From Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

NOVEL METHODS FOR MAPPING GENOME ORGANIZATION AND GENOME FRAGILITY IN THE 3D SPACE OF THE NUCLEUS

Tomasz Kallas

Stockholm 2020

Cover illustration: Pawel Kallas

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by US-AB

ISBN 978-91-8016-045-2

Novel methods for mapping genome organization and genome fragility in the 3D space of the nucleus

THESIS FOR DOCTORAL DEGREE (Ph.D.)

The thesis will be defended at Jacob Berzelius hall, Karolinska Institutet, Berzelius väg 3, Stockholm

December 4th, 2020, at 08:30 a.m.

By

Tomasz Kallas

Principal Supervisor:
Dr. Magda Bienko
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Genome Biology

Co-supervisor(s):
Professor Björn Högberg
Department of Medical Biochemistry and Biophysics
Division of Biomaterials

Opponent:
Associate Professor Andreas Lennartsson
Karolinska Institutet
Department of Biosciences and Nutrition

Examination Board:
Associate Professor Marianne Farnebo
Karolinska Institutet
Department of Cell and Molecular Biology

Professor Masood Kamali-Moghaddam
Uppsala Universitet
Department of Immunology, Genetics and Pathology

Associate Professor Theodoros Foukakis
Karolinska Institutet
Department of Oncology-Pathology

- To everyone who has supported me throughout this journey

ABSTRACT

For over a century, scientists have been trying to understand how the DNA molecule that is so tightly packed in the micro space of the nucleus sustains critical cellular processes such as transcription, replication or the maintenance of genetic information. Despite the huge effort of researchers around the world, we know relatively little. This is in part due to the lack of methods that bring the right throughput and resolution to the study of the higher-order spatial arrangement of the genome. For example, the radial arrangement of the chromatin in mammalian cells remains largely unrevealed as well as the processes that lead to genome instability, which could potentially lead to the development of cancer. It is therefore crucial to develop tools that would allow us to precisely map both the location and frequency of DNA double-strand breaks (DSBs) along the genome, as well as to study them in the right 3D context of the chromatin.

In order to fill this gap in, this thesis describes two methods which we have developed in order to map genome organization and genome fragility in the 3D space of the nucleus.

In **paper I**, we developed GPSeq (Genome Loci Positioning by Sequencing) as a genome-wide technique for mapping radial arrangement of the genome in mammalian cells. We showed that GPSeq accurately generates maps of the radial organization of the human genome at 1 Mb and 100 kb resolutions, thus allowing us to reveal unique radial patterns of various genetic and epigenomic traits, gene expression, A and B subcompartments as well as radial arrangements of DSBs, cancer mutations or germline variants.

In **paper II**, we developed BLISS (Breaks Labelling *In Situ* and Sequencing) as a genome-wide technique to quantitatively profile DSBs distribution in cells. We demonstrated that BLISS can be successfully applied to samples with either low number of cells or to tissue sections and yet accurately detect DSBs. We showed the sensitivity of BLISS by estimating off-target activity of two nucleases- Cas9 and Cpf1 in CRISPR system and demonstrated that Cpf1 is more specific when compared to Cas9.

LIST OF SCIENTIFIC PAPERS

THESIS PUBLICATIONS

I. Girelli G. *, Custodio J. *, **Kallas T.** *, Agostini F., Wernersson E., Spanjaard B., Mota A., Kolbeinsdottir S., Gelali E., Crosetto N., Bienko M. (2020) **GPSeq reveals the radial organization of chromatin in the cell nucleus.** *Nat. Biotechnol*. 38, 1184-1193.

II. Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., **Kallas, T.**, Custodio, J., Wernersson, E., Li, Y., Gao, L., Federova, Y., Zetsche, B., Zhang, F., Bienko, M., Crosetto, N. (2017). **BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks.** *Nat. Commun*. 8, 15058.

OTHER PUBLICATIONS

Mirzazadeh R., **Kallas T.**, Bienko M., Crosetto N. (2018). **Genome-Wide Profiling of DNA Double-Strand Breaks by the BLESS and BLISS Methods**. In: Muzi-Falconi M., Brown G. (eds) Genome Instability. *Methods Mol Biol*, vol 1672. Humana Press, New York, NY.

* These authors contributed equally to the work.

1 INTRODUCTION TO 3D GENOME ORGANIZATION IN EUKARYOTIC CELLS 1

1.1 Packaging of DNA inside the nucleus 1
1.1..1 Folding of the chromatin fiber 1
1.1..2 Hetero- vs. eu-chromatin 2
1.1..3 TADs and sub-TADs 3
1.1..4 Long-range contacts 4
1.1..5 Compartmentalization of the genome 5
1.1..6 Chromosome territories 6

1.2 Dynamics of genome organization and chromosome repositioning throughout the cell cycle 7

1.3 Fragility of the genome 9
1.3..1 Chromatin dynamics during DSBs repair 11

1.4 Methods for mapping the 3D genome 12
1.4..1 DNA-FISH 12
1.4..2 Hi-C 13
1.4..3 Dip-C 14
1.4..4 SPRITE 15
1.4..5 DamID 16
1.4..6 GAM 17
1.4..7 TSA-seq 18
1.4..8 ATAC-seq 19
1.4..9 ChIP-seq 20
1.4..10 Methodological consideration 21

2 DOCTORAL THESIS 23

2.1 Aims of the study 23

2.2 Methods 24
2.2..1 Cells and tissue 24
2.2..2 Sample preparation for GPSeq, YFISH and BLISS 24
2.2..3 Workflow for GPSeq and BLISS 25
2.2..4 3D DNA FISH, immunofluorescence and imaging 27
2.2..5 GPSeq, YFISH and BLISS adapters 28
2.2..6 Sequencing and data processing 28

2.3 Summary of research papers 29
2.3..1 Paper I: GPSeq reveals the radial organization of chromatin in the cell nucleus 29
2.3..2 Paper II: BLISS is a versatile and quantitative method to map DNA breaks in low-input cell and tissue samples 38

3 DISCUSSION AND CONCLUSIONS 44

4 ACKNOWLEDGEMENTS 46

5 REFERENCES 51

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
gDNA	Genomic DNA
RNA	Ribonucleic acid
3C	Chromosome conformation capture
4C	Circular chromosome conformation capture
5C	Chromosome conformation capture carbon copy
CT	Chromosomal territory
CTCF	CCCTC-binding factor
DSB	Double strand break
FISH	Fluorescence *in situ* hybridization
GPSeq	Genomic loci Positioning by Sequencing
IVT	*In vitro* transcription
LAD	Lamina-associated domain
PCR	Polymerase chain reaction
PFA	Para-formaldehyde
TAD	Topologically associated domain
TSS	Transcription start site
UMI	Unique molecular identifier
YFISH	Y- Fluorescence *in situ* Hybridization where the adapter is Y-shaped
BLISS	Breaks labelling *in situ* and sequencing
DSB	DNA double-strand break
TOPII	Type II topoisomerase
Cas9	CRISPR associated protein 9
Cpf1	CRISPR from *Prevotella* and *Francisella* 1
γH2AX	Histone H2A, variant X, phosphorylated at Ser139
NHEJ	Non-homologous end joining
HR	Homologous end joining
BLESS	Breaks labelling enriched on streptavidin and sequencing

END-seq	DSB ends sequencing
GUIDEseq	Genome -wide unbiased identification of DSBs enabled by sequencing
Digenome-seq	Cas9-digested whole genome sequencing
Hi-C	High-throughput chromosome conformation capture
Dip-C	Diploid chromatin conformation capture
SPRITE	Split-pool recognition of interactions by tag extension
DamID	DNA adenine methyltransferease identification
GAM	Genome architecture mapping
TSA-seq	Tyramide signal amplification sequencing
ATAC-seq	Assay of transposase accessible chromatin sequencing
ChIP-seq	Chromatin immunoprecipitation sequencing

1 INTRODUCTION TO 3D GENOME ORGANIZATION IN EUKARYOTIC CELLS

Since more than a hundred years, the fascinating question of how two meters of DNA can fit into the cell nucleus, and how this influences key nuclear processes such as transcription, has intrigued scientists. In this review, I will present an overview of the studies that have addressed this fundamental question and of the main insights they led to.

1.1 PACKAGING OF DNA INSIDE THE NUCLEUS

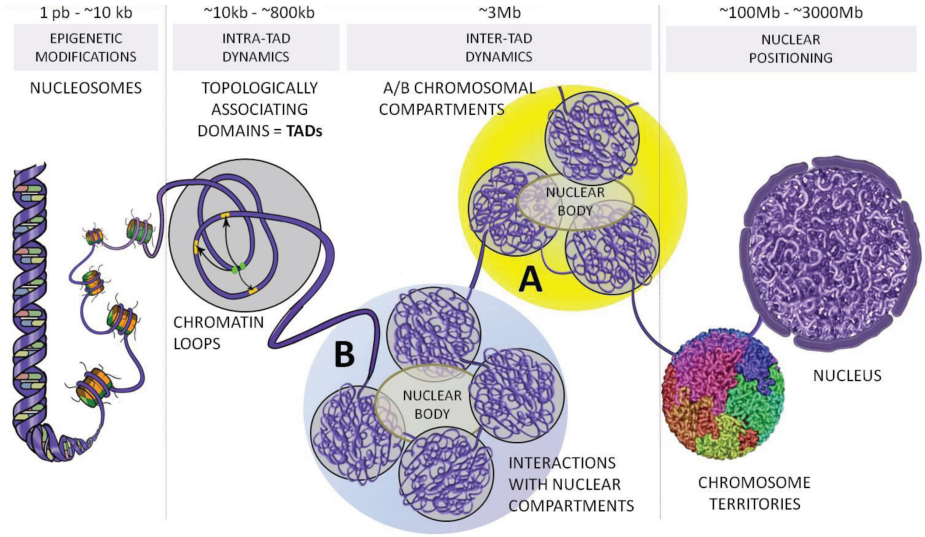

Fig. 1. Schematic representation of hierarchical genome organization in eukaryotic cells. Image adapted from Ea et al., Genes, 2015.

The human genome consists of around 20,000 protein-coding genes hidden in 3.2 billion base pairs (A+T, C+G) of deoxyribonucleic acid (DNA) that is spatially organized in a hierarchical fashion in order to fit in a nuclear volume of roughly 1000 μm^3 (**fig. 1.**) [1]. At the smallest level, the DNA is wrapped around histone proteins to form nucleosomes, and is then folded into chromatin fibers, chromosome domains, compartments and finally the chromosome.

1.1..1 Folding of the chromatin fiber

At the smallest scale, 146 bp of DNA is rolled around an octamer of histones (H2A, H2B, H3 and H4) to form a nucleosome. In 1976, Finch and Klug observed by electron microscopy that the chromatin isolated from a rat liver is organized into supercoils or solenoids (**fig. 2A**), measuring around 33 nm in diameter [2]. Many years later, in 2005, Schalch and colleagues

looking at DNA organization in chicken erythrocytes, reported unique interactions among singular nucleosomes that steer the folding of so-called nucleosomal array (the primary structure of chromatin) into a secondary structure referred to as the 30 nm fiber, and then further into larger-scale configurations or tertiary structures to eventually build a whole chromosome [3]. Based on these observations, they proposed a chromatin folding model, the so-called zigzag model, in which two adjacent rows of two nucleosomes with the linker DNA segments (joining DNA) intersect each other (**fig. 2B.**). On the other hand, the group led by Prof. Maeshima showed that human mitotic chromosomes are mostly built of unevenly folded nucleosome fibers rather than 30 nm chromatin fibers, as no regular structure above 11 nm was detected [4]. Another compelling and more recent observation was reported by the group of Prof. O'Shea [5]. According to their study, in interphase and in mitotic intact human cells, chromatin appeared rather flexible and arranged into 5 to 24 nm fibers. This is the first study reporting a chromatin structure in intact human cells, indicating that chromatin is folded into smaller and more heterogenous structures than the 30 nm fiber.

