatid specific chromatin structure in the absence of sororin using scsHi-C. In order to do that, I decided to employ the auxin inducible degron (AID) system. The AID system makes use of the conditional ubiquitination of proteins harboring a specific protein tag ("AID-tag") by a plant-derived E3-ligase [221]. This sys-

²²Panel (a) and (b) and their caption are adapted from ED Fig. 7a and b of [182]. Panel (c) caption is adapted from ED Fig. 7f of [182]

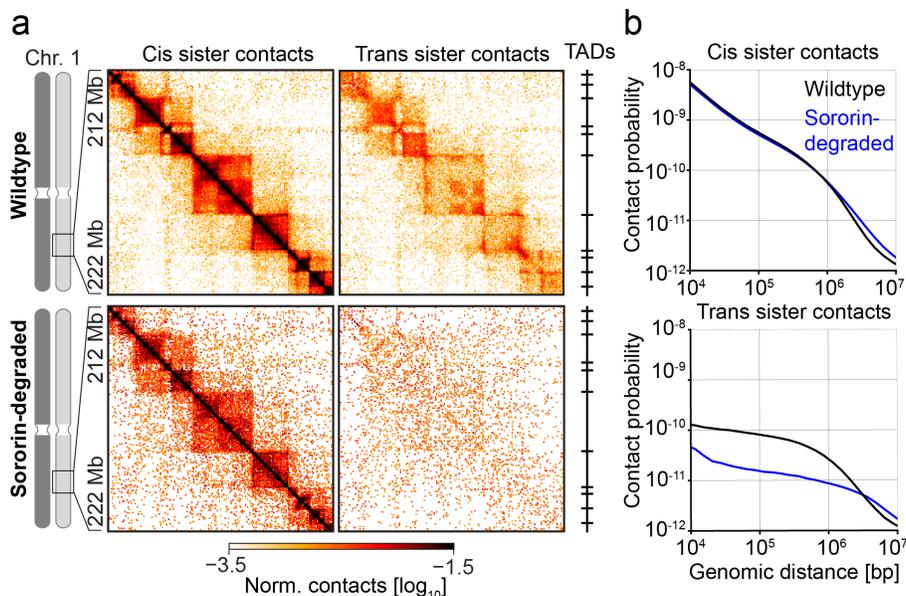


Figure 3.29: **scsHi-C analysis of Sororin-degraded HeLa cells.**²³(a) Cis-sister and trans-sister contacts of merged Sororin-degraded samples (from $n = 3$ biologically independent experiments) at a representative region on chromosome 1 are displayed alongside the location of TAD boundaries (see 2.21.9 for details). Bin size of the matrix is 40 kb. (b) Average contact probability over different genomic distances for cis-sister and trans-sister contacts of the different G2 sample shown in (a).

tem allows to target tagged proteins for proteasomal degradation by addition of Indol-3-acetic acid, an auxin plant hormone. So to deplete sororin, I obtained a HeLa Kyoto cell line from the lab of Jan-Michael Peters that carried homozygously AID-tagged sororin alleles (subsequently called "Sororin-AID cell line") (Fig. 3.28a).

I first tested whether the synchronization scheme (see 3.3.3) that was developed for WT cells would also yield a synchronous G2 population for the Sororin-AID cell line. I decided to degrade sororin prior to entry into the S-Phase that establishes the 4sT label since this would allow to completely prevent the establishment of a stable cohesin population. When I performed the synchronization protocol and assessed synchrony via FACS it was apparent that the protocol developed for WT cells also yields a highly synchronous G2 population of the Sororin-AID cell line (Fig. 3.28b). So I went ahead and generated three replicates of G2 synchronized samples where sororin was degraded and sequenced them deeply (see Table 4).

Before performing in-depth analysis, I tested whether the generated replic-

²³Panel (a) and caption are adapted from Fig. 4a of [182]

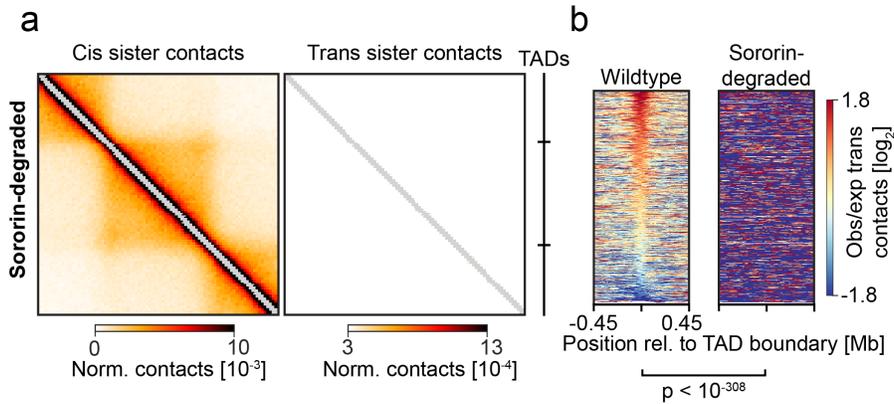


Figure 3.30: **TAD-level analysis of sororin-degraded HeLa cells.**²⁴(a) Average cis-sister and trans-sister contact environment around TAD centers of TADs between 300 and 500 kb in G2 synchronized cells with degraded Sororin. The panel shows ICE-normalized contacts binned at 10 kb in a window of 900 kb. (b) Stack-up of average observed/expected values within sliding windows of 100 kb around TAD boundaries of G2 wildtype samples and Sororin-degraded samples. The panel shows windows of 900 kb. The rows are sorted based on the center enrichment of the G2 wildtype condition. P-values were calculated using a two-sided Mann-Whitney U test performed on the $n = 5801$ values in the center columns of the respective stack-up matrix.

ates were reproducible using HiCRep, which suggested highly reproducible Hi-C data sets for all read types examined (Fig. 3.28c). When I looked at heatmap representations of cis-sister contacts derived from the sororin-degraded samples they were very similar to the same regions displayed for the G2 WT sample (Fig. 3.29a). Indeed, TADs were present to a similar degree regardless of sororin degradation state based on visual inspection of example regions and pileups of TAD centers (Fig. 3.30a; compare 3.6.3 for WT examples). When I looked at scaling plots of cis-sister contacts of sororin-degraded samples, they also showed very similar behavior to cis-sister WT data sets. However, when I looked more closely, cis-sister contacts were slightly enriched at longer genomic ranges compared to WT (Fig. 3.29b). This potentially hints at increased chromatin mobility due to the absence of cohesive cohesin. To sum up, sororin is not required to a large extent to shape the familiar appearance of within sister contact patterns.

3.8.2 Sororin is required for sister chromatid alignment in G2

When I turned my focus to trans-sister contacts, the data for sororin-degraded samples looked very different from WT data sets: When looking at heatmap

²⁴Panel (a) and (b) and caption are adapted from Fig. 4c and d of [182]

representations, trans-sister contacts were much less pronounced, suggesting a dealignment of sister chromatids (Fig. 3.29a). This notion was corroborated by scaling plots, where trans-sister contacts are strongly depleted locally up to 1 Mb, before reaching WT levels again (Fig. 3.29b). To test whether this depletion is also detectable on average TAD representations, I calculated pileups of TAD centers similar to the analysis performed for WT samples (see 3.6.3). Also here, a strong depletion of trans-sister contacts was visible, resulting in complete loss of TAD specific trans-sister patterns upon sororin degradation (Fig. 3.30a). In order to clarify whether TAD boundary enrichment might still persist, I calculated stack ups of profiles around TAD boundaries, representing observed/expected values. Also here, no detectable enrichment was visible for the sororin-degraded sample (Fig. 3.30b). Taken together, this suggests that sororin degradation leads to massive dealignment of sister chromatids up to regions of 1 Mb of separation. Moreover, not only are trans-sister contacts much less pronounced upon sororin degradation, they are also not enriched at TAD boundaries anymore, suggesting that cohesive cohesin is necessary for enrichment of trans-sister contacts at the boundaries of TADs.

3.9 Loop extrusion is required for local separation of sister chromatids

Theoretical work using polymer simulations suggests that loop extrusion is a central principle of sister chromatid organization [30]. In these studies, the authors showed that a completely entangled pair of sister chromatids can be separated completely by an appropriated amount of simulated loop extrusion factors (LEFs). Given that loop extrusion correctly predicts and recapitulates many aspects of interphase chromatin organization as measured by conventional Hi-C [32], I wondered whether I could test the predictions of the loop extrusion model on sister chromatid specific structure using scsHi-C.

The group of Jan-Michael Peters and others recently showed that the SMC protein cohesin is able to perform loop extrusion *in vitro* [97, 115]. Through these experiments it became apparent that NIPBL is an essential cofactor that is required for cohesin mediated loop extrusion. I thus hypothesized that if I degrade NIPBL in cells that are synchronized to the G2 cell cycle stage, I would obtain cells that harbor cohesive cohesin similarly to WT, but would lack the ability perform cohesin mediated loop extrusion. In such a situation, one could directly test whether loop extrusion is responsible for the local separation of sister chromatids that was observed in WT cells and thus whether loop extrusion has the potential to separate sister chromatids *in vivo*.

To this end, I obtained a HeLa cell line from the lab of Jan-Michael Peters that harbored homozygously AID-tagged NIPBL alleles (Fig. 3.31a). I first wanted to test whether the developed synchronization scheme (see 3.3.3) permits also synchronization of this cell line to the G2 cell cycle stage. So I went ahead and performed FACS analysis of such synchronized samples and found that the synchrony was comparable to WT cells (Fig. 3.31b). Since my hypotheses depended on the assumption of functional cohesion, I wanted to test

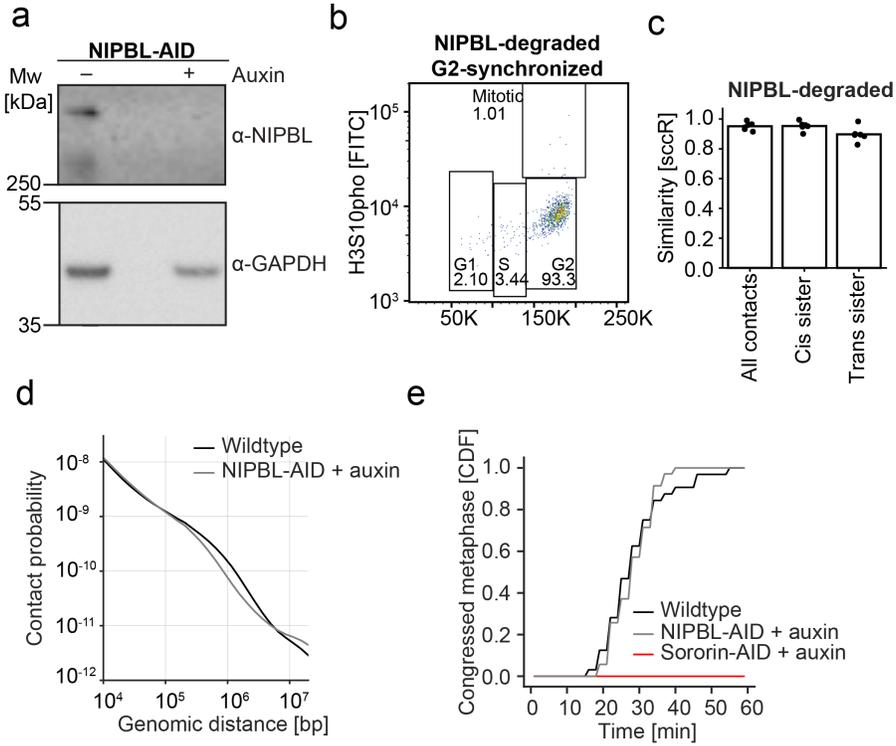


Figure 3.31: **Characterization of NIPBL-AID HeLa cell line.**²⁵(a) Western blot for NIPBL and GAPDH of HeLa Sororin-NIPBL cells synchronized to G2 and either treated with auxin (+) or H2O (-) (b) Cell cycle analysis of HeLa Sororin-NIPBL cells synchronized to G2 and treated with auxin. Panel shows a FACS plot of cells stained for pH3S10 to mark mitotic cells and propidium iodide to measure DNA content. Gates for different cell cycle stages are shown and the indicated numbers reflect percentage of cells that were measured. (c) HiCrep analysis of all, cis-sister and trans-sister contacts of all replicates of NIPBL-AID cells treated with auxin. Bars show the mean of all comparisons. (d) Contact probability of all contacts at different genomic distances of HeLa NIPBL-AID cells synchronized to G2 ($n = 4$ biologically independent experiments) that were treated with auxin and HeLa WT cells synchronized to G2 ($n = 11$ biologically independent experiments). (e) Chromosome congression analysis of NIPBL- and Sororin-depleted cells. Panel shows the cumulative frequency of cells congression their chromosomes in metaphase after entering mitosis in a RO3306 wash-out. Pooled replicates are shown from $n = 2$ biologically independent experiments