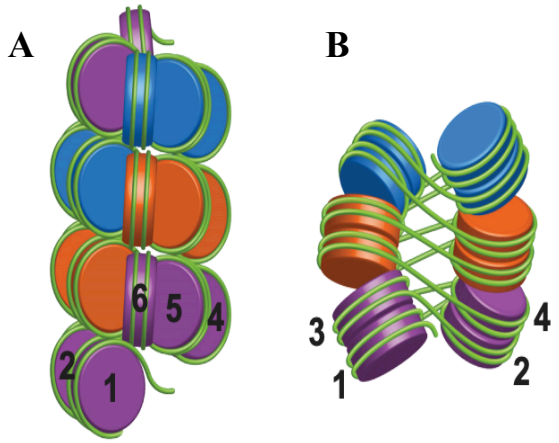

Fig. 2. Schematic representation of two chromatin folding models: (**A**) solenoid with hexanucleosomal unit structure and (**B**) zigzag with tetranucleosomal structure unit. Image adapted from Hsieh et al, Molecular Cell, 2020.

1.1..2 Hetero- vs. eu-chromatin

Already back in 1928, studying nuclei of a moss, Emil Heitz identified chromosomal regions that do not follow post-mitotic decondensation, and he defined these regions as heterochromatin [6,7]. In contrast, he termed euchromatin the regions that do decondense and diffuse in the interphase nucleus. Furthermore, he observed similar densely stained heterochromatin in *Drosophila* cells (using acetocarmine or fuchsin dyes), and later showed that *Drosophila* heterochromatin has a low gene density. Subsequently, Dr. Hsu reported that heterochromatin localizes towards the periphery of the nucleus, whereas euchromatin occupies more central regions [8]. Based on that observation, he proposed that the role of heterochromatin is to serve as a 'bodyguard'- a protector of the euchromatin from harming agents such as mutagens or viruses. Gilbert and colleagues further reported that the peripheral, compact heterochromatin is associated with a low density of genes, which are mostly inactive [9]. Heterochromatin can be further subdivided into two types: constitutive and facultative. This division is based on two histone modifications, namely methylation of histone H3 at lysine 9 (H3K9me) and methylation of histone H3 at lysine 27 (meH3K27me), that are associated with

gene silencing. Constitutive heterochromatin comprises highly repetitive DNA elements, such as pericentromeric DNA, and contains relatively high amounts of meH3K9 [10], whereas facultative heterochromatin consists of specific sequences of single-copy genes and is associated with H3K27me [11]. The latter type of heterochromatin can be either condensed or decondensed, depending on the cell cycle phase, and is compatible with gene expression, whereas constitutive heterochromatin is condensed throughout the cell cycle and is not transcribed. Both heterochromatin types are preferentially found at the nuclear periphery and around nucleoli [10–12]. Conversely, open euchromatin is characterized by a high density of transcriptionally active genes.

Exceptions to this chromatin arrangement were observed in different cell types. In retinal rod cells of nocturnal mammals, heterochromatin was found to be concentrated in the central parts of the nucleus, while euchromatin was shown to be located at the nuclear periphery [13]. It has been proposed that these nuclei with inverted structure serve as microlenses allowing nocturnal animals to see in a low-light environment. A similar observation has been reported in human neutrophils, where the accessible euchromatin is located at the nuclear periphery, to be released to fight pathogens in a process called NETosis (neutrophil extracellular trap) [14]. At last, a similar phenomenon has been described in mouse olfactory neurons, where facultative heterochromatin co-localizes with constitutive heterochromatin in the central part of the nucleus, and plays an important role in regulating exclusive expression of single odorant receptors genes [15].

Worth mentioning, there are several key players such as membrane proteins- Lamin A, B, C, and Lamin receptors- that tether heterochromatin to the nuclear periphery and around nucleoli [12].

1.1..3 TADs and sub-TADs

Genome-wide contact frequency matrices such as the ones generated by Hi-C, a DNA proximity ligation assay followed by genome-wide sequencing, revealed that, in interphase cells, chromatin folds into topologically associating domains (TADs) or contact domains [16–18]. TADs range in size from 100 kb to 5 Mb (with an average size of 1 Mb), and are defined as linear stretches of genome that have an elevated contact frequency when compared to the contact frequency with regions across their boundaries. The boundaries regions between TADs are found to be enriched in binding sites for architectural proteins such as the insulator binding protein CTCF (CCCTC-binding factor) and cohesin, suggesting a role for these regions in the formation of TADs [17–19]. Interestingly, many of the TADs boundaries are found to be conserved among species, as is the case, for example, between human and mouse ESCs [20]. TADs are subdivided into smaller domains, so-called sub-TADs, that are around 200-300 kb in size [21]. These sub-TADs were shown to be assembled before TADs that encompass them, being already formed in prometaphase or ana/telophase, whereas TADs start to be detected only in early G1 [22]. Although TADs are relatively conserved in a variety of cell types and organisms,

the folding that occurs inside of them is reported to be more variable [17–20]. Intra-TAD topology might be the outcome of cell-type specific transcription regulation, as TADs often hold tissue-specific genes and the enhancers associated with them. Therefore, TADs are currently considered to be the determinants of the range of enhancer functionality. Using single-cell Hi-C, it was shown that contacts within TADs are more similar and recurrent in different cells than contacts that cross TAD borders (inter-domain contacts) and those that occur between two different chromosomes (trans-chromosomal contacts). Furthermore, it was found that TADs containing active genes happen to be positioned at the surface of their chromosomal territories [21].

1.1..4 Long-range contacts

Genomic loci positioned tens of kilobases up to megabases away, can undergo long-range interactions. These contacts can be found either on the same chromosome (cis) or between two chromosomes (trans). One of the most well-studied case of long-range chromatin contacts is the interaction between enhancers and their target genes in the β-globin locus [23,24]. While examining the mouse and human β-globin loci during erythroid differentiation, it was found that the locus forms a looped configuration that arises from interactions between distal regulatory elements positioned on either end of the locus, and that eventually lead to the expression of β-globin genes. Ever since, long-range cis- and trans- interactions have been identified that regulate the activity of enhancers and their target promoters, the function of insulators, gene transcription, and imprinting and X-chromosome inactivation [23–27].

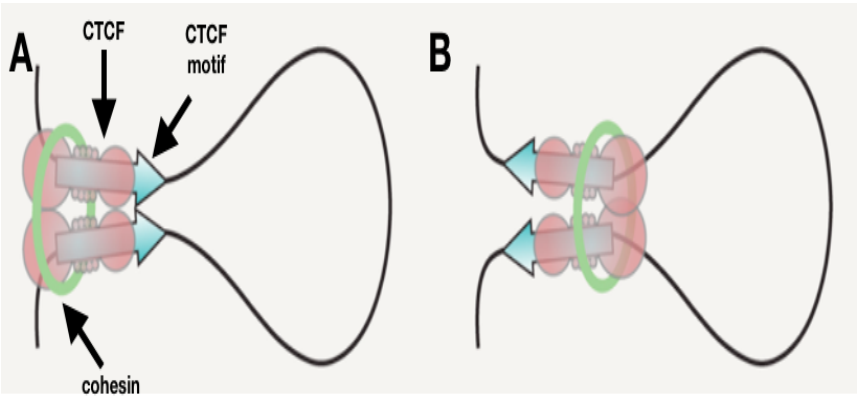

Fig. 3. Schematic representation of orientation-based CTCF loop formation. (**A**) Convergent orientation of CTCF two motifs. (**B**) Divergent orientation of CTCF two motifs. Image adapted from Nichols and Corces, Cell, 2015.

Sofueva and colleagues [28], showed that, in order to form a long-range chromatin loop, two key protein players are needed, namely CTCF and the ring-shaped cohesin complex, a member of the Structural Maintenance of Chromosomes (SMC) protein complex. In a study performed on wild-type and cohesin-deficient neural stem cells (NSCs), cohesin-deficient cells were found to form significantly fewer long-range contacts than wild-type cells. Similar results were obtained by Rao and colleagues, where they observed disappearing of loop domains after experimental cohesin degradation, whereas reintroducing cohesin recovered their formation [29]. The loss of cohesin had an impact on the global architecture of chromatin as well, leading to the relaxation of TAD domains (decrease in intra-domain contacts and increase in contacts with neighboring

domains) and overall chromatin decompaction. However, the loss of cohesin had a small effect on transcription with only a very few genes, mostly located around super-enhancers, being down-regulated. These observations indicate that CTCF alone is not enough to form long-range loops and that it needs to cooperate with functional cohesion [29]. Specifically, a model was proposed in which CTCF recruits cohesin to form a border-to-border loop between TAD boundaries [30]. Cohesin attaches to one CTCF that defines one of the TAD borders, and extrudes a loop until it reaches the second CTCF at the end of a boundary. The presence of two bound CTCF motifs is not sufficient, however, to initiate the formation of a loop. The orientation of these motifs plays a significant role in this process (**fig. 3.**). Specifically, it was shown that if two CTCF-bound motifs are oriented in a convergent manner (i.e., when they are pointing to one another on the linear genome) then the chromatin loop can be identified in 65% of all cases, while when the motifs are divergent or in the same orientation, respectively 1% and 34% of the CTCF-bound motifs form a loop [31]. Rao and colleagues found that 92% of all formed loops are between CTCF sites with convergent motifs, confirming prior observations [18,31]. Deletion of selected CTCF binding sites leads to disruption of chromatin loops [31]. A recent study, provided another insight into the loop formation [32]. By real-time imaging of loop extrusion, it was observed that a single condensin complex, another SMC family member, is able to pull out up to 1500 bp of DNA per second, using only a moderate amount of ATP that serves as fuel for the condensin motor function. Importantly, the loop extrusion was found to be asymmetric, meaning that the condensin can start reeling the DNA string only from one side, which is in alignment with the model described above [30].

1.1..5 Compartmentalization of the genome

Based on Hi-C experiments, the genome can be arbitrarily divided into two compartments referred to as A and B [33] . The A compartment contains genomic regions with high GC content that are enriched for genes, transcriptional activity, early replicating DNA, and that are overall considered to be active and open, as deduced from their accessibility to DNaseI digestion as well as the presence of active chromatin marks (H3K36me3, H3K79me2, H3K27ac and H3K4me1 H3K36me3 and H3K27ac3). On the other hand, genomic regions that make up the B compartment are generally gene-poor, contain late-replicating domains, are decorated with high levels of silencing marks (H3K9me3), and exhibit lower DNaseI accessibility. Together, these characteristics suggest that A compartments represent euchromatin and B compartments heterochromatin. There is a high correlation between the regions that compose B compartments and the regions of the genome that are known as lamina-associated domains (LADs), which have a high probability to reside in contact with the nuclear lamina, and are generally gene poor and devoid of gene activity. These observations indicate close proximity of the B compartment to the periphery of the nucleus, which is not observed for the A compartment [17]. Importantly, genomic regions belonging to the same compartment on the same chromosome interact with each more frequently than with regions in the opposite compartment [33]. With the higher resolution data, in 2014, Rao and colleagues partitioned further A and B compartments into six

subcompartments, namely A1, A2 and B1, B2, B3, B4 [18]. These subcompartments are enriched/marked with distinct genetic and epigenetic features. As A compartment, A1 and A2 subcompartments are gene rich, have highly expressed genes and active chromatin marks. Their presence is not observed neither at the nuclear lamina nor at the nucleolus-associated domains (NADs). The difference between these two subcompartments can be found while looking at certain features. For example, the time of replication for A1 is shorter than for A2, finishing at the beginning of S phase, while for A2 it continues until the middle of S phase. It was also shown that A2 compared to A1, is more enriched in H3K9me3, has lower GC content and consists of longer genes. On the other hand, B subcompartments exhibit similar patterns to B compartment. Unlike B2 and B3, B1 subcompartment is enriched with H3K27me3 and is depleted from H3K36me3, indicating facultative heterochromatin. B2, due to its enrichment in pericentromeric sequences, can be associated with constitutive heterochromatin. The replication timing also differs for these subcompartments. While the B1 type peaks at the middle of S phase, the B2 and B3 replicate by the end of the S phase. While both, B2 and B3 are present at the nuclear lamina, only B2 shows enrichment in NADs, with B3 being strongly depleted at these domains/regions. On the other hand, B4 subcompartment was found to be enriched in KRAB-ZNF superfamily genes that have been decorated with both active (H3K36me3) and inactive (H3K9me3, H4K20me3) histone marks [18] and is entirely located on chr19.