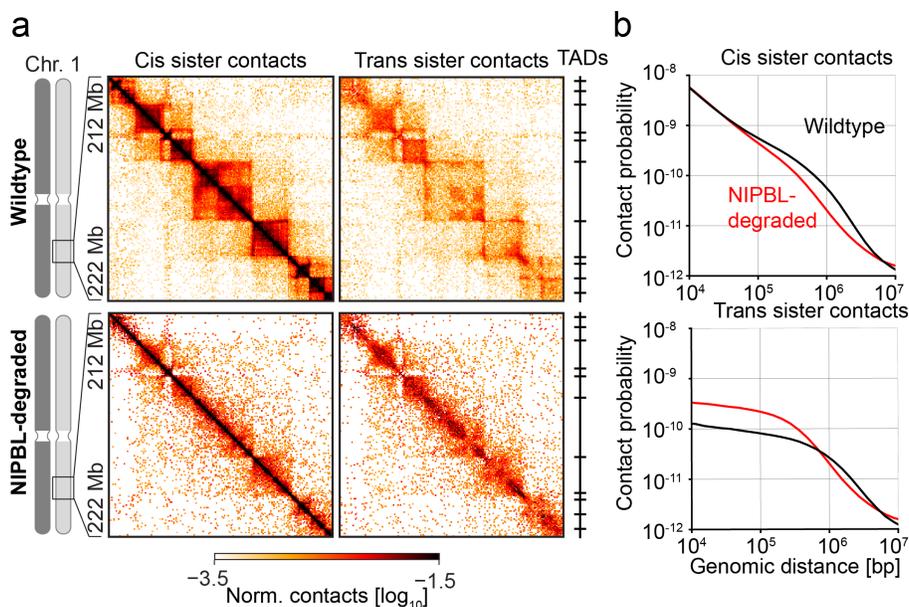


Figure 3.32: **scsHi-C analysis of NIPBL-degraded HeLa cells.**²⁶(a) Cis-sister and trans-sister contacts of merged NIPBL-degraded samples (from $n = 4$ biologically independent experiments) at a representative region on chromosome 1 are displayed alongside the location of TAD boundaries (see Method section for details). Bin size of the matrix is 40 kb. (b) Average contact probability over different genomic distances for cis-sister and trans-sister contacts of the different G2 sample shown in (a).

whether cohesive cohesin is unperturbed under these experimental conditions. To this end, I synchronized WT and NIPBL-AID cells without the addition of auxin to G2 (see 3.3.3) and released them into mitosis via wash-out of RO3306. As a positive control, I included Sororin-AID cells that were treated with auxin. If cohesion were perturbed in the NIPBL-AID cell line, congression of chromosomes at the metaphase plate would be severely delayed [222]. However, this experiment suggested that the cumulative fraction of cells that exhibit aligned metaphase chromosomes with progressive time was almost identical between the WT and the NIPBL-AID cell line (Fig. 3.31e). This suggests that cohesion is not perturbed as a result of basal NIPBL degradation. Thus, I went ahead and generated four scsHi-C replicates of NIPBL-degraded G2 synchronized samples and subjected them to deep sequencing (see Table 3).

HiCRep analysis showed that the generated replicates were highly reproducible for all read types examined and I thus proceeded with in-depth analysis

²⁵Panel (a), (b), (c) and (d) and their caption are adapted from ED Fig. 7a, b, c and d of [182]

²⁶Panel (a) and caption are adapted from Fig. 4a of [182]

(Fig. 3.31c). Prior Hi-C experiments performed on NIPBL deficient samples showed a characteristic loss of signatures that were related to loop extrusion: TAD signatures were much less pronounced, Hi-C loops were severely diminished and scaling plots exhibited a characteristic "linearization" [74]. In order to test whether the NIPBL-degraded samples generated here also show this signature, I first generated scaling plots of all contacts and observed the mentioned "linearization" connected to loss of TADs (Fig. 3.31d). Moreover, inspection of individual regions of cis-sister contacts contact maps suggested an almost complete loss of TAD associated structures (Fig. 3.32a). Having established that the NIPBL degradation recapitulates published signatures of loop extrusion loss, I explored trans-sister patterns.

Interestingly, trans-sister contacts were highly increased and much more restricted to the main diagonal as compared to WT samples, indicative of increased sister chromatid alignment (Fig. 3.32a). This behavior is corroborated by genome-wide scaling plots, where the "contact point" between cis-sister and trans-sister contacts can be detected at much lower genomic separations as in the WT sample (Fig. 3.32b). Taken together, this suggests that loop extrusion is necessary for the local separation of sister chromatids seen in WT cells.

I next wanted to test the influence of NIPBL degradation on the observed TAD specific trans-sister patterns. So I performed pileups of TAD centers as for WT cells (see 3.6.3). Also here, the increased amount of trans-sister contacts throughout the entire TAD is evident (Fig. 3.33a). Since the overall level of trans-sister contacts was increased, it could be that trans-sister contacts still enriched at TAD-boundaries with elevated baseline. To test this, I calculated stack ups of lineprofiles around TAD boundaries. This analysis suggested that the TAD boundary enrichment was much less pronounced for NIPBL degraded cells as compared to WT samples (Fig. 3.33b).

In order to test whether NIPBL degradation has an influence on the highly paired and highly unpaired domains that I detected in WT samples (see 3.7), I calculated stack ups of these domains sorted by size representing the pairing score. Interestingly, while highly unpaired domains were much less depleted of contacts upon NIPBL degradation, highly paired domains still exhibited strong enrichment of trans-sister contacts, suggesting that a mechanism independent of loop extrusion is involved in their establishment (Fig. 3.33c, d).

To sum up, these data suggest that loop extrusion is capable of separating sister chromatids *in vivo*, leading to local separation in the G2 cell cycle stage. Furthermore, NIPBL is necessary to position trans-sister contact foci at TAD-boundaries. Finally, trans-sister contact landscape is not completely uniform in NIPBL degradation conditions, suggesting that loop extrusion is not the only force shaping sister chromatid contacts.

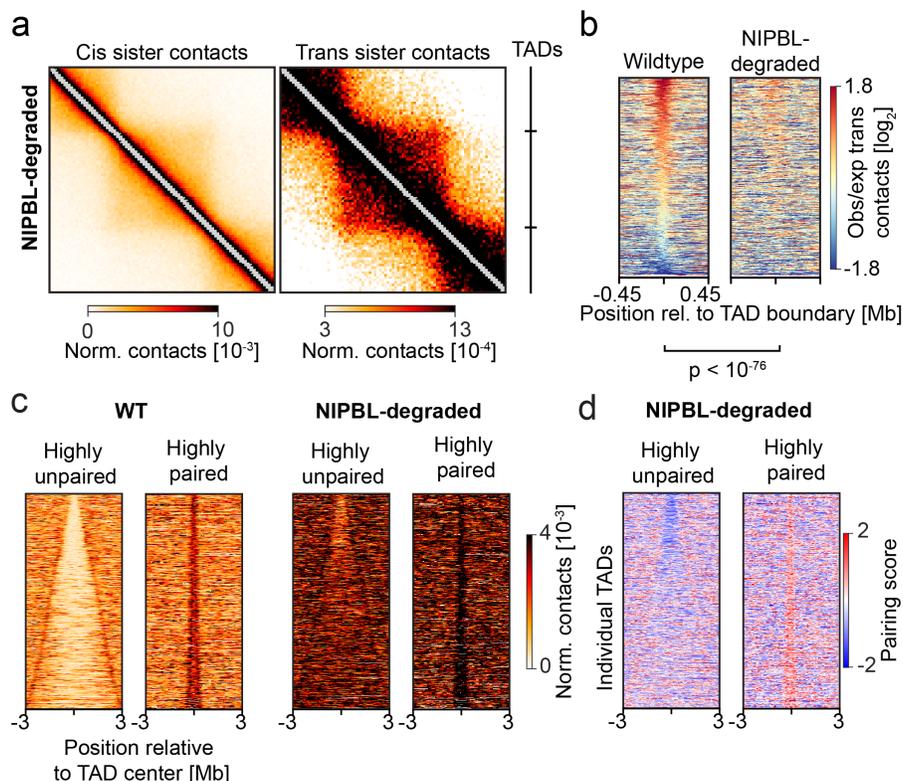


Figure 3.33: **TAD-level analysis of NIPBL-degraded HeLa cells.** (a) Average cis-sister and trans-sister contact environment around TAD centers of TADs between 300 and 500 kb in G2 synchronized cells with degraded NIPBL. The panel shows ICE-normalized contacts binned at 10 kb in a window of 900 kb. (b) Stack-up of average observed/expected values within sliding windows of 100 kb around TAD boundaries of G2 wildtype samples and NIPBL-degraded samples. The panel shows windows of 900 kb. The rows are sorted based on the center enrichment of the G2 wildtype condition. P-values were calculated using a two-sided Mann-Whitney U test performed on the $n = 5801$ values in the center columns of the respective stack-up matrix. (c) Stack-up of trans-sister corrected interaction frequency along TADs that are defined as highly paired or highly unpaired in WT cells, sorted by the size of TADs for NIPBL-degraded cells ($n = 4$ biologically independent, merged experiments) and WT cells ($n = 11$ biologically independent, merged experiments). Shown are windows of 6 Mb around the center of the respective TADs. Pairing scores were calculated within a sliding window of 200 kb on a Hi-C matrix with 20 kb bin size. (d) Stack-up of trans-sister pairing score along TADs that are highly paired or highly unpaired in WT cells, sorted by the size of TADs for NIPBL-degraded cells ($n = 4$ biologically independent, merged experiments). Shown are windows of 6 Mb around the center of the respective TADs. Pairing scores were calculated within a sliding window of 200 kb on a Hi-C matrix with 20 kb bin size.

4 Discussion

4.1 4sT is a new DNA label with many applications

Within this work, I established that 4-thio-thymidine (4sT) is a synthetic nucleotide that can be incorporated at high density into genomic DNA of a variety of different cell types, both cancer-derived and non-transformed. I showed that 4-thio-thymidine has low toxicity and does not induce DNA damage, which is noteworthy since other thiol containing nucleotides such as 6-thio-guanosine are being used for cancer therapy, partly due to their ability to cause DNA aberrations [204, 223]. Interestingly, 4-thio-thymidine is treated equivalently to thymidine in a cellular context since it base-pairs efficiently with adenosine and does not cause elevated point mutation rates. This is most likely due to its thymidine-like hydrogen bonding characteristics, since DNA-polymerases recognize mostly the shape of newly incorporated base-pairs [67]. 4sT thus constitutes a tool to incorporate a new chemical handle - a thiol group - into genomic DNA.

The cellular introduction of new functional groups in a biocompatible manner has led to great technological advances in the past as exemplified by the advent of unnatural amino acids [224]. In this context the usefulness of such an approach is governed in part by the chemical versatility of the introduced functional group. Here, thiols are among the most useful groups as a multitude of different reactions are available to functionalize these handles [209]. Thus, 4sT is a novel artificial nucleotide that allows functionalization of genomic DNA via introduction of a useful chemical group.