1.1..6 Chromosome territories

At the largest scale, human DNA is stored in the form of 23 pairs of chromosomes, ranging in size from about 40,000,000 to 250,000,000 base pairs. Each chromosome occupies a discrete space (1-2 µm), so-called chromosome territory (CT), in the cell nucleus during interphase (**fig. 4.**) [34]. Already in 1885, Carl Rabl proposed a model that stated that plant chromosomes retain an anaphase-like arrangement, occupying small and distinct areas throughout the cell cycle [35]. He reported that, in interphase, centromeres are clustered near one pole of the nucleus whereas telomeres are facing the opposite pole. This chromosome organization was from then on referred to as Rabl configuration, and was later also described in a variety of plant and animal cell nuclei [36]. Using single-cell Hi-C, it was shown in G1 phase of mouse embryonic stem cells (mESCs), chromosomes indeed adopt the Rabl configuration (**fig. 5.**) [37].

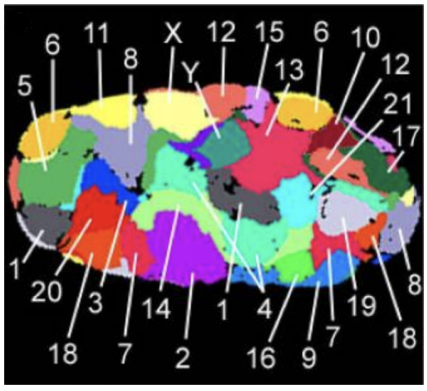

Fig. 4. CTs of all chromosomes in human fibroblasts nucleus. Image adapted from Bolzer et al., PLoS Biol, 2005.

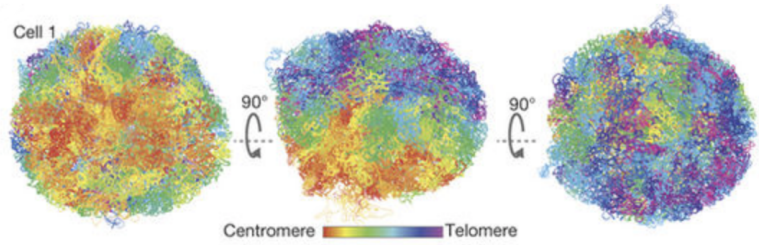

Fig. 5. Rabl configuration observed in mESCs, on a single-cell level in three different orientations, with chromosomes painted from red (centromeres) to purple (telomeres). Image adapted from Stevens et al., Nature 2017.

1.2 DYNAMICS OF GENOME ORGANIZATION AND CHROMOSOME REPOSITIONING THROUGHOUT THE CELL CYCLE

In a pioneering work in 1982, the group of Prof. Thomas Cremer, tested the Rabl model on micro-irradiated fibroblasts isolated from female Chinese hamster cells, and observed that the chromosomal arrangement in the interphase nuclei was preserved throughout the cell cycle [38]. In the same study, the authors found that both homologues of chromosomes (chr) 1, 2, 3 and 5 were positioned in close proximity. In a later study addressing the nuclear positioning of chr 3, 7, 8, 13, 17, 21, X and Y homologs in human Sertoli cells and dividing lymphocytes, it was found that in Sertoli cells, the autosomal pairs were associated closely or entirely together, whereas the opposite was seen for the sex chromosomes [39]. In lymphocytes, in contrast, no association was detected between the homologs of chr 3, 7, 8, 17, although chromosomes 13 and 21 resembled the distribution observed in Sertoli cells more closely. In 1985, using an *in situ* hybridization approach on human-mouse hybrid cell lines, it was found that each chromosome occupies a distinct domain that is spatially defined in the nucleus [40]. Further evidence of the existence of chromosome territories (CTs) was delivered by others [41–45]. In 1995, it was reported that homologs of prometaphase chromosome 7, 8, 16 and X in human diploid fibroblasts (HDFs), aneuploid HeLa, and K562 cells are localized on opposite sides of the prometaphase rosette, indicating an antiparallel organization of these chromosomes. Interestingly, chromosome 16 was found to occupy similar spatial positions in both normal diploid fibroblasts and HeLa cells, implying that at least some chromosomes might retain their normal arrangement in aneuploid cells such as HeLa [46].

One of the proposed models for radial positioning of chromosomes in interphase cells states that chromosomes are ordered based on their size, namely, small CTs (for example chr14, 15, 19-22) are found in the center of the nucleus, whereas large chromosomes reside at the periphery of the nucleus (for instance chr1, 2, 3, 4, 13) [47–49]. Another proposed model states that CTs are ordered by gene density, with gene-rich chromosomes positioned towards the nuclear interior and gene-poor chromosomes occupying more peripheral locations [50,51]. Indeed,

by performing fluorescent *in situ* hybridization (FISH) on human primary lymphocytes, fibroblasts, a lymphoblastoid cell line, and fibrosarcoma cells, it was demonstrated that chromosome 18 (85 Mb), which is gene-poor, is more peripheral and organized in a more compact territory than chromosome 19 (67 Mb), which occupies the inner part of the nucleus in all these cell types, despite being of a similar size as chr18 but being gene-rich [50].

To test the possible effect of transcription on spatial genome architecture, human primary lymphocytes were treated with inhibitors of RNA polymerases I and II [50]. Upon drug treatment, the chromosome territory of chromosome 19 but not 18 decreased in size compared to the untreated controls. Remarkably, this treatment did not have any impact on the radial positioning of both chromosomes. In another study, in human male and female lymphoblastoid cells, chr 1, 16, 17, 19 and 22, which are gene-rich, were situated in the center of nucleus, while chr 3, 4, 5, 7, 8, 11, 13 and 18, which are gene-poor, were positioned close to the nuclear periphery [45]. In the same study, no relation between the size of chromosomes and their location was observed. Similar gene-density related radial arrangement of chr18 and 19 was found in embryonic stem cells (ESCs) [52]. In the same study it was shown that some chromosomal regions, such as the whole p arm of chr12, are positioned more towards the center of the nucleus in ESCs, but not in differentiated cells. Noteworthy, this part of the chromosome contains clustered of pluripotency genes such as NANOG and Oct4. This observation suggests that chromosomal arrangement might be an important factor in controlling expression of selected genes. The same group observed that, compared to differentiated cells, ESCs presented less centromeres at the periphery of the nucleus but more around nucleoli. In differentiated cells, centromeres were associated with the nuclear periphery which mirrors what is reported for most of human cells [53–56]. For instance, in human lymphocytes, both centromeres and entire chromosome territories for chr1, 11, 12, 18 and X were found to be located in the periphery of the nucleus, while chr17 and 20 were found to occupy more internal regions. In conclusion, multiple lines of observations imply that chromosomes can occupy preferential radial locations in the nucleus of mammalian cells, and that this arrangement might depend on gene density [45,50,57].

Another model posits that chromosome-specific centromere clustering of three or more centromeres, might play a significant role in genome organization [55]. In a Hi-C based study, centromeres were found in the central part of the nucleus, which is in concordance with prior observations in ESCs [52]. Both Hi-C and DNA FISH revealed that the formation of centromere clusters is more frequent for some chromosomes such as chr1, 9 and 21, but not for example for chr 2, 3, 6. Interestingly, it was found that if a centromere is present in a large centromere cluster, it is more prone to be found in the center of the nucleus than if the same centromere is located in a smaller cluster. These observations suggest a possible role of centromere clustering in chromosome positioning towards the central part of nucleus. Furthermore, it was reported that the deposition of histone modifications specific to open chromatin, onto subcentromeric regions of the chromosome correlates positively with the frequency of these centromeres to form stable clusters [55].

Another important factor that seems to determine the position of chromosomes in the nucleus is the cell cycle phase. For example, it was shown that in G0 (quiescent) and senescent cells, chr 18 changes its position from a peripheral to a more central location in the nucleus. Upon exiting the G0 phase and re-entry into the cell cycle, the chromosome returns to the nuclear periphery within the first two to four hours of the first G1 phase. In contrast, the location of chr 19 remained central throughout the cell cycle [50]. In line with these observations, the spatial radial organization of chromosome territories of human chr 7 and 10 in HeLa cells was shown to be preserved from the mid of G1 to late G2/early prophase, while most of the relocation events occurred during prometaphase [58]. In summary, all above discussed studies indicate that chromosomes occupy preferential radial positions in cell nucleus. However, at present, it remains unclear whether these preferences arise only due to specific characteristics of chromosomes such as gene density or size, or whether other factors play a role.

Despite the enormous progress in the field of genome organization, the mechanisms that regulate the chromatin structure and how it can be established, reset and maintained, remain largely unexplored.

1.3 FRAGILITY OF THE GENOME

Arising evidence suggests a possible link between genome organization and chromosome fragility.

It has been proposed that genome architecture drives chromosome fragility. In the recent study of Canela and colleagues [59] it was shown that loop anchors bound by architectural proteins, CTCF and cohesin, are particularly vulnerable to DNA double-strand breaks (DSBs), making a connection between DSBs formation and chromosome organization. Using END-seq technique to map DSBs genome-wide at nucleotide resolution, it was found that these preferential vulnerable regions are linked to topoisomerase II (TOP2) recruitment to loop anchors. The role of this enzyme is to resolve DNA entanglements in response to torsional stress generated by transcription, replication or chromatin compaction (**fig. 6.**) [60–62]. TOP2 cuts one fragment of DNA and passes the other one through the break, followed by sealing the breakage. In addition, TOP2B as one of the two isoforms of TOP2, was found to be enriched in CTCF and cohesin bound sites that flank TADs [59,63] and to physically interacts with both CTCF and cohesin [64,65]. It has also been reported that the position of TOP2 on chromatin requires pre-existing binding of cohesin to DNA [66]. Furthermore, the data obtained with ChIP-seq technique linked TOP2B occupancy with DNase I hypersensitivity sites that mark open chromatin regions [65]. These findings suggest a possible role of TOP2B in resolving topological constrains associated with genome architecture. However, it was reported that topoisomerase II activity might lead to creation of chromosomal translocations, deletions or insertions [67,68]. Indeed, DSBs at loop anchors are enriched for previously identified breakpoint clusters that are commonly involved in oncogenic chromosomal translocations [59]. These observations indicate a possible threat that chromatin loops could impose on genome integrity.

As described above, it is evident that DNA loops influence DSBs formation, but whether these breaks impose any changes on chromatin structure remained mostly unclear until recent years. Using live cell microscopy to track chromatin loci, the group of Christopher Zimmer [69] looked at DSBs-induced chromatin remodeling in buddying yeast. The study reported that DNA damage that occurred either inside of the DNA loop or at TAD boundaries induces the stiffening of a chromatin fiber. Other studies demonstrated involvement of architectural proteins such as cohesin complex and its regulators including NIPBL in the chromatin loops reorganization upon DSB induction [70–72]. An interesting observation was made by Caron and colleagues [73] who looked at the function of cohesin in the response to DSB formation in human cells. Specifically, it was shown that cohesin is recruited to sites of DNA lesions and participates in the maintenance of transcription of active genes within chromatin domains that have been remodeled as a response to the DNA break. While examining the role of CTCF in the maintenance of the genome stability and DNA repair in human cells, it was found that this insulator protein is engaged in homologous recombination repair of DSBs and similarly to cohesin, is recruited to DNA break sites [74]. This study provided a clear link between genome organization and its stability. Moreover, using Hi-C and high-throughput, genome-wide translocation sequencing (HTGTS) technique, it was reported that nuclear organization is linked to the formation of chromosomal translocations. While examining mouse *lgh* locus upon ionizing radiation, it was shown that if it gets broken, the probability of it fusing to another broken DNA end would correlate with the spatial proximity between the two broken sites [75]. The closer the two ends would be in 3D, the higher the chances of them fusing with each other. Taking all these studies together, it is believed that chromosome architecture, especially chromatin looping, at least in part drives the formation of DSBs and their repair processes, providing a strong link between chromatin conformation and genome stability. It remains unclear whether and how DSBs and other mutational process, in turn, affect 3D genome organization.

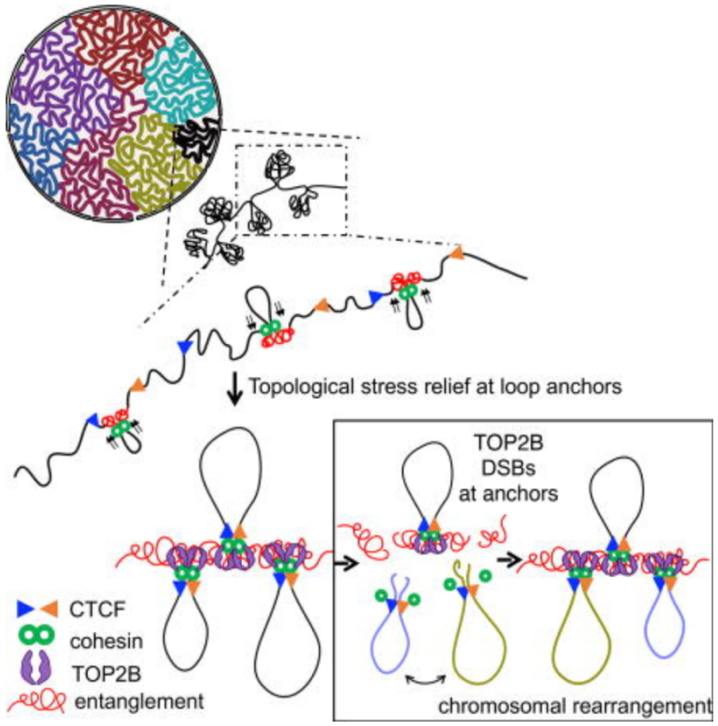

Fig. 6. Schematic representation of TOP2 role in resolving DNA entanglements in response to torsional stress. Image adapted from Canela et al, Cell, 2017.

1.3..1 Chromatin dynamics during DSBs repair

DNA damage can occur anywhere in the genome, and as it was reported by Janssen and colleagues that DSBs in both hetero- and eu-chromatin are repaired with the same kinetics [76]. However, DSBs repair processes that occur in the heterochromatin exhibit unique temporal and spatial characteristics that differ from the ones found in euchromatin [77].

Two main pathways that are available for the repair of DSBs are homologous recombination (HR) and non-homologous end joining (NHEJ). In the case of HR, a homologous DNA template is used to accurate repair a break, whereas NHEJ involves direct ligation of the two broken DNA ends, often resulting in mutations at the break site [78].

It was observed that the HR repair of heterochromatic DSBs can be challenging due to the fact that heterochromatin mostly consists of highly repetitive sequences that are prone to aberrant recombination which eventually can lead to genome instability [77,79,80]. Despite this threat, HR is the main repair pathway that is used by heterochromatic DSBs in both *Drosophila* and mammalian cells [76,77,80]. Since heterochromatin is usually compacted and is believed to be poorly accessible to RNA polymerase and DNA repair proteins, it was speculated that this can impose a barrier for a proper repair od DSBs. To overcome this limitation, it was reported that DNA accessibility within heterochromatin changes in response to DNA damage [81].

It was shown in human mammary and skin cells, that after 1h of applying ionizing radiation (IR), both repair (γH2AX) and checkpoint (53BP1) proteins are recruited to the DNA lesions, which leads to the formation of so-called radiation-induced foci [77]. These foci are non-randomly distributed in the space of the nucleus, being preferentially found at the periphery of heterochromatin [82,83]. Their distribution points to the exclusion of these foci from heterochromatin, once the break occurred, indicating relocalization of heterochromatic RIF to the euchromatic region during DSB repair [77]. Moreover, Chiolo and colleagues observed that shortly upon IR-induced damage in *Drosophila* cells, heterochromatin undergoes dynamic changes. These changes include global relaxation, expansion and protrusions of heterochromatic domains. These processes are followed by relocation of breaks towards euchromatic space at the periphery of heterochromatic domains in order to expose these lesions to the repair centers [77]. Furthermore, these DSBs are next relocated to nuclear pores to continue homologous recombination [80]. The same process of relocation was reported in mouse cells [84,85]. Proteins such as γH2AX, that are involved in this heterochromatin expansion are as well engaged in DSB signaling and relocalization. On the other hand, proteins that are involved in the late events of HR repair, such as Rad51, are recruited only after their relocalization to the periphery [77]. Heterochromatin relaxation that requires either the activity of ataxia telangiectasia mutated (ATM) kinase or the release of the heterochromatin protein HP1-β, facilitates the efficient repair of a DNA break [86,87]. These spatial and temporal dynamics that occur in heterochromatin upon damage induction, prevent repetitive sequences from aberrant recombination [76]. Taking these observations together, it has been proposed that both nuclear architecture control nuclear dynamics of the DNA repair processes.

1.4 METHODS FOR MAPPING THE 3D GENOME

In order to understand the role of the chromatin architecture in gene regulation, is important to resolve both, the chromosomes folding as well as their location in the nuclear space. Various techniques have been developed to investigate the nuclear organization, each coming with some strengths and drawbacks. In this chapter, I will present an overview of the methods that helped to reveal new aspects of the 3D genome organization, demonstrating the complexity of the nucleus.

1.4..1 DNA-FISH

Before the development of sequencing-based techniques, florescent *in situ* hybridization of DNA (DNA FISH) was the only method that allowed for the visualization of the 3D chromosome organization. Since its development in 1969 [88], DNA FISH has been the most commonly used imaging technique to study chromatin contacts in fixed cells.

In DNA FISH, cells are typically fixed with 4% paraformaldehyde, followed by permeabilization that allows for the probe (complementary DNA sequence conjugated with florescence dyes) to enter the nucleus. Next, using a combination of heat and formamide, DNA is denatured, allowing for the hybridization of fluorescent probes to their complementary target region of interest in the genome. Subsequently, using microscopy the fluorescent signal is detected.

Using DNA FISH, Barbieri and colleagues studied the contribution of the contacts between gene and promoter to the nuclear organization. The authors demonstrated that in mouse ESCs, in gene-dense region located on a *HoxB* locus, state-dependent (active and poised) promoter interactions are the key players in DNA folding [89]. Bottiger and colleagues applied DNA FISH to *Drosophila* in order to detect chromatin compaction of chromatin characterized by different epigenetic states [90]. Finn and colleagues applied hiFISH, a DNA FISH based technique, to visualize high cell-to-cell variability in the spatial organization of the human genome [91]. Cryo-FISH, another high-resolution variation of the traditional FISH, was used to detect long-range contacts (>100 kb) that occur between TADs [92]. In this approach, FISH probes are hybridized to thin 0,1-0,2 um thick cryosections and imaged using florescent microscopy [93,94]. This FISH approach was applied, for example, to study intermingling between chromosomal territories [93]. The power of DNA FISH arises from its ability to provide single-cell resolution but it is limited to the simultaneous visualization of a small number of genomic loci. To overcome this challenge, Wang and colleagues developed a multiplexed FISH that allows for sequential visualization of many genomic regions in the fixed nuclei [95]. This approach was applied to human fibroblast cells (IMR90) to track spatial organization of TADs and compartments in individual interphase chromosomes, confirming that chromatin is arranged into large contact domains [95]. On the other hand, higher resolution maps (30 kb) were obtained with super-resolution imaging of oligopaint-based FISH [96]. In this approach, short in length and high in

specificity fluorophore-tagged oligonucleotides (oligopaints) were hybridized to 30 kb DNA segments of IMR90 cells, revealing diverse chromatin configurations in single cells.

Both, the power and accuracy of DNA FISH techniques to visualize genomic regions is directly linked to the preservation of genomic structures, the resolution of the used microscope as well as the size of the sequence of interest. It was shown, that the denaturation step used in the DNA FISH procedure might introduce alterations in chromatin ultrastructure [97]. The other disadvantage is the high cost of DNA FISH which arises from the fact that the protocol is probe- and sample -specific, and in order to optimize it, a large number of typically very expensive commercial probes is needed. However, recent development of iFISH platform minimizes the cost and time of probes production, making DNA FISH more accessible for the 3D genome community [98].

1.4..2 Hi-C

Hi-C (high-throughput chromosome conformation capture) was the first technique to create a genome-wide map of chromatin contacts in mammalian genome [33].

In the original Hi-C, cells are fixed with paraformaldehyde, followed by fragmentation of DNA with restriction enzyme of choice. The fragmented DNA that have sticky ends are filled in with biotinylated cytosine nucleotide. Then, under diluted conditions, DNA fragments are ligated and many of them form chimeras (when the two ligation partners come from distant genomic regions, as opposed to the immediately adjacent ones). Next, genomic DNA is isolated, mechanically sheered and using streptavidin beads, biotinylated fragments are pulled down. Captured fragments are then amplified and subjected to pair-end high-throughput sequencing. Streptavidin step minimizes the number of unligated events being sequenced, leading to about 50-70% of sequencing reads being mapped to ligated restriction fragments [33].

Since its first development in 2009, Hi-C protocol has been further developed by many groups, leading to the improvement of this technique. One of the Hi-C modifications is called DNase-Hi-C and uses DNase I to fragment chromatin [99]. This technology was developed to overcome the resolution limitations that are associated with restriction enzyme digestion that were reported for the original Hi-C [100]. It was observed that using DNase I, both efficiency and resolution (40 kb) were highly improved compared to the original Hi-C (1 Mb) [99]. Another Hi-C based technique that omits the use of the restriction enzyme to fragment chromatin is Micro-C. Developed by Hsieh and colleagues, Micro-C involves the micrococcal nuclease (MNase) digestion of the chromatin and therefore allowing to determine structures at the single nucleosome resolution (between 200 bp and 4 kb) [101]. Using Micro-C on mouse ESCs, the authors confirmed the evidence of 30 nm chromatin fiber as well as supported the zigzag chromatin folding model [3,102]. Moreover, this technique identified dynamic and small (5- 10 kb) chromatin structures such as self-interacting domains due to the detection of boundaries

that were not observed with the original Hi-C [33]. These structures seem to be functionally related to gene activity and genome organization [102]. Another interesting improved version of Hi-C, named *in situ* Hi-C was reported by Rao and colleagues, in which the authors omitted the use of sodium dodecyl sulfate (SDS) in their protocol in order to better preserve the 3D structure of the nucleus [18]. In the original Hi-C, SDS was used to disrupt the nuclear membrane and it was thought that it remains partially in the solution during ligation step. It was observed that in *in situ* Hi-C, due to the lack of SDS, the amount of random ligations decreased, whereas signal-to-noise ratio improved. In addition, *in situ* Hi-C is faster than the traditional Hi-C and it does not require dilution of the crosslinked DNA, therefore allowing to perform the experiment in smaller volumes. As a result, the efficiency of both ligation and DNA isolation is improved. All these steps led to generation of higher resolution interaction maps (1 kb) compared to the original Hi-C. Since all the described Hi-C approaches are bulk assays and require over 100 million cells, it was impossible to provide spatial information at the single-cell level. To overcome this limitation, Nagano and colleagues developed single-cell Hi-C that enables creation of interaction maps at the single-cell level [21]. In this approach, chromatin is crosslinked using formaldehyde and then ligated, followed by the digestion with endonuclease. Next, single nuclei are isolated from the population of cells, and then DNA libraries are prepared from each nucleus. The advantage of this technology is that it can be applied to a rare material, as it was shown in the study on oocyte-to-zygote transition [103], reveling unique chromatin reorganization. One of the drawbacks of the single-cell Hi-C is the low recovery of chromatin interactions. This might be the result of inefficient enzymatic digestion, ligation or DNA extraction, which can contribute to the higher cost of sequencing as it might require higher depth to reach a desired resolution.