In the context of this work, I focused on a specific reaction that would allow sequencing based detection of 4sT incorporation. The goal was to exploit the biocompatibility of 4sT to label cellular DNA and then - after DNA extraction - to use the introduced functional handle to generate a sequencing signal. To this end, I employed OsO_4 -based chemistry to mediate oxidative hydrolysis of 4sT to 5-methyl-cytosine. This chemistry was initially developed for 4-thio-uracil [200] and we found - in collaboration with the group of Ronald Micura from the University of Innsbruck - that it extends readily to the DNA analog 4sT. The conversion of 4sT to 5-methyl-cytosine is very efficient and causes a point mutation upon sequencing of the modified position, thus generating a signal that marks the initial position of 4sT and allows quantitative detection of the incorporated label. We therefore have developed a system that allows incorporation of a synthetic nucleotide and a downstream approach that allows efficient, sequencing-based detection.

While 4sT is a very promising novel DNA label, there are limitations that may impact its usefulness in the future. The main issue that I encountered was that the incorporation density of 2.5 % of 4sT relative to thymidine is quite low. In particular, since many approaches will rely on classifying reads into a "labelled" and "unlabelled" population, the low incorporation density means that expensive long-read sequencing is necessary and that a lot of reads cannot be classified since the absence of mutations is not informative in such a context. Moreover, while 4sT is non-toxic at the concentrations tested, it does decrease

the rate of proliferation of HeLa cells. This could mean that 4sT has some effect on biological function that was not probed for here and could hamper certain applications. However, these limitations are minor and 4sT will be applicable in a wide range of different biological contexts.

4.1.1 Potential applications

In general, 4sT will be useful for every application, where identical DNA molecules need to be distinguished in a cellular context. For 4sT to be applicable, two prerequisites are necessary. First, the experimental set-up will need to include DNA replication prior to analysis since 4sT incorporation is dependent on successful replication. Second, the target cells need to be able to incorporate 4sT to a sufficient degree and should permit cell cycle synchronization, since this is necessary to collect cells in a homogeneous labelling state. If these prerequisites are met, 4sT can be used to distinguish labelled from unlabelled DNA, permitting a variety of methodological advancements.

One potential application of 4sT labelling could be the study of epigenetic inheritance [225]. This subject is concerned with elucidating how epigenetic state, defined as the collection of histone modifications and chromatin associated proteins, can be transmitted to daughter cells. A key issue here is that this epigenetic state needs to be restored after DNA is replicated, a process that is notoriously difficult to study with conventional methods [226, 227]. 4sT incorporation via DNA replication could allow to distinguish between replicated and unreplicated DNA, which could enable probing of epigenetic state via ChIP-seq and Cut-and-Run [228] before and after replication. These approaches could be extended to single-cell experiments, where individual replication forks could be observed and the dynamics of histone mark reestablishment investigated.

Another field that could benefit from distinction between identical DNA molecules via 4sT is the study of DNA replication. Here, marking newly replicated DNA via 4sT could allow mapping of replication timing with unprecedented signal-to-noise ratio and in a single cell context. Additionally, if other thiol-containing nucleotides, such as 6-thio-guanosine, could be used to generate a second type of point mutation, this could allow chromatin fiber assays to be performed via sequencing. In the classical version, CldU and IdU are incorporated into cellular DNA successively and detected via immunostaining on stretched DNA fibers [229]. These analyses allow replication fork direction and speed determination. A sequencing version of this assay would allow to map these metrics genome-wide, potentially elucidating the role of epigenetic state in this process.

In summary, 4sT in combination with OsO₄-based chemistry is a great tool to distinguish identical DNA molecules via replication based incorporation. Within this work the most relevant application was in the context of distinguishing sister chromatids for chromatin structure analysis.

4.2 scsHi-C: Sister chromatid sensitive chromatin structure analysis

The structural analysis of sister chromatids has been a significant hurdle in the past due to the fact that they are chemically identical and thus share the same sequence. This is because most techniques that probe the 3D-structure of chromatin are dependent on differences in sequence either because they are reliant on sequencing in the case of Hi-C [21] or on hybridization of detection probes in the case of FISH [37, 127]. In order to solve this problem and investigate sister chromatid specific chromatin structure, these identical DNA sequences need to be made distinct. 4sT is an ideal tool to achieve this goal since sister chromatids can easily incorporate it via DNA replication and can thus be distinguished in a sequencing based experiment.

4.2.1 Strand-specific labelling allows high-yield, low-noise sister chromatid Hi-C

The most natural approach to achieve sister chromatid specific labelling would have been to adopt the synchronization scheme from Nagasaka et al. [158] (Fig. 3.1a): In this approach, cells would have been labelled for 10 days to achieve almost complete labelling with 4sT, followed by two subsequent DNA replications in the absence of 4sT to obtain cells that contain one sister chromatid with one 4sT-labelled strand and one sister chromatid that is unlabelled. However, this idea has several downsides. First, the prolonged labelling with 4sT might increase potential subtle negative effects of this artificial nucleotide on cellular function. Moreover, in this synchronization scheme only one of four DNA strands is labelled by 4sT, likely resulting in poor signal-to-noise characteristics.

In order to solve this problem, I decided to employ a different approach. My idea was to exploit the fact that DNA replication is semi-conservative [199], yielding one labelled strand per sister chromatid after one round of DNA replication in 4sT-containing medium. However, these sister chromatids are not identical, since they have incorporated the 4sT label on different DNA strands: One sister chromatid will have incorporated the label on the Watson strand and the other on the Crick strand. Since sequencing technologies can easily distinguish the strandedness of DNA fragments, this synchronization scheme does not only drastically reduce the time to generate samples, but also increases the amount of 4sT in the final material and thus the signal-to-noise ratio compared to the approach employed by Nagasaka et al. [158].

Signal-to-noise ratio is a particular concern for sister chromatid resolved Hi-C since trans-sister contacts may only constitute a small fraction of all contacts because sister chromatids are distinct polymers that potentially diffuse far apart. However, prior approaches to achieve sister chromatid discrimination suffer from extremely low signal-to-noise ratio. Specifically, a recent study uses BrdU to enable sequencing based discrimination of sister chromatids [230]. Here, the authors employ a similar labelling scheme as described above for 4sT with BrdU and then use the fact that BrdU containing DNA can be destroyed using UV-

light [231]. By sequencing the remaining DNA, the strandedness of the DNA fragments allows sister chromatid assignment. This approach has theoretical merit, but in practice exhibits a very high noise ground with up to 50 % of sister chromatid contacts constituting unspecific signal. This is most likely because the approach measures the remaining DNA after UV-treatment, but has no way of determining whether the DNA destruction was efficient or how likely a given read constitutes a true sister chromatid contact. Here, the scsHi-C approach using 4sT is at an advantage since it measures the presence of a signal - the 4sT induced point-mutation - and can decide on a threshold of point-mutations above which to consider a read trustworthy. This results in lower overall yield, but the reads that are retained are of high quality and informative about sister chromatid specific chromatin structure. scsHi-C thus achieves very high signal-to-noise ratio based on all metrics used, both in the detection of 4sT labelled reads and the discrimination between cis-sister and trans-sister contacts.

Of course the fact that this high signal-to-noise ratio is bought with a decrease of yield results in higher experimental costs and prohibits certain types of analyses that are dependent on very high resolution such as probing promoter-enhancer contacts or gene-specific chromatin structure.

4.2.2 Principle of strand-specific labelling is transferable

The principle for sister chromatid specific chromatin structure analyses described in 4.2.1 is not limited to 4sT, but can potentially be transferred to other artificial nucleotides. Here, two prerequisites need to be fulfilled: First, the nucleotide should allow efficient incorporation into genomic DNA without perturbing cellular function. Second, a way of detecting the artificial nucleotide using sequencing based approaches needs to be available. If those conditions are met a sister chromatid specific labelling scheme for chromatin structure analysis can be implemented.

In this context, a particularly attractive candidate could be BrdU. It has been established in the molecular biology community for years that BrdU can be incorporated into cellular DNA without significantly perturbing cellular function [232, 233]. Moreover, mass spec experiments that I performed suggest that BrdU can be incorporated to up to 80 % relative to thymine, which is more than 30-fold higher than 4sT. Additionally, BrdU can be easily detected via nanopore sequencing by training a classifier to distinguish current traces from thymidine and BrdU [234]. While the number of reads that can be obtained from a nanopore instrument is still an order of magnitude lower than from comparable Illumina devices, Hi-C has been successfully performed using nanopore sequencing by measuring higher-order concatamers of chromatin contacts rather than pairwise contacts [235]. Using this trick, a single nanopore read can be expanded into 10s of Hi-C contacts. Since the high labelling density of BrdU will likely allow sister chromatid identity assignment of every nanopore read, this approach could enable high-resolution sister chromatid specific chromatin structure analysis, potentially elucidating promoter-enhancer architecture across sister chromatids.

4.3 The role of sister chromatid alignment in G2

The only biological system where the degree of alignment of similar chromosomes has been probed is in the context of homolog pairing in *D. melanogaster* [195, 236]. Here, Hi-C was performed leveraging single nucleotide polymorphisms to distinguish between homologous chromosomes. The authors found that in this system the homologous chromosomes are tightly aligned on large stretches of the genome in a so-called "railroad" configuration. Interestingly, scsHi-C analysis of G2 synchronized HeLa cells showed that sister chromatids also exhibit extensive alignment, but at a much larger scale: Sister chromatids are locally separated and intermix only at distances above 3 Mb. This finding suggests that sister chromatids do not exhibit a "railroad"-like alignment of local chromatin features, but are rather loosely held together. Nonetheless, the genome-wide proximity of sister chromatids in G2 is remarkable and highlights an important conundrum.

Based on the current understanding of sister chromatid cohesion, it is only required during mitosis: Here, it mediates mechanical resistance for alignment of chromosomes at the metaphase plate and holds sister chromatids together until ordered segregation at the metaphase-to-anaphase transition [67, Chapter 17]. However, at this time in the cell-cycle of human cells, most cohesive cohesin is removed from chromosomal arms by the prophase pathway and all the mentioned functions are performed by cohesive cohesin at the centromere [98]. So what is the role of genome-wide cohesion and sister chromatid alignment during G2, when most of it is removed again during Prophase?

4.3.1 Sister chromatid alignment and DNA damage repair

When cellular DNA is subject to a double strand break, the only way to repair this lesion again without causing potential mutations is homology directed DNA repair (HDR). Here, the lost information is copied from the other sister chromatid via resection of DNA surrounding the break site, homology search and re-synthesis of DNA using the other sister chromatid as a template [175]. In order for this process to work, the same genomic loci on both sister chromatids need to come into close proximity. Given the enormous size of the human genome, this is a daunting task and global alignment of sister chromatids during S-Phase and G2 might constitute the biological solution to this problem.

The connection between sister chromatid cohesion and DNA damage repair has been studied extensively in *S. cerevisiae*: Here, it was shown that perturbations of cohesin in diploid cells leads to HDR using the homolog chromosome as a template more frequently [150]. Moreover, in different studies, perturbation of cohesin led to decreased error-free repair of stalled replication forks, a defect that could be rescued by artificially tethering sister chromatids together [146, 147]. It is therefore conceivable that the proximity between sister chromatids influences the choice of the DNA repair pathway. Interestingly, other studies showed that certain genomic contexts favor DSB repair via NHEJ over HDR [176, 177]. It is tempting to speculate that local differences in sister chromatid

proximity might contribute to these differences. To sum up, there is evidence to speculate that sister chromatid alignment in G2 in human cells might have evolved to allow efficient HDR of DNA breaks, a claim that will be very interesting to probe in future.

4.4 Implications of TAD-level organization

Close inspection of G2 scsHi-C contact maps suggests that trans-sister contacts are not uniformly distributed along the genome, but rather exhibit striking patterns such as filled domains and focal accumulations at the main diagonal. Interestingly, by visual inspection it seems that these patterns are bounded by cis-sister TADs: The overall intensity of trans-sister contacts changes on a TAD-level and cis-sister TADs bound the genomic extent of increased trans-sister signal before decaying sharply. If true, this notion would suggest that there may be similar processes that shape cis-sister and trans-sister contacts, but affecting within sister and between sister organization in a different way.