Overall, both original Hi-C and its derivatives contributed enormously to our knowledge about 3D organization of the genome, including genome folding, gene regulations, or how regulatory elements are linked to the nuclear topology as summarized in the introduction chapter.

1.4..3 Dip-C

It was pointed out that the resolution of single-cell methods that are used to investigate 3D genome is limited due to various reasons such as high cell-to-cell variability or technical problems [37]. This limitation, for instance, challenges the identification of chromatin interactions involved in gene regulation [104]. Recently, a technique called Dip-C was developed to detect large number of intra- and inter-chromosomal contacts in single diploid cells [105].

In Dip-C, that stands for diploid chromatin conformation capture, cells are fixed with 2% of paraformaldehyde, followed by digestion with the MboI restriction enzyme. Next, DNA fragments that are positioned in a close proximity ligate to each other, followed by sorting of single cells. Subsequently, cells are lysed, genomic DNA is isolated, ligation products are amplified using whole-genome multiplex end-tagging amplification technique (META) and

sequenced. Dip-C increased the sensitivity and resolution over single-cell Hi-C [21] by avoiding inefficient biotin-associated steps, and by applying META procedure that led to more efficient amplification of ligation products. As a result, the amount of detected interactions increased on average to 1 million per cell in comparison to 10 thousands in single-cell Hi-C [21,105].

It was reported that, in both human lymphoblastoid cells (GM12878) and human peripheral blood mononuclear cells, Dip-C can detect topological differences between pairs of alleles undergoing genomic imprinting at 20 kb resolution [105].

Unlike single-cell Hi-C that would first require a cell line to be haplotyped [21], Dip-C can be applied to study genome organization, chromosome intermingling and enhancer-promoter interactions in diploid single cells.

1.4..4 SPRITE

To overcome challenges that arise from proximity ligation, in 2018, Quinodoz and colleagues developed a sequencing-based technique called SPRITE [106]. SPRITE, that stands for split-pool recognition of interactions by tag extension, allows for genome-wide measurement of higher-order DNA-DNA or DNA-RNA interactions within the nucleus. In this assay, cells are fixed in suspension with paraformaldehyde to ensure the crosslinking of chromatin. After fixation, nuclei are isolated and chromatin is mechanically sonicated and further fragmented by DNAse digestion. Next, all crosslinked complexes are split into 96-well plate and then, in each well, all DNA molecules are tagged with a unique barcode sequence. After ligation of a specific barcode, all complexes are pooled together into one well. The procedure of splitting, pooling and barcoding is repeated several (3-5) times and enables all the molecules present in one interacting complex to be tagged with a complex-specific barcode sequence. Subsequently, all ligated events are sequenced and generated reads are clustered by their shared barcodes.

Using SPRITE, it was shown that in both mouse embryonic stem cells and in human lymphoblastoid cells (GM12878), transcriptionally active regions are organized around nuclear speckles, while inactive DNA regions are positioned around nucleoli. In addition, similar to ligation-proximity assays, SPRTIE recapitulated known higher-order genome structures, such as chromosome territories, compartments, TADs and DNA loops [106].

Despite the power of SPRITE, this technique possesses several limitations. One of them might arise from the length of crosslinked fragments (up to 1 kb), that could lead to artifacts during library preparation, thus eventually lowering the representation of inter-chromosomal contacts. Another disadvantage emerges from the fact that SPRITE is a bulk assay that requires millions of cells in order to create comprehensive genome structure maps, therefore limiting its application to precious material or being applied to single cells.

Similar to TSA-seq [107] (see below), this method identifies the importance of nuclear bodies in shaping the higher-order genome structure in the nucleus [106].

1.4..5 DamID

Another technology that maps chromatin contacts with intra-nuclear structures is DamID. DamID, that stands for DNA adenine methyltransferease identification, is a technique that was first used to generate a map of genome-wide interactions between DNA loci and protein of interest [108]. In this approach, *E. coli* adenine methyltransferease (Dam) is fused to the DNA binding domain of a protein of interest that is expressed at low levels. This binding results in local DNA methylation, specifically, of adenines in the GATC motif of nearby sites where this protein interacts with DNA. Next, genomic DNA is isolated and treated with methylation-sensitive endonuclease (DpnI), that can only cut methylated GATC. Enzymatic digestion is followed by ligation of PCR adapters to the cut DNA to ensure that only methylated fragments are amplified. The ligation is followed by digestion with DpnII, that cuts only remaining unmethylated GATC sequences, excluding them from amplification. Finally, using PCR, only methylated fragments are amplified, followed by sequencing [108–110].

In 2008, using the above protocol, Guelen and colleagues created a contacts map of the whole genome with nuclear lamina (NL) [111]. It was shown that in human fibroblasts, genome-lamina interactions occur through so-called lamina associated domains (LADs) and that around 35% of the genome interacts with the nuclear lamina. DamID, on average, requires 100 thousands cells in order to reconstruct a map of lamina contacts. To investigate how much of the genome interacts with NL at the single-cell level, Kind and colleagues reported a modified version of the DamID - single-cell DamID [112]. The design of the technique is similar to the original DamID, thought with some differences, such as: 1) the use of clonal cell lines with a more controlled Dam-LmnB1 expression; 2) the use of the flow sorting of single cells and 3) the use of the 96-well plate format. Using single-cell DamID, it was shown that in human myeloid leukemia cells (KBM7), all studied cells share the same 15% of DNA that interacts with lamina, indicating that there are some fixed contact regions. The group also observed that some regions of the genome tend to be more variable than others, forming lamina contacts in some cells but not in others [112]. These findings provide spatial information about chromatin with regard to the nuclear lamina. Another interesting version of DamID, named as DamC was reported by Redolfi and colleagues [113]. DamC involves DNA methylation-based identification of chromosomal interactions with next-generation sequencing together with modeling of methylation kinetics. Using DamC on mouse ESCs, the authors for the first time ever provided an evidence that both TADs and CTCF loops exist not only *in vitro* but as well *in vivo* [113].

The power of DamID arises from different aspects, such as ligation-free design, no use of paraformaldehyde for chromatin crosslinking, no chromatin immunoprecipitation or antibody targeting, that could contribute to some biases, as in the case of the ChIP-seq method [114] (see below).

As with all methods, DamID has limitations. For instance, the resolution depends on the distribution of GATC in the studied genome. It was also reported that Dam can methylate adenines that are up to 5 kb away from the targeted motif, limiting the spatial accuracy of the

method [108]. The location of chromatin can be accurately estimated only when genomic loci are close to the nuclear lamina but not to other nuclear compartments, such as splicing speckles. It is due to the fact that the majority of chromatin that surrounds these sub-nuclear structures does not bind directly to the tagged protein [115]. Another drawback is that DamID requires transgenic cell lines or animals that could express Dam-fusion protein of interest, as methylation of adenine does not happen in eukaryotes, with the exception of *D. melanogaster* [116] or *C. elegans* [117]. On this note, DamID could not be applied to human biopsies. Furthermore, this technology can only be applied to DNA but not to RNA.

1.4..6 GAM

Another ligation-free technology for mapping genome-wide chromatin contacts was developed in 2017 by Beagrie et al. [92]. The technique is called GAM and stands for genome architecture mapping.

In GAM, that was initially applied to mouse ESCs, cells are first fixed with two concentrations of paraformaldehyde (4 and 8%), then pelleted to be subsequently embedded in sucrose to prevent damage while freezing in liquid nitrogen. After freezing, cells are cryo-sectioned at random orientation to slices of a typical thickness of 0.22 μm. Next, using laser microdissection, single nuclear slices are isolated. Subsequently, from each slice, genomic DNA is extracted and amplified using whole-genome amplification (WGA) method. Finally, during library preparation, DNA from each slice is uniquely indexed with sequencing adapters and then pooled together with DNA coming from all sections, followed by high-throughput sequencing.

Sequencing data are then used to identify chromatin interactions between pairs or even triplets of loci. Regions that are in a close proximity are expected to be more often found in the same nuclear slice than more distant loci, which serves as the basis for the analysis of the GAM data.

Using GAM, interactions between super-enhancers that are distant tens of megabases in the space of nucleus were found [92]. This technology also confirmed the presence of TADs, that were observed in Hi-C maps [17,92].

The resolution of the GAM technique depends on the number of cut nuclear slices processed. It was reported that with 400 sections equaling to 46 cells, it was possible to generate interaction maps at 30 kb resolution, similar to the one obtained with Hi-C that used millions of cells with the same sequencing depth [17,92]. Unlike ligation-based assays, GAM does not require restriction enzymes, thus avoiding any possible biases toward regions with a high density of cut sites [92,100]. It is worth to mention that GAM could be applied to the clinical samples, given that no genetic modification of the studied material is needed here.

Despite its multiple advantages, GAM has limitations as well. One of them is the time that is needed to prepare the material, including cryo-sectioning of cells and nuclei laser

microdissection, therefore limiting its application to a high number of cells. Another drawback is the fact that this technology requires that all cells should have relatively consistent shape and size in order to rely on the information provided by the size of a given nuclear slice [92,118].

1.4..7 TSA-seq

One of the recently developed techniques that contributes to defining the position of chromatin in 3D space of nucleus is TSA-seq [107]. Unlike, DamID [108–110], TSA-seq provides information about interactions between DNA and various intra-nuclear bodies or sub-compartments, such as splicing speckles [107].

TSA-seq, that stands for tyramide signal amplification sequencing, enables genome-wide measurement of chromatin distances to defined sub-nuclear structures [107]. In this assay, a given nuclear body is targeted by an antibody (specific to that nuclear body)-coupled horseradish peroxidase that generates highly active biotin-conjugated tyramide free radicals. Any genomic sequence that is in a close proximity to what the antibodies recognize will be biotinylated. Importantly, the closer the chromatin is to the targeted sub-nuclear structure, the more it will be biotinylated. In TSA-seq, cells are first crosslinked with 1.6% of paraformaldehyde and then stained with specific antibody. Once the labelling is done, DNA is isolated and mechanically sheared. Subsequently, the biotinylated DNA is pulled down followed by high-throughput sequencing. Generated reads are then used to create a map of distances between all DNA loci and their nearest intra-nuclear body of interest.

TSA-seq targeting nuclear speckles in human leukemia cells (K562) identified two types of transcriptionally active hubs, called “hot zones”. Both hubs are characterized by the same features, such as enrichment in RNA pol II (RNAPII), active histone marks, high gene expression and gene density. The difference found, for instance, was their nuclear location. Genomic distance to the nearest LAD was found to be higher for type I than for type II. Type I was associated with A1 subcompartment, previously identified by Hi-C [18], whereas type II spanned A2 domain. It was reported that genes that occupy more internal locations, near nuclear speckles (type I), tend to be more active than genes that are in closer distance to the nuclear lamina (type II) [107].

TSA-seq is one of the first techniques to map actual distances between chromatin and nuclear subcompartments. In contrast to DamID [108–110], TSA-seq can map chromatin proximity to various nuclear bodies or sub-compartments, including, but not limited to, nuclear lamina [107].

Despite multiple valuable features, TSA-seq is not without shortcomings. One of them is the high amount of the starting material, which is over 100 million cells, therefore not allowing to study chromatin dynamics at single-cell resolution. This technology seems to be both costly and time-consuming, as it requires sequential antibody staining of different batches of cells for over 1 month [107].

1.4..8 ATAC-seq

An interesting assay to map chromatin accessibility at the genome-wide level was developed in 2013 by Buenrostro and colleagues and was named ATAC-seq [119]. In ATAC-seq, that stands for assay of transposase accessible chromatin sequencing, cells are first lysed and then treated with hyperactive Tn5 transposase. This transposase binds to open chromatin and simultaneously cuts and tags genome with sequencing adaptors. Next, DNA is purified using spin columns and transposed DNA fragments are amplified with PCR. Subsequently, PCR products are processed through the library preparation step, followed by paired-end high-throughput sequencing. As a result, generated sequencing data are used to identify regions that are either occupied by nucleosomes or are nucleosome-free, therefore measuring the openness of the genome.

ATAC-seq that was applied together with ChIP-seq to human lymphoblastoid cells (GM12878), identified several classes of transcription factor (TF) binding modes with respect to the closest nucleosome [119]. One of these classes is characterized by TFs that avoid nucleosome in their close proximity (180 bp away), another class is driven by factors that bind close to nucleosome, specifically at sequences to which CTCF and cohesin-complex subunits bind. The third class of a binding mode was found to be specific to nucleosome-associated DNA and involves factors involved in chromatin remodeling, such as CHD1, SIN3A or RNA poll II. It was also shown, that transcriptional start sites (TSSs) tend to be enriched in nucleosome-free fragments. These findings provide genome-wide information about nucleosome position and transcription factor binding [119].

Some of the advantages of the original ATAC-seq protocol included low number of cells, short time of preparation, no use of paraformaldehyde (even though it can be used if needed) and paired-end sequencing. It was shown, that this assay was able to capture open chromatin sites with as little as 500 cells and to generate complex libraries with 50 thousands cells [119]. This sample reduction was a big improvement over DNase-seq that required 50 million cells [120]. The ATAC-seq protocol was designed in a way that it could be completed within 3 h, giving this assay an advantage over any other technique that measured either chromatin accessibility, like DNase-seq [120] or TF binding, such as ChIP-seq [121,122], that take several days to complete. The simplicity of the technique could as well be found in the fact that this protocol does not require chromatin fragmentation nor antibodies use [119]. Another advantage of ATAC-seq is the benefit of using paired-end sequencing, that leads to more precise alignment of reads that are mapped over the repetitive regions in the genome [119].

On the other hand, one of the drawbacks of this technique is the fact that Tn5 has a bias towards A/T regions [123]. The presence of mitochondrial DNA is another shortcoming of ATAC-seq. It was shown that 20-80% of total sequencing reads generated with this assay, come from mitochondria, thus decreasing the complexity of libraries [124].

Recent improvements of the original ATAC-seq led to development of the single-cell ATAC-seq (scATAC-seq) that allows to look at the chromatin state in individual cells [125]. In this approach, after the cells are lysed and treated with Tn5, single cells are isolated using microfluidic platform. Next, DNA is purified from each cell and barcoded with a unique adaptor, followed by single-cell library preparation and then pooled together for high-throughput sequencing. Using 254 single cells (GM12878), scATAC-seq recreated maps of accessible regions obtained with a population-based ATAC-seq (10 thousands cells) [119], thus showing its high accuracy. scATAC-seq has been used to study the chromatin accessibility landscape at the level of individual cell, hence revealing the mechanisms regulating the cell-to-cell heterogeneity [125].

1.4..9 ChIP-seq

ChIP-seq is another interesting genome-wide method that maps genomic regions that are bound by a particular protein, for example transcription factors.

ChIP-seq, that was developed in 2007, stands for chromatin immunoprecipitation sequencing [121]. In this method, cells are fixed with 1% of paraformaldehyde to crosslink proteins around DNA, followed by DNA extraction. Next, DNA is mechanically sheered and pulled down with a bead-labelled antibody, targeting the protein of interest. Subsequently, using magnet, antibody-protein-DNA complex is immunoprecipitated, followed by reversal of the crosslinking. Next, released DNA is tagged with sequencing adapters and amplified using PCR. Finally, DNA libraries are sequenced using high-throughput platforms.

As a result, generated sequencing data are used to identify specific DNA sequences that are bound to proteins of interest, therefore allowing to measure TF binding sites [121,122] or histone modifications [126,127]. ChIP-seq has been used, for instance, to study the location of super-enhancers in multiple tissues, showing that they span tens of kb of DNA sequence and are enriched in master TFs and the Mediator complex. These findings indicate that super-enhancers play a major role in orchestrating gene expression patterns due to the involvement of master TFs in specifying cell lineages [128,129]. ChIP-seq was also used in genome-wide mapping of CTCF binding sites, providing evidence that CTCF plays a key role in the barrier activity of insulator sequences which leads to separation of active and inactive chromatin domains [130]. This technique has also been applied to study the recruitment of repaired proteins, such as γH2A.X, to the location where double-strand break occurred [131]. Recent development of single-cell ChIP-seq showed that in human fibroblasts, ESs and hematopoietic progenitor cells, cell subpopulations are defined by the state of chromatin, thus revealing epigenetic heterogeneity between individual cells [132].

An interesting version of ChIP-seq, named ChIA-PET was reported by Fullwood and colleagues [133]. ChIA-PET that stands for chromatin interaction analysis with paired-end tag sequencing was developed to overcome the limitation of ChIP-seq to study long-range chromatin interactions to determine, for example, the target genes of the distal TF binding sites. ChIA-PET has been used, for instance, to identify different types of loci interactions, such as enhancer-promoter, enhancer-enhancer and promoter-promoter interactions, showing that 40% of enhancers are not involved in the regulation of their nearest promoters. These findings refute the previous assumption made by ChIP-seq, that the TF binding sites regulate their nearest genes [121,133]. Using ChIA-PET, Demare and colleagues demonstrated on multiple types of tissue, that 65% of chromatin interactions that are bound by SMC1A cohesin subunit are co-occupied by another architectural protein - CTCF [134].

Despite its power, ChIP-seq has some limitations. The complexity of libraries generated with ChIP-seq, is closely related to the quality of antibody used, time of the crosslinking, the amount of initial material, chromatin fragmentation, or over-amplification with PCR [135]. Testing all the above factors make ChIP-seq laborious and time consuming. Regarding mechanical sheering, it was shown, that heterochromatin that is typically not associated with TF binding, happens to be more resistant to the fragmentation than euchromatin, thus leading to biases [136]. The use of unspecific or low-quality antibodies can be another drawback of ChIP-seq, which DamID does not have [108–110].

Overall, Chip-seq has been particularly useful in mapping the occupancy of histone variants as well the occupancy of histones carrying various post-translational modifications.

1.4..10 Methodological consideration

DNA FISH and sequencing-based approaches capture different aspects of chromosome structure due to their different sensitivity and resolution, which are assay-specific [14,45,52,55]. Although there is often a good correlation between sequencing and imaging data some studies show differences [137]. Therefore, the data coming from either of these approaches should be validated by either new techniques or by each other.

2 DOCTORAL THESIS

2.1 AIMS OF THE STUDY

The overall aim of this thesis was to develop methods for mapping genome organization and genome fragility in the 3D space of the nucleus.

The specific aims of the constituent papers were:

Paper I:

- To establish a novel genome-wide technique to assess the radial organization of the genome in the human cell nucleus

Paper II:

- To establish a novel genome-wide technique to precisley profile the location and frequecy of DSBs in low-input samples

2.2 METHODS

2.2..1 Cells and tissue

For GPSeq and YFISH:

HAP1 cells were obtained from Horizon Discovery (cat. no. C859) and cultured in Iscove's Modified Dulbecco's Medium (IMDM, Merck, cat. no. 51471C) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific, cat. no. F2442). GM06990 cells were obtained from the Coriell Cell Repository (cat. no. GM06990) and cultured in Roswell Park Memorial Institute Medium 1640 supplemented with 2 mM L-glutamine (RPMI, Sigma, cat. no. R8758) and 15% fetal bovine serum (Thermo Fisher Scientific, cat. no. F2442). Cells were maintained in a humidified environment at 37 °C, containing 5% CO_2. All cells tested negative for Mycoplasma using MycoAlert™ Mycoplasma Detection Kit (Lonza, cat. no. LT07-118).

For BLISS:

KBM7 cells were obtained from Oscar Fernandez-Capetillo (SciLifeLab, Stockholm, Sweden); U2OS cells from Mats Nilsson (SciLifeLab); HEK 293 cells from ATCC, mESCs from Simon Elsaesser (SciLifeLab). KBM7 cells were grown in Iscove's modified Dulbecco's medium (Life Technologies, cat. no. 10829018), supplemented with 10% fetal bovine serum (FBS, Gibco, cat. no. F2442); U2OS cells were grown in DMEM medium (Life Technologies, cat. no. D0819), supplemented with 10% FBS; HEK 293 T in DMEM supplemented with 10% FBS; and mESCs were grown in minimal essential medium (Sigma, cat. no. M2279), supplemented with 20% FBS, 1% GlutaMAX (Gibco, cat. no. 35050061), 1% non-essential amino acids (Gibco, catalogue number 11140035), 1% sodium pyruvate (Gibco, cat. no. 11360070) and 0.2% β-mercaptoethanol, in the presence of leukemia inhibitory factor (Sigma catalogue number L5158-5UG) corresponding to 1,000 U ml^{-1}. All cells tested negative for Mycoplasma using MycoAlert™ Mycoplasma Detection Kit (Lonza, cat. no. LT07-118). Mouse liver biopsies were obtained from wild-type, 6-week-old C57/BL6 male mice that were sacrificed following the guidelines in the MIT protocol 0414-027-17 'Modeling and Treating Genetic Disease Using Targeted Genome Engineering' (IACUC AWA A3125-01, IACUC 0411-040-14, approval date 5/16/2013).

2.2..2 Sample preparation for GPSeq, YFISH and BLISS

For GPSeq and YFISH:

In the case of HAP1, cells were seeded directly onto coverslips placed in 6-well plates. In the case of GM06990, which grow in suspension, cells were spotted onto coverslips pre-coated with Poly-L-Lysine (Sigma, cat. no. P8920-100 ml) and placed inside a 6-well plate and

incubated for 10 min at room temperature (RT). Next, cells were fixed in 0.4X PBS (Thermo Fisher Scientific, cat. no. AM9625)/4% paraformaldehyde (EMS, cat. no. 15710) for 10 min at RT, followed by quenching of unreacted paraformaldehyde in 1X PBS/125 mM glycine for 5 min at RT. Subsequently cells were washed three times, 5 min each, with 1X PBS/0.05% Triton X-100 at RT and permeabilized in 1X PBS/0.5% Triton X-100 for 20 min at RT. Following overnight incubation in 1X PBS/20% glycerol at RT, cells were subjected to four cycles of freeze-and-thaw in liquid nitrogen, and then washed three times, 5 min each in 1X PBS/0.05% Triton X-100 at RT. Afterwards, cells were incubated in 0.1 N HCl for 5 min at RT and quickly rinsed them twice in 1X PBS/0.05% Triton X-100 at RT. Lastly, cells were rinsed in 2X SSC buffer (Thermo Fisher Scientific, cat. no. AM9763) and stored in 2X SCC/0.05% NaN_3 at 4 °C, until further processing.

2.2..3 Workflow for GPSeq and BLISS

For GPSeq:

A detailed step-by-step protocol has been published in "Nature Protocol Exchange" (https://doi.org/10.21203/rs.3.pex-570/v1). Briefly, fixed cells were subjected to the restriction digestion using either HindIII (NEB, cat. no. R3104S) or MboI (NEB, cat. no. R0147M) enzymes at 37 °C, for a defined time, ranging from 1 to 30 min (MboI) and to 6h (HindIII). The digestion reaction was stopped by placing samples in ice-cold buffer (1X PBS/50 mM EDTA/0.01% Triton X-100) and washed number of times. Next, cells were dephosphorylated using calf intestinal alkaline phosphatase (Promega, cat. no. M1821) for 2 h at 37 °C, followed by the ligation of either YFISH or GPSeq adaptors.