One of the main visual hallmarks of trans-sister contact maps are the focal accumulation of signal at the diagonal that coincides with the boundaries of TADs. Genome-wide analysis confirms this and suggests that trans-sister contacts are enriched roughly 2-fold at TAD-boundaries. Along the lines of the argument presented in 4.3.1, this might result in increased choice of error-free repair pathways of DSBs due to the increased inter-sister proximity at these sites. Interestingly, TAD-boundaries are subject to increased rates of translocations and topoisomerase II occupancy [237–239], suggesting a high degree of DNA damage. This might be because of torsional stress that accumulates at these sites due to the loop extrusion action of cohesin. It is therefore possible that the increased interaction of sister chromatids at TAD boundaries might help to avoid elevated mutation rates by increasing the rate of error-free repair pathways.

4.4.1 Consequence for gene expression

Previous reports have suggested that regulatory interactions between promoters and enhancers often happen within the confines of TADs as genes occurring in the same TAD are often co-regulated [134]. This behavior has a plausible explanation when I take into account what is known about mechanisms that shape TADs. Based on modeling and perturbation studies, TADs are most likely the result of cohesin mediated loop extrusion, which is confined by CTCF boundaries that both constitute a roadblock for cohesin and stabilize its binding to DNA [32, 124, 129]. The signature of TADs visible in Hi-C maps is thought to arise through observing a population of cells that contains multiple cohesin proteins that are *en route* of extrusion, capturing different intermediate states [32]. It is easy to imagine how such a collection of small loops might facilitate the interaction between TAD-internal sequences and their collective boundary - marked by CTCF - might prohibit interactions with external regions. Indeed,

CTCF has been described before the advent of Hi-C as an insulator element that is capable of blocking promoter-enhancer interactions [118].

However, the same mechanism cannot work across sister chromatids since loop extrusion is a cis-acting process. But we know that regulatory interactions between different chromatids are possible as has been described extensively in *D. melanogaster* in the context of transvection using genetic means and in cultured cells using transgenes that were asymmetrically inserted into homologous chromosomes [84, 132]. It is therefore conceivable that enhancer-promoter interactions might cross insulating boundaries across sister chromatids. This could have dramatic consequences since aberrant activation of genes for example at a wrong developmental time-point could cause severe issues. But especially cells in developing organisms cycle very rapidly, spending a large fraction of their time in a state with two sister chromatids. It therefore is likely that organisms have evolved a mechanism to avoid such aberrant activation.

The increased trans-sister interaction at TAD-boundaries might constitute such a mechanism: If sister chromatids are "clamped" together and aligned at TAD-boundaries, this will decrease interactions that cross TAD-boundaries between sister chromatids. Such a mechanism could therefore contribute to avoiding aberrant regulatory interactions between elements on different sister chromatids and thus enable precise gene expression control across the cell-cycle.

4.5 Functional role of trans-sister contact heterogeneity

As described in more detail in 4.4, cis-sister TADs are units of trans-sister organization. However, these units are very heterogeneous with some TADs exhibiting strong trans-sister interaction while others are practically devoid of trans-sister signal. Interestingly, some TADs show asymmetry with regards to their trans-sister signal: In some cases there is high trans-sister interaction frequency on one border, which diminishes gradually throughout the domain, whereas in other cases there are dominant stripes on one border of the TAD, but not on the other (Fig. 4.1a,b). This heterogeneity poses the question of what mechanism is responsible for it and what functional implications it might have.

4.5.1 Mechanisms that determine trans-sister TAD state

The first obvious correlation that I found was between TAD size and trans-sister contact frequency. Here, small domains tend to contain many trans-sister contacts whereas large domains contain very little. It is interesting to speculate that loop extrusion as a mechanism for TAD formation might be the cause of this. As described more in depth in 4.4.1, TADs are often thought of as an average snapshot of a collection of loops that are in the process of being extruded. If we assume this model, smaller TADs will consist of smaller loops that arise within a stretch of DNA that is "clamped" together across sister chromatids (see 4.4). This will naturally cause smaller separation between sister chromatids as the maximum possible distance is bounded by twice the

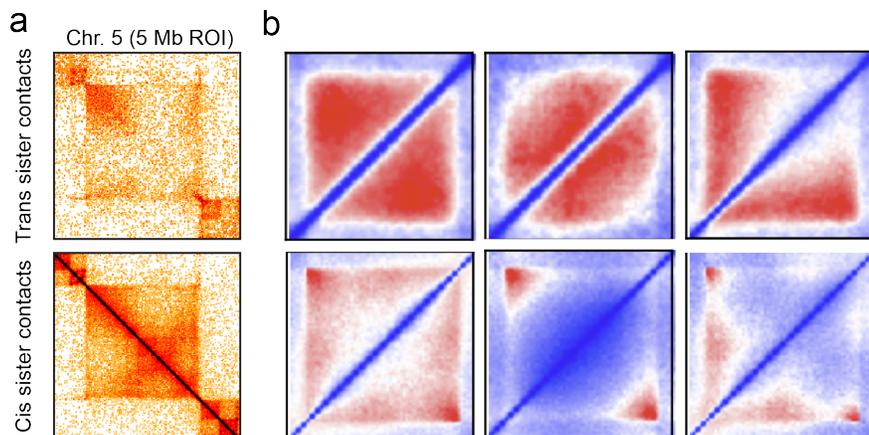


Figure 4.1: **Heterogeneity of trans-sister contacts.** (a) Example region on Chromosome 5 of scsHi-C data of HeLa Kyoto cells synchronized to G2. While cis-sister contacts are relatively uniform, trans-sister contacts show complex patterns such as gradients, depletion and focal enrichment. (b) Average classes of trans-sister contacts within domains. Domains were called based on foreground enrichment and clustered to reveal common behavior (see 2.21.17 for details).

largest loop size (one loop on each sister chromatid). While this mechanism may contribute to the different trans-sister behavior of TADs it cannot explain the asymmetry within TADs that I observed.

I reasoned that different chromatin state and its implication on chromatin structure might contribute to trans-sister contact behavior. And indeed, correlation analyses and predictive models show that trimethylation of lysine 27 on histone H3 (H3K27me3) as well as the methyltransferase that catalyzes this modification are strong influencers of trans-sister interactions. Interestingly, based on feature importance analyses of my predictive model, H3K27me3 is a much stronger predictor of trans-sister interactions than TAD-size, suggesting that epigenetic state may be the main contributor to determining the trans-sister TAD behavior.

4.5.2 Functional implications

The importance of H3K27me3 in predicting trans-sister interactions is very interesting since polycomb proteins, which are associated with this modification, have been linked to trans-homolog effects in *D. melanogaster* [84]. Indeed, a large fraction of polycomb proteins have been discovered based on a phenomenon called pairing-sensitive silencing. This phenomenon refers to enhancing gene silencing of two homologous genes across homologous chromosomes and is dependent on polycomb-response elements (PREs), the binding sites of polycomb

proteins [83]. This suggests that polycomb proteins might mediate both physical and functional interactions across homologous chromosomes in *D. melanogaster*.

It is interesting to speculate that polycomb proteins might assume a similar role in mediating synergistic silencing across sister chromatids. In particular in the context of gene dosage compensation: During DNA replication, each regulatory element and each promoter are duplicated, yielding potentially twice the amount of transcripts. In processes where exact levels of a protein are required, such as for signaling molecules and their receptors during development, it is therefore vital to not increase protein levels two-fold. Increased silencing of repressed domains across sister chromatids mediated by polycomb proteins might be a mechanism to achieve this. It will be therefore very interesting to probe this hypothesis with targeted perturbations and gene expression measurements in the future.

4.6 Sororin is necessary for sister chromatid alignment in G2

Sororin is necessary for sister chromatid cohesion in mammalian cells and has been shown to be constantly required as degradation in G2 - after cohesion establishment - also leads to defects in mitosis [170, 171]. However, thus far no study has looked at the influence of sororin on chromatin structure, partly because we did not have any suitable method to probe sister chromatid specific chromatin architecture. scsHi-C solves this problem and suggests that sororin is required for the extensive sister chromatid alignment observed in WT G2 HeLa Kyoto cells. This opens up new possibilities to study the effect of sister chromatid alignment on various biological functions since alignment is completely lost in this perturbation condition.

4.6.1 Role of sororin in gene expression regulation and HDR

One such biological function might be the synchronization of gene expression state across sister chromatids. As mentioned in 4.5.2, sister chromatid alignment in the context of polycomb domains might be necessary for synergistic repression in the context of gene dosage compensation. Interestingly, these trans-sister interaction domains do not exist in sororin-degraded G2 cells, suggesting a dependence on both sororin and H3K27me3. If this is true, I would predict a genome-wide change in gene expression state upon sororin degradation in G2 synchronized cells, potentially unraveling a new role of sororin besides mediating cohesion in mitosis. It will be therefore very interesting to investigate this hypothesis by performing gene expression measurements in sororin degraded cells.

Another potential role for sororin might be in the context of DNA repair. As mentioned in 4.3.1, the alignment of sister chromatids in G2 might predominantly exist to increase availability of the second sister chromatid as a DNA repair template. Loss of sister chromatid alignment might therefore switch the preferred repair pathway of DSBs in G2 away from HDR and force cells to use

mutagenic alternatives such as NHEJ. This model might provide a mechanistic explanation of the involvement of cohesive cohesin in DSB repair and why cohesin degradation in *S. cerevisiae* causes increased NHEJ dependent DSB repair [148, 149]. Sororin degradation is thus an important tool that will allow investigation of the role of sister chromatid alignment in both gene expression synchronization and DSB repair.

4.7 Influence of loop extrusion on trans-sister contacts

Loop extrusion - the process of successively forming loops on a DNA template - has been proposed as a mechanism that would allow sister chromatid separation in the context of mitosis [30]. Here, the prime candidate to perform such a function *in vivo* is condensin since it is necessary for chromatin compaction and chromosome individualization and has been shown to be able to perform loop extrusion *in vitro* [88, 110, 115]. However, there is also a loop extrusion factor active during interphase - the SMC protein cohesin - that has been shown to be required for TAD and loop formation, most likely through its loop extrusion activity [41, 102]. I reasoned that if loop extrusion is capable of sister chromatid separation that this then should also happen in interphase. I performed NIPBL degradation on G2 synchronized HeLa Kyoto cells to test this hypothesis. This perturbation is particularly suited to study the effect of interphase loop extrusion on sister chromatid separation since it has a dual prohibitive function: First, NIPBL was shown to be an essential cofactor for cohesin mediated loop extrusion *in vitro* [97, 115] and its loss is therefore expected to abrogate this function *in vivo*. Second, NIPBL has been described as a loading factor of cohesin [96] and thus its perturbation in combination with prolonged incubation will result in Wapl mediated unloading of transiently bound cohesin from chromatin, while cohesive cohesin is protected in these conditions.

scsHi-C experiments under these conditions showed that NIPBL degradation caused a massive increase of sister chromatid proximity and alignment, abrogating the local separation that was observed in WT G2 cells. These results suggest that loop extrusion is capable of separating sister chromatids not only in mitosis, but also in interphase and constitute the first experimental evidence that loop extrusion causes sister separation.

Interestingly, not only local separation of sister chromatids was lost, but also accumulation of trans-sister contacts at TAD-boundaries. An attractive mechanism for this could be that loop extrusion mediated by transiently bound cohesin is capable of pushing the long residence time cohesive cohesin towards TAD boundaries, where it mediates trans-sister interactions. It is not known whether cohesive cohesin can be translocated along chromosomes, but prior work in mouse embryonal fibroblasts (MEFs) suggests that cohesin can in principle be moved by motors acting on chromatin such as RNA polymerase II [240]. Such a mechanism could potentially also explain why H3K27me3 positive facultative heterochromatin exhibits enriched trans-sister interactions: Dense chromatin could pose an obstacle to cohesin mediated loop extrusion, which would cause cohesive cohesin to enrich there. In this context, ChIP-seq experiments probing

sororin occupancy in the presence and absence of NIPBL will be very informative in the future and could elucidate key mechanisms that shape sister chromatid topology.

While loop extrusion is a likely cause of local sister chromatid separation in G2, the question about its biological function remains. Specifically, if sister chromatid alignment is necessary for DNA repair as postulated in 4.3.1, why do sisters exhibit local separation in the size range of TADs? One potential role might be to avoid interference with gene expression regulation: By separating TAD internal regions, units of gene expression could avoid trans-sister interactions. This might be to prohibit synergistic activation of actively transcribed regions, which could result in dramatically increased dosage of regulatory proteins. Dense trans-sister TADs that exhibit trans-sister silencing as proposed in 4.5.2 would then be the other side of this principle, where repression requires synergy. To sum up, it is clear that trans-sister topology is complex and needs to accommodate many different biological needs, thus constituting an interesting future research topic.