For YFISH:

YFISH adapters were ligated to the introduced cut sites in a T4 DNA ligase reaction mix (Thermo Fisher Scientific, cat. no. EL0014) for 18 h at 16 °C. Next, the unligated adapters were washed away with the high-salt buffer (10 mM Tris-HCl/1M NaCl /0.5% Triton X-100 pH 8), followed by the hybridization of a labelled oligonucleotide at 200 nM for 18 h at 30 °C. Afterwards, unhybridized oligonucleotides were washed away and cells were stained with 0.1 ng/μ Hoechst 33342 (Thermo Fisher Scientific, cat. no. H3570) for 30 min at 30 °C. Next, samples were mounted in ProLong Gold Antifade Mountant (Thermo Fisher Scientific, cat. no. P36930) and imaged either with wide-field epifluorescence microscopy or STED microscopy.

For GPSeq:

GPSeq adapters were ligated to the introduced cut sites in a T4 DNA ligase reaction mix (Thermo Fisher Scientific, cat. no. EL0014) for 18 h at 16 °C. Next, the unligated adapters were washed away with the high-salt buffer (10 mM Tris-HCl/1M NaCl /0.5% Triton X-100 pH 8), followed by scrapping off cells from the coverslips that were then digested with

Proteinase K (NEB, cat. no. P8107S), for 18 h at 56 °C. Subsequently, gDNA was extracted, sonicated and *in vitro* transcribed using T7 RNA polymerase (Thermo Fisher Scientific, cat. no. AM1334-5) for 14 h at 37 °C. Afterwards, RNA product was processed through the library preparation using the modified TruSeq Small RNA Library Preparation kit (Illumina, cat. no. RS-200-0012) and sequenced on the NextSeq 500 platform (Illumina).

For BLISS:

A detailed step-by-step BLISS protocol is provided in "Nature Protocol Exchange" (https://doi.org/10.21203/rs.2.1448/v2). In case of cell lines, we either grew them directly onto 13 mm coverslips (VWR, cat. no. 631-0148) or we spotted them onto coverslips pre-coated with poly-L-lysine (Sigma, cat. no. P8920-100ML). For Cas9 and Cpf1 experiments, cells were grown in the 24-well plate and fixed with 4% paraformaldehyde. For mouse liver tissues, we appplied two approaches: (1) Tissue cryopreservation and sectioning: liver biopsies were first fixed in paraformaldehyde 4% for 1 h at 25 °C and then immersed in a sucrose gardient (15% overnight and then 30% until the tissue sank) before embedding in optimal cutting temperature medium (OCT). Next, 30 µm-thick tissue sections were mounted onto microscope slides, dried at room temperature (RT) and stored at 4 °C before further processing. (2) Preparation of nuclei suspensions: freshly extracted liver biopsies were cut into small pieces and transferred into tube containing nucleus isolation buffer (NaCl 146 mM, Tris-HCl 10 mM, $CaCl_2$ 1 mM, $MgCl_2$ 21 mM, bovine serum albumin 0.05%, Nonidet P-40 0.2% pH 7.8). Samples were incubated for 15–40 min until the tissue fragments became transparent, subsequently the nuclei were centrifuged for 5 min at 500 *g* and then re-suspended in 200–500 µl of 1 × PBS. 100 µm of nuclei suspension were dispensed onto a 13 mm coverlisp, previously coated with poly-L-lysine-coated and incubated for 10 min at RT. Next, 100 µl of paraformaldehyde 8% in 1 × PBS was gently added and incubated for 10 min at RT, followed by two washes in 1 × PBS at RT. The samples were stored in 1 × PBS at 4 °C, until further processing.

-*Cas or Cpf1 expression constructs and transfections*

The following plasmids were used: Cas9-BLISS located within both the *EMX1* locus (5′-GAGTCCGAGCAGAAGAAGAAgGG-3′) and the *VEGFA* gene locus (5′-GGTGAGTGAGTGTGTGCGTG tGG-3′). In case of Cpf1, the same expression vector was used to clone AsCpf1 and LbCpf1 along with their cognate sgRNAs. For transfection procedure, cells were first plated in 24-well plates pre-coated with poly-D-lysine (Merck Millipore, cat. no. A003E) and grown until they reached 60–70% confluence. Next, to each well a mix of 2 µl of Lipofectamine 2000 (Life Technologies, cat. no. 11668019) and 500 ng of Cas9 plasmid in 100 µl total of OptiMEM (Gibco, cat. no. 31985062) was applied.

2.2..4 3D DNA FISH, immunofluorescence and imaging

3D DNA FISH

DNA FISH probes were designed and produced according to iFISH protocol [98]. Briefly, samples were incubated for 1 h at 37 °C with the pre-hybridization buffer (2X SSC/5Denhardt's solution/50 mM sodium phosphate buffer/1 mM EDTA/100 ng/ml ssDNA/50% formamide, pH 7.5–8.0), followed by incubation with the first hybridization mix, containing single-locus probes (up to 6 at the same time at the final concentration of 3.2 nM per probe). This process included DNA denaturation for 3 min at 75 °C on a heating block, followed by the incubation of samples for 18 h at 37 °C. Next, samples were washed and incubated for 3 h at 30 °C in a humidity chamber with the second hybridization mix, containing the secondary fluorescently labeled oligos (one color per locus, up to 6 colors in the same mix, each at the final concentration of 20 nM). Subsequently, samples were washed and stained with 0.1 ng/µl Hoechst 33342 for 30 min at 30 °C. Afterwards, samples were imaged using wide-field microscopy.

Immunofluorescence to detect markers of DSBs

Immunostaining of DSB marker γH2A.X was performed using a mouse anti-phospho-Histone H2A.X (Ser139) primary antibody (Millipore, cat. no. 05-636), diluted 1:200 in a blocking buffer (1X PBS/0.1% Tween 20/1% BSA), for 1 h at RT, followed by the incubation with the secondary antibody diluted 1:500 in blocking buffer, for 1 h at RT. Afterwards, stained samples were imaged every 0.4 µm using Z stack module to cover the entire volume of nuclei with an LSM 780 confocal microscope (Zeiss) equipped with a 63X oil objective.

Gradual antibody diffusion

For the gradual antibody diffusion, samples were fixed according to the YFISH and GPSeq protocol, described above. Next, samples were subjected to the incubation with the diluted 1:500 primary antibody in a blocking buffer, for a defined time, ranging from 1 to 30 min or 1h at RT. Directly after incubation, cells were post-fixed with 1X PBS/4% PFA for 10 min at RT, followed by incubation with the secondary antibody diluted 1:500. in blocking buffer for 1 h at RT.

For this procedure, following antibodies were used: rabbit Anti-Histone H2A (Cell Signalling Technology, cat. no. 12349S), Anti-rabbit IgG ATTO 488 conjugate (Abcam, cat. no. ab150077) and Anti-rabbit IgG Alexa Fluor 647 conjugate (Abcam, cat. no. ab150075).

Imaging

For YFISH, 3D DNA FISH and gradual antibody diffusion, samples were imaged using a wide-field epifluorescence microscopy, equipped with a 100x 1.45 NA objective mounted on a custom-built Eclipse Ti-E inverted microscope system (Nikon) controlled by the NIS Elements software (Nikon) and equipped with an iXON Ultra 888 ECCD camera (Andor Technology). For YFISH, 3D DNA FISH and gradual antibody diffusion, multiple image stacks per sample were acquired, each consisting of 40–110 focal planes every 0.2 or 0.3 μm.

2.2..5 GPSeq, YFISH and BLISS adapters

Individual oligonucleotides (oligos) were purchased from Integrated DNA Technologies in desalted form at 100 mM concentration in nuclease-free water. UMIs were generated by random incorporation of the four standard dNTPs using the 'Machine mixing' option. Each forward oligo was diluted to 10 mM and phosphorylated with T4 Polynucleotide Kinase (NEB, cat. no. M0201) for 1 h at 37 °C, followed by incubation with the corresponding reverse oligo for 5 min at 95 °C and gradually cooled down to 25 °C over a period of 45 min (–1.55 °C/min).

2.2..6 Sequencing and data processing

Sequencing for both GPSeq and BLISS, was performed on Illumina NextSeq 500 platform using NextSeq 500/550 High Output v2 kit (75 cycles). Based on the RA5 index sequences, raw sequencing reads were demultiplexed either automatically, using Illumina BaseSpace Sequence Hub cloud, or manually, using bcl2fastq (v2.18). Demultiplexed reads were then used to generate FastQ files. A custom-built pipeline was used to scan for reads containing the full prefix, UMI_barcode_restriction site (GPSeq) and for UMI_barcode (BLISS). Next, the reads were trimmed in order to remove prefix and using *bwa-mem* aligned to the reference genomes: for GPSeq, human (Grch37/hg19), and for BLISS both human (Grch37/hg19) and mouse (Ncbi37/mm9). Only reads with the mapping quality ≥ 30 were kept and were further filtered based on the quality of UMI sequence to identify and remove PCR duplicates. Afterwards, the BED file containing the genomic coordinate associated with a number of unique UMIs was generated.

2.3 SUMMARY OF RESEARCH PAPERS

2.3..1 Paper I: GPSeq reveals the radial organization of chromatin in the cell nucleus

As discussed in the earlier section, "Dynamics of genome organization and chromosome repositioning throughout the cell cycle", it is evident that both chromosomes and sub-chromosomal regions are non-randomly organized and occupy preferential radial positions in cell nucleus. However, at present, it remains unclear whether these preferences arise only due to specific characteristics of chromosomes such as gene density or size, or whether other factors play a role. It as well remains unknown whether transcription activity affects the radiality or *vice-versa*.

Given the fact that most of our knowledge regarding the radial positioning of chromosomes and sub-chromosomal regions comes from microscopy studies on a few selected loci (as summarized in the introduction chapter) there was a need for higher throughput techniques that can probe the radial position of many DNA loci simultaneously, ideally genome-wide. Even though several sequencing-based methods have been developed to map genome organization in human nuclei, each comes with some drawbacks as described in the introduction chapter. To overcome these limitations, we developed a technique that allows for mapping of radial locations genome-wide, termed as Genomic Loci Position by Sequencing (GPSeq) [138] . GPSeq, is based on digestion of genomic DNA in fixed cells attached to coverslips, using increasing pulses of restriction endonucleases, followed by retrieval of the restriction sites that have been cut using Next Generation Sequencing (NGS) (**fig. 7., 8A-C.**).