4.8 Conclusion and outlook

In this thesis, I presented a novel synthetic nucleotide that allows sequencing based detection using a novel conversion chemistry. Furthermore, I showed how to use this nucleotide in the context of sister chromatid sensitive Hi-C to enable sister chromatid discrimination in the analysis of chromosome conformation. I used this novel method to elucidate the conformation of replicated chromosomes in G2 HeLa cells and was able to explain some aspects of it using perturbations of cohesin regulators. scsHi-C is thus a valuable tool that will allow quantification of sister chromatid alignment in a variety of contexts in the future.

One such aspect will be the study of sister chromatid resolution in mitosis. Prior studies have looked at the evolution of sister chromatid resolution from mitotic entry until metaphase and gained important insights [158]. However, this studies used a microscopy based read-out of labelled chromatids and thus only represent bulk measurements, which are limited in resolution and cannot distinguish different genomic locations. Here, scsHi-C could provide a much deeper understanding of this process, allowing unprecedented resolution and discrimination between different loci. This could allow elucidating the role of epigenetic state in resolution dynamics and the potential interplay between sister chromatid resolution and cohesion.

Furthermore, scsHi-C will enable the study of homology directed DNA repair in greater detail than before. Specifically, measurement of sister chromatid distance will allow for the first time to visualize how cells search for homologous regions in the context of HDR and permit to probe the involvement of different protein factors. Moreover, correlation between sister chromatid topology and DNA repair choice pathway could help to explain why NHEJ is the preferred mechanism for some genomic contexts and may allow exploitation of mutagenic repair pathway choice in the context of cancer therapy.

5 References

- [1] Eric S. Lander et al. “Initial sequencing and analysis of the human genome”. In: *Nature* 409.6822 (Feb. 2001).
- [2] J. Craig Venter et al. “The sequence of the human genome”. In: *Science* 291.5507 (Feb. 2001).
- [3] Yasuo Tsunaka et al. “Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle”. In: *Nucleic Acids Research* 33.10 (June 2005).
- [4] Isabel Garcia-Saez et al. “Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation”. In: *Molecular Cell* 72.5 (Dec. 2018).
- [5] Karolin Luger et al. “Crystal structure of the nucleosome core particle at 2.8 Å resolution”. In: *Nature* 389.6648 (1997).
- [6] Gina Arents and Evangelos N. Moudrianakis. “The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization”. In: *Proceedings of the National Academy of Sciences of the United States of America* 92.24 (Nov. 1995).
- [7] James R. Davie. “Covalent modifications of histones: Expression from chromatin templates”. In: *Current Opinion in Genetics and Development* 8.2 (Apr. 1998).
- [8] Roger D. Kornberg. “Structure of Chromatin”. In: *Annual Review of Biochemistry* 46.1 (June 1977).
- [9] C. L.F. Woodcock, J. P. Safer and J. E. Stanchfield. “Structural repeating units in chromatin. I. Evidence for their general occurrence”. In: *Experimental Cell Research* 97.1 (Jan. 1976).
- [10] Robert T. Simpson, Fritz Thoma and Joel M. Brubaker. “Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: A model system for study of higher order structure”. In: *Cell* 42.3 (Oct. 1985).
- [11] Patricia M. Schwarz et al. “Reversible oligonucleosome self-association: Dependence on divalent cations and core histone tail domains”. In: *Biochemistry* 35.13 (Apr. 1996).
- [12] Chenyi Wu, Andrew Bassett and Andrew Travers. “A variable topology for the 30-nm chromatin fibre”. In: *EMBO Reports* 8.12 (Dec. 2007).
- [13] Feng Song et al. “Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units”. In: *Science* 344.6182 (Apr. 2014).
- [14] Thomas Schalch et al. “X-ray structure of a tetranucleosome and its implications for the chromatin fibre”. In: *Nature* 436.7047 (July 2005).
- [15] Benedetta Dorigo et al. “Nucleosome arrays reveal the two-start organization of the chromatin fiber”. In: *Science* 306.5701 (Nov. 2004).

- [16] Yasumasa Joti et al. “Chromosomes without a 30-nm chromatin fiber”. In: *Nucleus (United States)* 3.5 (2012).
- [17] Mikhail Eltsov et al. “Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.50 (Dec. 2008).
- [18] Kazuhiro Maeshima et al. “Nucleosomal arrays self-assemble into supra-molecular globular structures lacking 30-nm fibers”. In: *EMBO* 35.10 (May 2016).
- [19] Natashe Kireeva et al. “Visualization of early chromosome condensation: A hierarchical folding, axial glue model of chromosome structure”. In: *Journal of Cell Biology* 166.6 (Sept. 2004).
- [20] Quentin Szabo, Frédéric Bantignies and Giacomo Cavalli. *Principles of genome folding into topologically associating domains*. Apr. 2019.
- [21] Erez Lieberman-Aiden et al. “Comprehensive mapping of long-range interactions reveals folding principles of the human genome.” In: *Science* 326.5950 (2009).
- [22] Suhas S.P. Rao et al. “A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping”. In: *Cell* 159.7 (Dec. 2014).
- [23] Suhas S.P. Rao et al. “Cohesin Loss Eliminates All Loop Domains”. In: *Cell* 171.2 (2017).
- [24] Jesse R. Dixon et al. “Topological domains in mammalian genomes identified by analysis of chromatin interactions”. In: *Nature* (2012).
- [25] Wibke Schwarzer et al. “Two independent modes of chromatin organization revealed by cohesin removal”. In: *Nature* 551.7678 (Nov. 2017).
- [26] Tsung Han S. Hsieh et al. “Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C”. In: *Cell* (2015).
- [27] Natalia Naumova et al. “Organization of the mitotic chromosome.” In: *Science* 342.6161 (2013).
- [28] Johan H. Gibcus et al. “A pathway for mitotic chromosome formation”. In: *Science* 359.6376 (2018).
- [29] Natalia Sikorska and Tom Sexton. *Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated*. Feb. 2020.
- [30] Anton Goloborodko et al. “Compaction and segregation of sister chromatids via active loop extrusion”. In: *eLife* 5 (2016).
- [31] Anders S. Hansen et al. “CTCF and cohesin regulate chromatin loop stability with distinct dynamics”. In: *eLife* 6 (May 2017).
- [32] Geoffrey Fudenberg et al. “Formation of Chromosomal Domains by Loop Extrusion”. In: *Cell Reports* 15.9 (May 2016).

- [33] Marie Zufferey et al. “Comparison of computational methods for the identification of topologically associating domains”. In: *Genome Biology* 19.1 (Dec. 2018).
- [34] Emily Crane et al. “Condensin-driven remodelling of X chromosome topology during dosage compensation”. In: *Nature* 523.7559 (July 2015).
- [35] Lin An et al. “OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries”. In: *Genome Biology* 20.1 (Dec. 2019).
- [36] Jesse R. Dixon et al. “Chromatin architecture reorganization during stem cell differentiation”. In: *Nature* 518.7539 (Feb. 2015).
- [37] Bogdan Bintu et al. “Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells”. In: *Science* 362.6413 (2018).
- [38] Martin Falk et al. “Heterochromatin drives compartmentalization of inverted and conventional nuclei”. In: *Nature* 570.7761 (2019).
- [39] Johannes Nuebler et al. “Chromatin organization by an interplay of loop extrusion and compartmental segregation.” In: *Proceedings of the National Academy of Sciences of the United States of America* (July 2018).
- [40] Jonathan A. Beagan and Jennifer E. Phillips-Cremins. “On the existence and functionality of topologically associating domains”. In: *Nature Genetics* 52.1 (Jan. 2020).
- [41] Gordana Wutz et al. “Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins.” In: *The EMBO journal* 36.24 (Dec. 2017).
- [42] Thomas Cremer and Marion Cremer. “Chromosome territories.” In: *Cold Spring Harbor perspectives in biology* (2010).
- [43] Luis A. Parada, Philip G. McQueen and Tom Misteli. “Tissue-specific spatial organization of genomes.” In: *Genome biology* (2004).
- [44] A Grosberg et al. “Crumpled globule model of the three-dimensional structure of dna”. In: *EPL* 23.5 (1993).
- [45] Leonid A. Mirny. “The fractal globule as a model of chromatin architecture in the cell”. In: *Chromosome Research* 19.1 (2011).
- [46] Brian D. Strahl and C. David Allis. “The language of covalent histone modifications”. In: *Nature* 403.6765 (Jan. 2000).
- [47] T. Jenuwein and C. D. Allis. “Translating the histone code”. In: *Science* 293.5532 (Aug. 2001).
- [48] G Vidali, E L Gershey and V G Allfrey. “Chemical Studies of Histone Acetylation”. In: *Journal of Biological Chemistry* 243.24 (1968).
- [49] Min Hao Kuo and C. David Allis. “Roles of histone acetyltransferases and deacetylases in gene regulation”. In: *BioEssays* 20.8 (Aug. 1998).

- [50] Qinming Chen et al. “Regulation of Nucleosome Stacking and Chromatin Compaction by the Histone H4 N-Terminal Tail–H2A Acidic Patch Interaction”. In: *Journal of Molecular Biology* 429.13 (June 2017).
- [51] Roberto Sanchez and Ming Ming Zhou. *The role of human bromodomains in chromatin biology and gene transcription*. Sept. 2009.
- [52] Weiqun Shen et al. “Solution structure of human Brg1 bromodomain and its specific binding to acetylated histone tails”. In: *Biochemistry* 46.8 (Feb. 2007).
- [53] Eric L. Greer and Yang Shi. “Histone methylation: A dynamic mark in health, disease and inheritance”. In: *Nature Reviews Genetics* 13.5 (May 2012).
- [54] Kwangbeom Hyun et al. “Writing, erasing and reading histone lysine methylations”. In: *Experimental and Molecular Medicine* 49.4 (Apr. 2017).
- [55] David O. Jones, Ian G. Cowell and Prim B. Singh. “Mammalian chromodomain proteins: Their role in genome organisation and expression”. In: *BioEssays* 22.2 (Feb. 2000).
- [56] Maria Ninova, Katalin Fejes Tóth and Alexei A. Aravin. “The control of gene expression and cell identity by H3K9 trimethylation”. In: *Development (Cambridge)* 146.19 (Oct. 2019).
- [57] Helena Santos-Rosa et al. “Active genes are tri-methylated at K4 of histone H3”. In: *Nature* 419.6905 (Sept. 2002).
- [58] Michiel Vermeulen et al. “Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4”. In: *Cell* 131.1 (Oct. 2007).
- [59] Bryan M. Turner. “Cellular memory and the histone code”. In: *Cell* 111.3 (Nov. 2002).
- [60] Elisheva Javasky et al. “Study of mitotic chromatin supports a model of bookmarking by histone modifications and reveals nucleosome deposition patterns”. In: *Genome Research* 28.10 (Oct. 2018).
- [61] Joshua C. Black, Capucine Van Rechem and Johnathan R. Whetstine. “Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact”. In: *Molecular Cell* 48.4 (Nov. 2012).
- [62] Nicole J. Francis, Robert E. Kingston and Christopher L. Woodcock. “Chromatin compaction by a polycomb group protein complex”. In: *Science* 306.5701 (Nov. 2004).
- [63] Aaron J. Plys et al. “Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2”. In: *Genes and Development* 33.13-14 (July 2019).
- [64] Adam G. Larson et al. “Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin”. In: *Nature* (2017).
- [65] Bryan A Gibson et al. “Organization of Chromatin by Intrinsic and Regulated Phase Separation”. In: *Cell* 179 (2019).

- [66] Su Qin, Li Li and Jinrong Min. “The Chromodomain of Polycomb: Methylation Reader and Beyond”. In: *Polycomb Group Proteins*. Elsevier Inc., Jan. 2017.
- [67] Bruce Alberts et al. *Molecular Biology of the Cell*. 2017.
- [68] P H Lewis. “New mutants report”. In: *Drosophila Information Service* 21 (1947).
- [69] J. Simon, A. Chiang and W. Bender. “Ten different Polycomb group genes are required for spatial control of the *abdA* and *AbdB* homeotic products”. In: *Development* 114.2 (1992).
- [70] Divya Tej Sowpati, Senthilkumar Ramamoorthy and Rakesh K. Mishra. “Expansion of the polycomb system and evolution of complexity”. In: *Mechanisms of Development* 138 (Nov. 2015).
- [71] Shahram Golbabapour et al. “Gene silencing and polycomb group proteins: An overview of their structure, mechanisms and phylogenetics”. In: *OMICS A Journal of Integrative Biology* 17.6 (June 2013).
- [72] Hengbin Wang et al. “Role of histone H2A ubiquitination in Polycomb silencing”. In: *Nature* 431.7010 (Oct. 2004).
- [73] Ru Cao et al. “Role of histone H3 lysine 27 methylation in polycomb-group silencing”. In: *Science* 298.5595 (Nov. 2002).
- [74] Yuri B. Schwartz. “Cooperative Recruitment of Polycomb Complexes by Polycomb Response Elements”. In: *Polycomb Group Proteins*. Elsevier Inc., Jan. 2017.
- [75] Anne Laugesen, Jonas Westergaard Højfeldt and Kristian Helin. “Molecular Cell Review Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation”. In: *Molecular Cell* 74 (2019).
- [76] Wolfgang Fischle et al. “Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains”. In: *Genes and Development* 17.15 (Aug. 2003).
- [77] Wenlai Zhou et al. “Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation”. In: *Molecular Cell* 29.1 (Jan. 2008).
- [78] Ragnhild Eskeland et al. “Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination”. In: *Molecular Cell* 38.3 (May 2010).
- [79] Raphael Margueron et al. “Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms”. In: *Molecular Cell* 32.4 (Nov. 2008).
- [80] Chiara Lanzuolo et al. “Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex”. In: *Nature Cell Biology* 9.10 (Oct. 2007).

- [81] Maria A Ferraiuolo et al. “The three-dimensional architecture of Hox cluster silencing”. In: *Nucleic Acids Research* 38.21 (2010).
- [82] Charlotte Grimaud et al. “RNAi components are required for nuclear clustering of polycomb group response elements”. In: *Cell* 124.5 (Mar. 2006).
- [83] Judith Akassis. “Pairing-sensitive silencing, polycomb group response elements, and transposon homing in *Drosophila*”. In: *Advances in Genetics* 46 (Jan. 2002).
- [84] Frédéric Bantignies et al. “Inheritance of polycomb-dependent chromosomal interactions in *Drosophila*”. In: *Genes and Development* 17.19 (Oct. 2003).
- [85] James D P Rhodes et al. “Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells Correspondence”. In: *CellReports* 30 (2020).
- [86] Kobe C. Yuen and Jennifer L. Gerton. “Taking cohesin and condensin in context”. In: *PLOS Genetics* 14.1 (Jan. 2018). Ed. by Mónica P. Colaiácovo.
- [87] Frank Uhlmann. “SMC complexes: From DNA to chromosomes”. In: *Nature Reviews Molecular Cell Biology* 17.7 (July 2016).
- [88] Tatsuya Hirano, Ryuji Kobayashi and Michiko Hirano. “Condensins, chromosome condensation protein complexes containing XCAP- C, XCAP-E and a *Xenopus* homolog of the *Drosophila* Barren protein”. In: *Cell* 89.4 (May 1997).
- [89] Christine Michaelis and Rafal Ciosk. “Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids”. In: *Cell* 91 (1997).
- [90] David E. Anderson et al. “Condensin and cohesin display different arm conformations with characteristic hinge angles”. In: *Journal of Cell Biology* 156.3 (Feb. 2002).
- [91] Christian H. Haering et al. “Molecular architecture of SMC proteins and the yeast cohesin complex”. In: *Molecular Cell* 9.4 (2002).
- [92] Zhubing Shi et al. “Cryo-EM structure of the human cohesin-NIPBL-DNA complex”. In: *Science* 368.6498 (June 2020).
- [93] Byung Gil Lee et al. “Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism”. In: *Nature Structural and Molecular Biology* (July 2020).
- [94] Takao Ono et al. “Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells”. In: *Cell* 115.1 (2003).
- [95] Jonathan N. Wells et al. “Evolution of condensin and cohesin complexes driven by replacement of Kite by Hawk proteins”. In: *Current Biology* 27.1 (Jan. 2017).

- [96] Rafal Ciosk et al. “Cohesin’s Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins”. In: *Molecular Cell* 5.2 (Feb. 2000).
- [97] Iain F. Davidson et al. “DNA loop extrusion by human cohesin”. In: *Science* 366.6471 (Dec. 2019).
- [98] Irene C Waizenegger et al. “Two Distinct Pathways Remove Mammalian Cohesin from Chromosome Arms in Prophase and from Centromeres in Anaphase”. In: *Cell* 103.3 (Oct. 2000).
- [99] Stephanie Kueng et al. “Wapl Controls the Dynamic Association of Cohesin with Chromatin”. In: *Cell* 127.5 (2006).
- [100] Vincent Guacci, Douglas Koshland and Alexander Strunnikov. “A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*”. In: *Cell* 91.1 (Oct. 1997).
- [101] Kerstin S. Wendt et al. “Cohesin mediates transcriptional insulation by CCCTC-binding factor”. In: *Nature* 451.7180 (Feb. 2008).
- [102] Johanna Gassler et al. “A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture.” In: *The EMBO journal* 36.24 (Dec. 2017).
- [103] Prakash Arumugam et al. “ATP Hydrolysis Is Required for Cohesin’s Association with Chromosomes”. In: *Current Biology* 13.22 (Nov. 2003).
- [104] W. C. Earnshaw and U. K. Laemmli. “Architecture of metaphase chromosomes and chromosome scaffolds”. In: *Journal of Cell Biology* 96.1 (1983).
- [105] John F. Marko. “Micromechanical studies of mitotic chromosomes”. In: *Chromosome Research* 16.3 (May 2008).
- [106] H. G. Keyl. “Lampbrush chromosomes in spermatocytes of *Chironomus*”. In: *Chromosoma* 51.1 (Mar. 1975).
- [107] Franz Klein et al. “A Central Role for Cohesins in Sister Chromatid Cohesion, Formation of Axial Elements, and Recombination during Yeast Meiosis Genetic and biochemical analyses have identified”. In: *Cell* 98 (1999).
- [108] Kim Nasmyth. “Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis”. In: *Annual Review of Genetics* 35.1 (Dec. 2001).
- [109] Anton Goloborodko, John Marko and Leonid Mirny. “Mitotic chromosome compaction via active loop extrusion”. In: *bioRxiv* (June 2015).
- [110] Iain F Davidson et al. “Rapid movement and transcriptional re-localization of human cohesin on DNA”. In: *The EMBO Journal* 35.24 (2016).
- [111] Jorine M Eeftens et al. “Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism”. In: *The EMBO Journal* 36.23 (Dec. 2017).

- [112] Johannes Stigler et al. “Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin”. In: *Cell Reports* 15.5 (May 2016).
- [113] Mahipal Ganji et al. “Real-time imaging of DNA loop extrusion by condensin”. In: *Science* (Feb. 2018).
- [114] Stefan Golfier et al. “Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner”. In: *eLife* 9 (May 2020).
- [115] Yoori Kim et al. “Human cohesin compacts DNA by loop extrusion”. In: *Science* 366.6471 (Dec. 2019).
- [116] Edward J Banigan et al. “Chromosome organization by one-sided and two-sided loop extrusion”. In: *bioRxiv* (Apr. 2020).
- [117] Chin Tong Ong and Victor G. Corces. “CTCF: An architectural protein bridging genome topology and function”. In: *Nature Reviews Genetics* 15.4 (Mar. 2014).
- [118] Adam C. Bell, Adam G. West and Gary Felsenfeld. “The protein CTCF is required for the enhancer blocking activity of vertebrate insulators”. In: *Cell* 98.3 (Aug. 1999).
- [119] Chandrasekhar Kanduri et al. “Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive”. In: *Current Biology* 10.14 (July 2000).
- [120] Varun Narendra et al. “CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation”. In: *Science* 347.6225 (Feb. 2015).
- [121] Gordana Wutz et al. “ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinstag1 from WAPL”. In: *eLife* 9 (Feb. 2020).
- [122] Ya Guo et al. “CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function”. In: *Cell* 162.4 (Aug. 2015).
- [123] Anders S. Hansen. “CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism”. In: *Nucleus* 11.1 (Jan. 2020).
- [124] Yan Li et al. “The structural basis for cohesin–CTCF-anchored loops”. In: *Nature* 578.7795 (Feb. 2020).
- [125] E Heitz. *Das Heterochromatin der Moose*. Bornträger, 1928.
- [126] S C Elgin, S C R Elgin and J L Workman. *Chromatin Structure and Gene Expression*. Chromatin Structure and Gene Expression. Oxford University Press, 2000.
- [127] Alistair N. Boettiger et al. “Super-resolution imaging reveals distinct chromatin folding for different epigenetic states”. In: *Nature* (2016).

- [128] J. E. Brownell et al. “Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation”. In: *Cell* 84.6 (Mar. 1996).
- [129] Nils Krietenstein et al. “Ultrastructural Details of Mammalian Chromosome Architecture”. In: *Molecular Cell* 78.3 (May 2020).
- [130] Tsung Han S. Hsieh et al. “Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding”. In: *Molecular Cell* 78.3 (May 2020).
- [131] Robert A. Rollins, Patrick Morcillo and Dale Dorsett. “Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes”. In: *Genetics* 152.2 (June 1999).
- [132] Bomyi Lim et al. “Visualization of Transvection in Living Drosophila Embryos”. In: *Molecular Cell* 70.2 (Apr. 2018).
- [133] Eileen E.M. Furlong and Michael Levine. “Developmental enhancers and chromosome topology”. In: *Science* 361.6409 (Sept. 2018).
- [134] Elphège P. Nora et al. “Spatial partitioning of the regulatory landscape of the X-inactivation centre”. In: *Nature* 485.7398 (May 2012).
- [135] Orsolya Symmons et al. “The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances”. In: *Developmental Cell* 39.5 (Dec. 2016).
- [136] Iain Williamson et al. “Developmentally regulated Shh expression is robust to TAD perturbations”. In: *Development (Cambridge)* 146.19 (Oct. 2019).
- [137] Raymond R. Tice and Richard B. Setlow. “DNA Repair and Replication in Aging Organisms and Cells”. In: *Handbook of the Biology of Aging* (1985).
- [138] James E. Haber. “DNA recombination: The replication connection”. In: *Trends in Biochemical Sciences* 24.7 (July 1999).
- [139] Yumin Teng, Yachuan Yu and Raymond Waters. “The *Saccharomyces cerevisiae* histone acetyltransferase Gcn5 has a role in the photoreactivation and nucleotide excision repair of UV-induced cyclobutane pyrimidine dimers in the MFA2 gene”. In: *Journal of Molecular Biology* 316.3 (Feb. 2002).
- [140] Yachuan Yu et al. “UV irradiation stimulates histone acetylation and chromatin remodeling at a repressed yeast locus”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.24 (June 2005).
- [141] Kiyoe Ura et al. “ATP-dependent chromatin remodeling facilitates nucleotide excision repair of UV-induced DNA lesions in synthetic dinucleosomes”. In: *EMBO Journal* 20.8 (Apr. 2001).

- [142] Ryujiro Hara and Aziz Sancar. “The SWI/SNF Chromatin-Remodeling Factor Stimulates Repair by Human Excision Nuclease in the Mononucleosome Core Particle”. In: *Molecular and Cellular Biology* 22.19 (Oct. 2002).
- [143] Fena Ochs et al. “Stabilization of chromatin topology safeguards genome integrity”. In: *Nature* 574.7779 (Oct. 2019).
- [144] Ireneusz Litwin et al. “ Error-free DNA damage tolerance pathway is facilitated by the Irc5 translocase through cohesin ”. In: *The EMBO Journal* 37.18 (Sept. 2018).
- [145] Nan Wu et al. “Scc1 sumoylation by Mms21 promotes sister chromatid recombination through counteracting Wapl”. In: *Genes and Development* 26.13 (2012).
- [146] Mireille Tittel-Elmer et al. “Cohesin Association to Replication Sites Depends on Rad50 and Promotes Fork Restart”. In: *Molecular Cell* 48.1 (Oct. 2012).
- [147] Marco Fumasoni et al. “Error-Free DNA Damage Tolerance and Sister Chromatid Proximity during DNA Replication Rely on the Pol α /Primase/Ctf4 Complex”. In: *Molecular Cell* 57.5 (Mar. 2015).
- [148] Xiangduo Kong et al. “Distinct Functions of Human Cohesin-SA1 and Cohesin-SA2 in Double-Strand Break Repair”. In: *Molecular and Cellular Biology* 34.4 (2014).
- [149] Felipe Cortés-Ledesma and Andrés Aguilera. “Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange”. In: *EMBO Reports* 7.9 (Sept. 2006).
- [150] Shay Covo et al. “Cohesin is limiting for the suppression of DNA damage-induced recombination between homologous chromosomes”. In: *PLoS Genetics* 6.7 (July 2010).
- [151] Chris Norbury and Paul Nurse. “Animal Cell Cycles and Their Control”. In: *Annual Review of Biochemistry* 61.1 (June 1992).
- [152] M. A. Felix et al. “A post-ribosomal supernatant from activated *Xenopus* eggs that displays post-translationally regulated oscillation of its cdc2+ mitotic kinase activity”. In: *EMBO Journal* 8.10 (1989).
- [153] Marcos Malumbres. “Cyclin-dependent kinases”. In: *Genome Biology* 15.6 (June 2014).
- [154] Katherine Wickliffe et al. “The multiple layers of ubiquitin-dependent cell cycle control”. In: *Chemical Reviews* 109.4 (Apr. 2009).
- [155] Paul Batty and Daniel W. Gerlich. “Mitotic Chromosome Mechanics: How Cells Segregate Their Genome”. In: *Trends in Cell Biology* 29.9 (Sept. 2019).
- [156] J. Richard Mcintosh. “Mitosis”. In: *Cold Spring Harbor Perspectives in Biology* 8.9 (Sept. 2016).

- [157] Keishi Shintomi and Tatsuya Hirano. “Sister chromatid resolution: A cohesin releasing network and beyond”. In: *Chromosoma* 119.5 (Oct. 2010).
- [158] Kota Nagasaka et al. “Sister chromatid resolution is an intrinsic part of chromosome organization in prophase”. In: *Nature Cell Biology* 18.6 (2016).
- [159] Ewa Piskadlo, Alexandra Tavares and Raquel A Oliveira. “Metaphase chromosome structure is dynamically maintained by condensin I-directed DNA (de)catenation”. In: *eLife* 6 (May 2017).
- [160] Keiji Kimura et al. “Phosphorylation and activation of 13S condensin by Cdc2 in vitro”. In: *Science* 282.5388 (Oct. 1998).
- [161] Ana Losada, Michiko Hirano and Tatsuya Hirano. “Identification of Xenopus SMC protein complexes required for sister chromatid cohesion”. In: *Genes and Development* 12.13 (1998).
- [162] Daniel Gerlich et al. “Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication.” In: *Current biology* 16.15 (2006).
- [163] Erwan Watrin et al. “Human Scc4 Is Required for Cohesin Binding to Chromatin, Sister-Chromatid Cohesion, and Mitotic Progression”. In: *Current Biology* 16.9 (May 2006).
- [164] Kim Nasmyth and Christian H. Haering. “Cohesin: Its roles and mechanisms”. In: *Annual Review of Genetics* 43 (Dec. 2009).
- [165] Antonio Tedeschi et al. “Wapl is an essential regulator of chromatin structure and chromosome segregation”. In: *Nature* 501.7468 (Sept. 2013).
- [166] Robert V. Skibbens. “Establishment of Sister Chromatid Cohesion”. In: *Current Biology* 19.24 (Dec. 2009).
- [167] Dawn Bender et al. “Multivalent interaction of ESCO2 with the replication machinery is required for sister chromatid cohesion in vertebrates”. In: *Proceedings of the National Academy of Sciences of the United States of America* 117.2 (Jan. 2020).
- [168] James D P Rhodes et al. “Cohesin Can Remain Associated with Chromosomes during DNA Replication.” In: *Cell reports* 20.12 (Sept. 2017).
- [169] Tomoko Nishiyama et al. “Sororin mediates sister chromatid cohesion by antagonizing Wapl”. In: *Cell* 143.5 (2010).
- [170] Julia Schmitz et al. “Sororin Is Required for Stable Binding of Cohesin to Chromatin and for Sister Chromatid Cohesion in Interphase”. In: *Current Biology* 17.7 (Apr. 2007).
- [171] Rene Ladurner et al. “Sororin actively maintains sister chromatid cohesion”. In: *The EMBO Journal* 35.6 (2016).
- [172] Silke Hauf et al. “Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2”. In: *PLoS Biology* 3.3 (Mar. 2005).

- [173] Barry E McGuinness et al. “Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells”. In: *PLoS Biology* 3.3 (Mar. 2005). Ed. by R. Scott Hawley.
- [174] Jan Michael Peters. “The anaphase promoting complex/cyclosome: A machine designed to destroy”. In: *Nature Reviews Molecular Cell Biology* 7.9 (Sept. 2006).
- [175] Nicole Hustedt and Daniel Durocher. “The control of DNA repair by the cell cycle”. In: *Nature Cell Biology* 19.1 (Jan. 2017).
- [176] Ruben Schep et al. “Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance”. In: *bioRxiv* (May 2020).
- [177] Thomas Clouaire et al. “Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures”. In: *Molecular Cell* 72.2 (Oct. 2018).
- [178] Kazuyoshi Murata and Matthias Wolf. “Cryo-electron microscopy for structural analysis of dynamic biological macromolecules”. In: *Biochimica et Biophysica Acta* 1862.2 (Feb. 2018).
- [179] Horng D. Ou et al. “ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells”. In: *Science* 357.6349 (July 2017).
- [180] James O J Davies et al. “How best to identify chromosomal interactions: a comparison of approaches”. In: *Nature Methods* 14.2 (Feb. 2017).
- [181] Rugile Stanyte et al. “Dynamics of sister chromatid resolution during cell cycle progression.” In: *The Journal of cell biology* 217.6 (June 2018).
- [182] Michael Mitter et al. “Sister-chromatid-sensitive Hi-C reveals the conformation of replicated human chromosomes”. In: *bioRxiv* (Mar. 2020).
- [183] Aisha Yesbolatova et al. “Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole”. In: *Methods* 164-165 (July 2019).
- [184] Magdalena Morawska and Helle D. Ulrich. “An expanded tool kit for the auxin-inducible degron system in budding yeast”. In: *Yeast* 30.9 (Sept. 2013).
- [185] Matthias Samwer et al. “DNA Cross-Bridging Shapes a Single Nucleus from a Set of Mitotic Chromosomes”. In: *Cell* 170.5 (Aug. 2017).
- [186] Christoph Sommer et al. “A deep learning and novelty detection framework for rapid phenotyping in high-content screening”. In: *Molecular Biology of the Cell* 28.23 (Nov. 2017). Ed. by Charles Boone.
- [187] Michael Held et al. “CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging”. In: *Nature Methods* 7.9 (Sept. 2010).
- [188] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature Methods* 9.7 (July 2012).

- [189] Alexandra Lusser et al. “Thiouridine-to-Cytidine Conversion Sequencing (TUC-Seq) to Measure mRNA Transcription and Degradation Rates”. In: *Methods in molecular biology* 2062 (2020).
- [190] Veronika A Herzog et al. “Thiol-linked alkylation of RNA to assess expression dynamics”. In: *Nature Methods* 14.12 (Sept. 2017).
- [191] Nezar Abdennur and Leonid Mirny. “Cooler: scalable storage for Hi-C data and other genomically-labeled arrays”. In: *bioRxiv* (2019).
- [192] Maxim Imakaev et al. “Iterative correction of Hi-C data reveals hallmarks of chromosome organization”. In: *Nature Methods* 9.10 (2012).
- [193] Nathan C. Sheffield and Christoph Bock. “LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor”. In: *Bioinformatics* 32.4 (2016).
- [194] Tao Yang et al. “HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient”. In: *Genome Research* 27.11 (Nov. 2017).
- [195] Jelena Erceg et al. “The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos”. In: *Nature Communications* 10.1 (Dec. 2019).
- [196] Jian Zhong et al. “Purification of nanogram-range immunoprecipitated DNA in ChIP-seq application”. In: *BMC Genomics* 18.1 (Dec. 2017).
- [197] The ENCODE Project Consortium. “An integrated encyclopedia of DNA elements in the human genome”. In: *Nature* 489.7414 (Sept. 2012).
- [198] R. S. Hansen et al. “Sequencing newly replicated DNA reveals widespread plasticity in human replication timing”. In: *Proceedings of the National Academy of Sciences* (2010).
- [199] M Meselson and F W Stahl. “The Replication of DNA in *Escherichia coli*.” In: *Proceedings of the National Academy of Sciences of the United States of America* 44.7 (July 1958).
- [200] Christian Riml et al. “Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing.” eng. In: *Angewandte Chemie (International ed. in English)* 56.43 (Oct. 2017).
- [201] Matthias Muhar et al. “SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis”. In: *Science* 360.6390 (May 2018).
- [202] Andrew Massey, Yao Zhong Xu and Peter Karran. “Photoactivation of DNA thiobases as a potential novel therapeutic option”. In: *Current Biology* 11.14 (July 2001).
- [203] Andrew Massey, Yao Zhong Xu and Peter Karran. “Ambiguous coding is required for the lethal interaction between methylated DNA bases and DNA mismatch repair”. In: *DNA Repair* 1.4 (Apr. 2002).
- [204] Gertrude B. Elion. “The purine path to chemotherapy”. In: *Science* 244.4900 (1989).

- [205] Ann M. De Mazière, Willem J. Hage and Geertje A. Ubbels. “A method for staining of cell nuclei in *Xenopus laevis* embryos with cyanine dyes for whole-mount confocal laser scanning microscopy”. In: *Journal of Histochemistry and Cytochemistry* 44.4 (1996).
- [206] K. R. Hande. “Etoposide: Four decades of development of a topoisomerase II inhibitor”. In: *European Journal of Cancer* 34.10 (Sept. 1998).
- [207] G. Bjursell and P. Reichard. “Effects of thymidine on deoxyribonucleoside triphosphate pools and deoxyribonucleic acid synthesis in Chinese hamster ovary cells.” In: *Journal of Biological Chemistry* 248.11 (June 1973).
- [208] L. J. Mah, A. El-Osta and T. C. Karagiannis. “ γ h2AX: A sensitive molecular marker of DNA damage and repair”. In: *Leukemia* 24.4 (Feb. 2010).
- [209] Charles E. Hoyle, Andrew B. Lowe and Christopher N. Bowman. “Thiol-click chemistry: A multifaceted toolbox for small molecule and polymer synthesis”. In: *Chemical Society Reviews* 39.4 (Mar. 2010).
- [210] Anna Bębenek and Izabela Ziuzia-Graczyk. *Fidelity of DNA replication—a matter of proofreading*. Oct. 2018.
- [211] David Mittelman and John H. Wilson. “The fractured genome of HeLa cells”. In: *Genome Biology* 14.4 (Apr. 2013).
- [212] Guo Chen and Xingming Deng. “Cell Synchronization by Double Thymidine Block”. In: *BIO-PROTOCOL* 8.17 (2018).
- [213] Andrey G. Baranovskiy et al. “Structural basis for inhibition of DNA replication by aphidicolin”. In: *Nucleic Acids Research* 42.22 (Dec. 2014).
- [214] L. T. Vassilev et al. “Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1”. In: *Proceedings of the National Academy of Sciences* 103.28 (2006).
- [215] Maxwell R Mumbach et al. “HiChIP: efficient and sensitive analysis of protein-directed genome architecture”. In: *Nature Methods* 13.11 (Nov. 2016).
- [216] Takashi Nagano et al. “Cell-cycle dynamics of chromosomal organization at single-cell resolution”. In: *Nature* 547.7661 (July 2017).
- [217] Aravind Subramanian et al. “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43 (Oct. 2005).
- [218] Tin Kam Ho. “Random decision forests”. In: *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. Vol. 1. IEEE Computer Society, 1995.
- [219] Leo Breiman. “Random forests”. In: *Machine Learning* 45.1 (Oct. 2001).

- [220] Susannah Rankin, Nagi G. Ayad and Marc W. Kirschner. “Sororin, a Substrate of the Anaphase- Promoting Complex, Is Required for Sister Chromatid Cohesion in Vertebrates”. In: *Molecular Cell* 18.2 (Apr. 2005).
- [221] Kohei Nishimura et al. “An auxin-based degron system for the rapid depletion of proteins in nonplant cells”. In: *Nature Methods* 6.12 (Dec. 2009).
- [222] Adrian Salic, Jennifer C. Waters and Timothy J. Mitchison. “Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis”. In: *Cell* 118.5 (Sept. 2004).
- [223] Peter Karran and Natalie Attard. “Thiopurines in current medical practice: Molecular mechanisms and contributions to therapy-related cancer”. In: *Nature Reviews Cancer* 8.1 (Jan. 2008).
- [224] L. Wang et al. “Expanding the genetic code of Escherichia coli”. In: *Science* 292.5516 (Apr. 2001).
- [225] Francisco M. Iglesias and Pablo D. Cerdán. “Maintaining epigenetic inheritance during DNA replication in plants”. In: *Frontiers in Plant Science* 7.FEB2016 (Feb. 2016).
- [226] Varija N. Budhavarapu, Myrriah Chavez and Jessica K. Tyler. “How is epigenetic information maintained through DNA replication?” In: *Epigenetics and Chromatin* 6.1 (Oct. 2013).
- [227] Nataliya Petryk et al. “MCM2 promotes symmetric inheritance of modified histones during DNA replication”. In: *Science* 361.6409 (Sept. 2018).
- [228] Peter J. Skene and Steven Henikoff. “An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites”. In: *eLife* 6 (Jan. 2017).
- [229] Rebecca A. Frum, Sumitra Deb and Swati Palit Deb. “Use of the DNA Fiber Spreading Technique to Detect the Effects of Mutant p53 on DNA Replication”. In: *Methods in molecular biology (Clifton, N.J.)* Vol. 962. NIH Public Access, 2013.
- [230] Marlies E. Oomen et al. “Detecting chromatin interactions along and between sister chromatids with SisterC”. In: *bioRxiv* (Mar. 2020).
- [231] Ester Falconer et al. “Identification of sister chromatids by DNA template strand sequences”. In: *Nature* 463.7277 (Jan. 2010).
- [232] Howard G. Gratzner. “Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication”. In: *Science* 218.4571 (1982).
- [233] M. D. Bick and R. L. Davidson. “Total substitution of bromodeoxyuridine for thymidine in the DNA of a bromodeoxyuridine dependent cell line”. In: *Proceedings of the National Academy of Sciences of the United States of America* 71.5 (1974).

- [234] Carolin A. Müller et al. “Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads”. In: *Nature Methods* 16.5 (May 2019).
- [235] Netha Ulahannan et al. “Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure”. In: *bioRxiv* (Nov. 2019).
- [236] Jumana AlHaj Abed et al. “Highly structured homolog pairing reflects functional organization of the Drosophila genome”. In: *Nature Communications* 10.1 (Dec. 2019).
- [237] Andres Canela et al. “Genome Organization Drives Chromosome Fragility”. In: *Cell* 170.3 (2017).
- [238] Andres Canela et al. “Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity”. In: *Molecular Cell* 75.2 (July 2019).
- [239] Liis Uusküla-Reimand et al. “Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders”. In: *Genome Biology* 17.1 (Aug. 2016).
- [240] Georg A. Busslinger et al. “Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl”. In: *Nature* 544.7651 (Apr. 2017).

6 Supplementary Material

6.1 Abbreviations

4sT	4-Thio-Thymidine
APC/C	Anaphase Promoting Complex/Cyclosome
AID	Auxin-Inducible Degron
bp	Basepair
BrdU	Bromodeoxyuridine
BSA	Bovine Serum Albumin
CdK	Cyclin Dependent Kinase
CTCF	CCCTC Binding Factor
CPM	Counts Per Million
DAPI	4,6-Diamidino-2-Phenylindol
dCTP	Deoxycytidinetriphosphate
dGTP	Deoxyguanosinetriphosphate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
dsDNA	Double Stranded Dna
DTT	Dithiothreitol
dTTP	Deoxythymidinetriphosphate
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscopy
EtOH	Ethanol
EZH2	Enhancer of Zeste Homolog 2
FACS	Fluorescence Assisted Cell Sorting
F-ara-EdU	2'-Deoxy-2'-Fluoro-5-Ethynyluridine
FISH	Fluorescence In Situ Hybridization
FITC	Fluorescein
FSC	Forward Scatter
GFP	Green Fluorescent Protein
gRNA	Guide RNA
GTF	General Transcription Factors
HAT	Histone Acetyl Transferase
HAWK	Heat Repeats Associated With Kleisins
HDR	Homology Directed Repair
Hi-C	High-Throughput Chromosome Conformation Capture
HMT	Histone Methyl Transferase
HP1	Heterochromatin Protein 1
HPLC	High Performance Liquid Chromatography
HR	Homologous Recombination
IAA	Iodoacetamide
ICCF	Iteratively Corrected Contact Frequency

IF	Immunofluorescence
kb	Kilobase
LEF	Loop Extrusion Factor
LOLA	Locus Overlap Analysis
MAD	Median Absolute Deviation
Mb	Megabase
MEF	Mouse Embryonal Fibroblasts
MeOH	Methanol
NEB	Nuclear Envelope Breakdown
NHEJ	Non-Homologous End Joining
ORI	Origin Of Replication
PDB	Protein Data Bank
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PE	Phycoerythrin
PI	Propidium Iodide
PRC1/2	Polycomb Repressive Complex 1/2
PRE	Polycomb Response Element
PTM	Post-Translational Modification
RNA	Ribonucleic Acid
ROI	Region Of Interest
ROS	Reactive Oxygen Species
RT	Room Temperature
SAC	Spindle Assembly Checkpoint
SAXS	Small-Angle X-Ray Scattering
scsHi-C	Sister Chromatid Sensitive Hi-C
SMC	Structural Maintenance of Chromosomes
SNP	Single Nucleotide Polymorphism
SSC	Side Scatter
SSD	Single Strand Damage
TAD	Topologically Associating Domain
UV	Ultraviolet
WT	Wildtype

6.2 Supplementary Tables

Table 1: Read statistics of G2 wild-type samples

	Total	Unique (HQ)	Cis-sister	Trans-sister
Replicate 1	3.08E+08	9.54E+07	1.26E+07	6.56E+05
Replicate 2	3.55E+08	1.01E+08	1.29E+07	6.28E+05
Replicate 3	2.24E+08	1.22E+08	1.62E+07	6.84E+05
Replicate 4	3.57E+08	1.39E+08	1.72E+07	9.97E+05
Replicate 5	3.37E+08	1.33E+08	1.76E+07	9.56E+05
Replicate 6	3.32E+08	1.70E+08	2.16E+07	8.18E+05
Replicate 7	3.34E+08	1.56E+08	1.63E+07	7.16E+05
Replicate 8	3.44E+08	1.80E+08	2.26E+07	1.11E+06
Replicate 9	4.14E+08	1.97E+08	2.41E+07	1.26E+06
Replicate 10	5.36E+08	2.10E+08	1.42E+07	1.10E+06
Replicate 11	2.33E+08	1.09E+08	9.85E+06	7.40E+05
Pooled	3.77E+09	1.61E+09	1.85E+08	9.66E+06

Table 2: Read statistics of prometaphase wild-type samples

	Total	Unique (HQ)	Cis-sister	Trans-sister
Replicate 1	6.72E+07	4.37E+07	3.19E+06	2.99E+05
Replicate 2	8.75E+07	5.10E+07	2.64E+06	2.23E+05
Pooled	1.55E+08	9.47E+07	5.84E+06	5.23E+05

Table 3: Read statistics of G2 NIPBL-degraded samples

	Total	Unique (HQ)	Cis-sister	Trans-sister
Replicate 1	5.51E+08	2.93E+08	1.49E+07	8.86E+05
Replicate 2	4.22E+08	2.27E+08	1.28E+07	9.73E+05
Replicate 3	7.15E+07	3.53E+07	1.10E+06	1.09E+05
Replicate 4	8.77E+07	2.73E+07	1.98E+06	2.16E+05
Pooled	1.13E+09	5.82E+08	3.07E+07	2.18E+06

Table 4: Read statistics of G2 Sororin-degraded samples

	Total	Unique (HQ)	Cis-sister	Trans-sister
Replicate 1	3.83E+08	2.19E+08	2.63E+07	8.86E+05
Replicate 2	3.95E+08	1.98E+08	2.34E+07	9.73E+05
Replicate 3	1.73E+08	1.03E+08	5.68E+06	3.44E+05
Pooled	9.51E+08	5.21E+08	5.54E+07	2.20E+06

Table 5: Cell lines used²⁷

Name	Genotype	Plasmid used	Resistance marker
HeLa Kyoto Sororin-AID	Tir1-3xMyc-T2A-Puro (Lentiviral integration) EGFP-AID-Sororin	Tir1-3xMyc-T2A-Puro-Lentivirus	Puromycin
HeLa Kyoto Nipbl-AID	Tir1-3xMyc-T2A-Puro (Lentiviral integration) AID-GFP-NIPBL	Tir1-3xMyc-T2A-Puro-Lentivirus AID-GFP-NIPBL-HR	Puromycin
HeLa Kyoto H2B-mCherry	H2B-mCherry (Lentiviral integration)	Lenti-H2B-mCherry	Blasticidin

Table 6: Plasmids used²⁸

Name	Insert	Prokaryotic resistance marker	Eukaryotic resistance marker
pRRL	Tir1-3xMyc-T2A-Puro	Ampicillin	Puromycin
AID-GFP-NIPBL-HR	AID-GFP-NIPBL Repair template	Ampicillin	-
Lenti-H2B-mCherry	H2B-mCherry	Ampicillin	Blasticidin
EGFP-AID-Sororin-HR	EGFP-AID-Sororin Repair template	Ampicillin	-

²⁷Table adapted from Table S6 of [182]²⁸Table adapted from Table S7 of [182]

Table 7: Chromatin features used for pairing score prediction

Data set	ID	Type	Info
CTCF	From JM Peters	ChIP-seq	G2 synchronized HeLa cells
TSS	GENCODE v32	Gene annotations	Transcriptional start sites
Genes	GENCODE v32	Gene annotations	Genes for hg19
SMC3	From JM Peters	ChIP-seq	G2 synchronized HeLa cells
Sororin	From JM Peters	ChIP-seq	G2 synchronized HeLa cells
H3K4me3	GSM733682	ChIP-seq	Bigwig and bed files
H3K9ac	GSM733756	ChIP-seq	Bigwig and bed files
EZH2	GSM1003520	ChIP-seq	Bigwig and bed files
H4K20me1	GSM733689	ChIP-seq	Bigwig and bed files
Chd2	GSM935432	ChIP-seq	Bigwig and bed files
H3K36me3	GSM733711	ChIP-seq	Bigwig and bed files
PolII	GSM733759	ChIP-seq	Bigwig and bed files
H3K79me2	GSM733669	ChIP-seq	Bigwig and bed files
H3K27me3	GSM733696	ChIP-seq	Bigwig and bed files
p300	GSM935500	ChIP-seq	Bigwig and bed files
H3K9me3	GSM1003480	ChIP-seq	Bigwig and bed files
Insulation	None	Insulation score	Called on G2 data set
Rep-timing	GSM923449	Replication Timing	Combined wiggle data set
Gro-seq	GSM1518913	Gro-seq data set	Gro-seq for HeLa cells
TADs	None	TAD calls	Location of TADs in G2 data set