In order to preserve the chromatin structure, cells are fixed with paraformaldehyde (PFA) followed by permeabilization with liquid nitrogen. The permeabilization step allows for the diffusion of the restriction enzyme into the nucleus. To prevent re-ligation of the cut sites, DNA is dephosphorylated before continuing to adapter ligation. The digested cut sites are then ligated with double-stranded oligonucleotide adapters, specially designed to allow linear amplification of the cut sites and their linear genomic neighborhood by *in vitro* transcription (IVT) (**fig. 7.**). The oligonucleotide adapters contain: 1) a sample-specific barcode sequence (one for each digestion time point), enabling sequencing of multiple samples in the same sequencing run; 2) an 8-nucleotide random Unique Molecular Identifier (UMI) sequence to quantify unique reads; 3) the Illumina RA5 adaptor sequence; and 4) the T7 promoter for T7 driven IVT. The linear amplification that is mediated by T7 polymerase reduces PCR amplification biases [139] and allows only for the amplification of fragments where the cut appeared. After ligation, genomic DNA (gDNA) is extracted, purified and fragmented in order to prevent any bias that IVT might introduce due to different DNA fragments length. Next, gDNA is normalized to the lowest concentration among all of the samples and equal amounts of DNA are used as input for IVT. After IVT, the amplified RNA is then processed to prepare sequencing library separately for

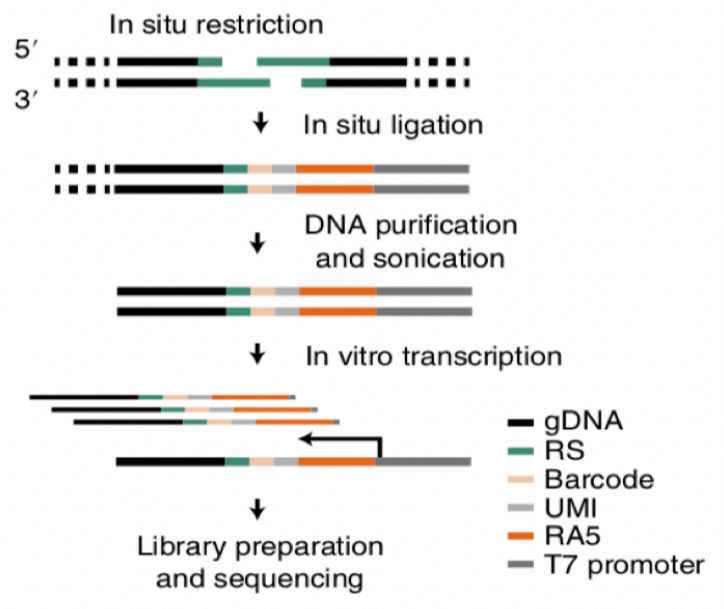

Fig. 7. Scheme of the GPSeq library preparation. Adapted from **paper I**.

each time point, which is then sequenced using the Illumina platform (NextSeq 500). The obtained sequencing data are processed through a custom-built computational pipeline in which demultiplexed sequencing reads are checked for read quality. Only reads that contain the proper adapter sequence are aligned against the reference genome. Mapped reads are then shifted, if necessary, to the nearest cut site position in the genome. Next, a UMI filtering step is performed in order to remove the PCR duplicates. The output of this analysis is a series of 'bed' files (text file that defines a feature track), one for each digestion time point, which are then used to estimate the centrality (radial rankings) of genomic loci.

Visualizing the cut sites by YFISH

To visualize the restriction enzyme diffusion, we developed a fluorescence in situ hybridization (FISH) assay called YFISH (**fig. 8.**), in which Y-shaped double-stranded adapters are ligated to each cut site and followed by hybridization of a fluorescently labeled oligonucleotide to the Y adapter. The samples are then imaged using wide-field epifluorescence microscopy or using STED (simulated emission depletion microscopy). The obtained images are deconvolved in order to improve their contrast and resolution. In order to quantify the signal coming from an increased time of enzymatic digestion, images are processed through a custom-built computational pipeline. In this pipeline, nuclei are automatically segmented in 3D by using signal coming from Hoechst 33342 staining. For each identified cell, the volume, shape, surface, flattened size (in Z-projection), sum of intensity, and average intensity are estimated. Then, only G1 nuclei are selected based on the DNA staining. GPSeq profiles are then built by calculating the mean, median, and mode of single-voxel intensity values (signal coming from digestion sites) falling in each bin of a binned normalized distance from the lamina to the center of the nucleus.

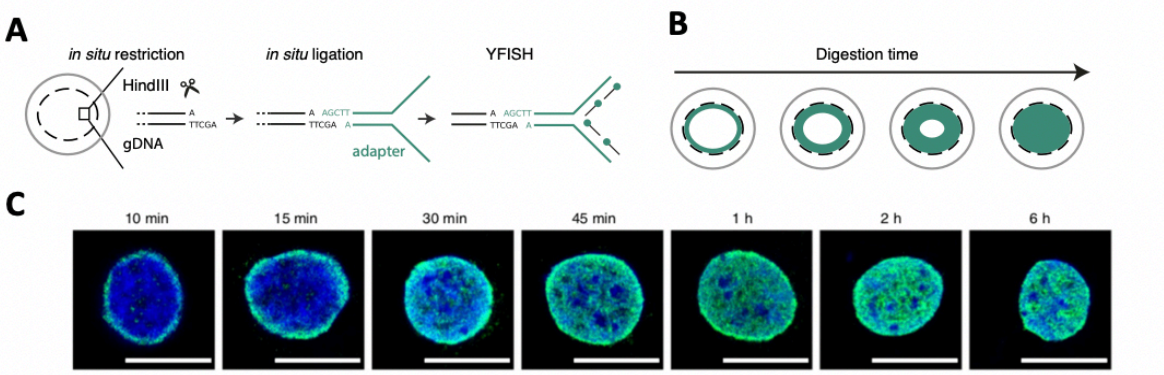

Fig. 8. GPSeq application. (**A**) Scheme of YFISH. Cells (grey circle) that were first fixed and their nuclei (dashed black circle) permeabilized are subjected to enzymatic digestion (for instance, with HindIII) for increasing times. To each cut site, introduced by restriction enzyme, the Y-shaped adapter (green) is ligated. Next, the fluorescently labelled oligos (black lines with green dots) are hybridized to their complementary sequence on the single-stranded arm of the Y-shaped adapter, which are then visualized using microscopy. (**B**) Scheme of gradual in situ digestion (green band) of the gDNA, starting from the periphery of the nucleus and progressing towards the nuclear interior. Each circle represents an individual sample. (**C**) Visual representation of the gradual gDNA enzymatic digestion obtained with STED microscopy. Green, signal coming from digestion of gDNA with HindIII. Blue, signal coming from a DNA staining. Hap1 cells. Scale bars, 10 μm. Adapted from **paper I**.

Sequencing experiments

We performed GPSeq on two cell lines, HAP1 (chronic myeloid leukemia, CML) and GM06990 (lymphoblastoid cells, LCL), of which the first is haploid and the second diploid. For HAP1 cells, we prepared two biological replicates, using either HindIII or MboI restriction enzymes, in total generating four samples. In the case of GM06990 cells, we prepared two replicates. We digested all samples either using 0.5 U/μl of the HindIII endonuclease for seven different digestion time points (10, 15, 30, 45 min and 1, 2, 6 h) or 0.5 U/μl MboI for four time points (1, 5, 10, 30 min).

Estimation of centrality

We devised three different ways of combining sequencing data coming from different time points to infer the radial positioning of genomic loci. The first approach compares only the last (2 h for HindIII, and 30 min for MboI) and the first digestion time point (10 min for HindIII, and 1 min for MboI). The second approach compares every condition with the first one, taking it as a reference point. The third approach compares each condition with the immediately previous one. The radial rankings are generated at two different resolutions, either using 1 Mb

sliding windows in steps of 100 kb (1 Mb resolution), or nonoverlapping 100 kb windows (100 kb resolution). In order to choose the metric that most accurately reflects the actual radial position of genomic regions, we compared radial rankings with 3D DNA FISH measurements (see below). After comparing the three approaches described above, we concluded that the third approach provides the most accurate estimate of radiality. Therefore, in all subsequent analyses, we used this approach and calculated what from now on we refer to as the GPSeq score.

High correlation between GPSeq and 3D DNA FISH data

In order to validate the technique, we performed 3D DNA FISH using evenly spaced 68 DNA probes [98] targeting 11 human chromosomes (one probe every 5 or 10 Mb), with each probe targeting a region of approximately 10 kb. For each probe, we measured the 3D distance of FISH dots from the nuclear edge, and then correlated it to the GPSeq score of a 1 Mb genomic window centered on the mid coordinate of the FISH probe. We found a high correlation (Pearson's correlation coefficient (PCC) =0.91 for both 1 Mb and 100 kb resolutions) between the 3D distances measured by DNA FISH and the corresponding GPSeq scores. This comparison confirmed the ability of GPSeq to detect the radial positioning of genomic loci.

High reproducibility of GPSeq score across biological replicates

Next, we looked at the reproducibility of the technique across all HAP1 biological replicates, and observed that the GPSeq score is highly consistent across all of them, at both 1 Mb and 100 kb resolution, with PCC= 0.87-0.97 (1 Mb) and PCC=0.62-0.83 (100 kb). Interestingly, we observed a high correlation between GPSeq scores obtained with HindIII and MboI, even though these two endonucleases have an opposite GC content bias.

GPSeq score anticorrelates with the DamID score

To further validate GPSeq technique, we compared GPSeq score (average among all four HAP1 replicates) with the DamID method used for calling the Lamina Associated Domains (LADs) on HAP1 cells [112]. As expected, we found a strong enrichment of these domains at the periphery of the nucleus as measured by GPSeq (**fig. 9A**), confirming the ability of GPSeq to map the radial position of genomic loci. Moreover, when we further explored the radial arrangement of LADs, we observed that constitutive inter-LAD (ciLADs) regions occupied nuclear interior, while constitutive LADs (cLADs) were found at the nuclear periphery. These observations indicate that the nuclear mid-section might be less conserved among different types of cells (**fig. 9B**).

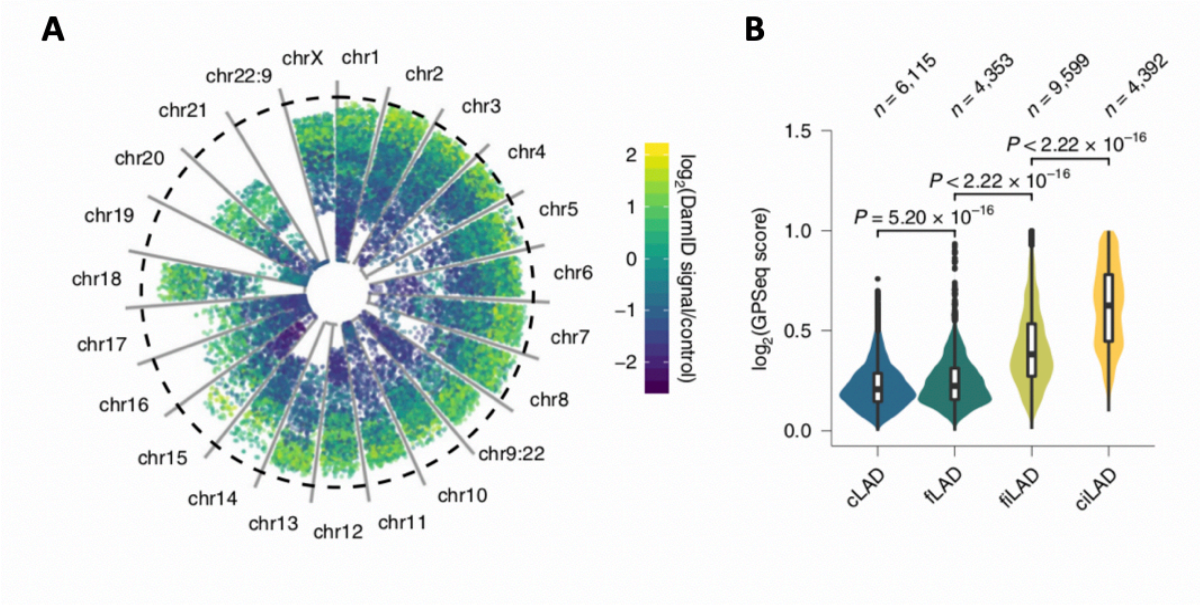

Fig. 9. GPSeq validation. (**A**) Radial distribution of lamin B DamID signal/control. Each dot represents 1 Mb genomic window. Nuclear lamina depicted as dashed black circle and nuclear center as a solid grey circle. (**B**) Radial arrangement of LADs, based on their GPSeq score: constitutive (cLADs) or facultative (fLADs) and constitutive or facultative inter-LADs (ciLADs and fiLADs). n indicates the number of analyzed genomic windows (1 Mb). Adapted from **paper I**.

GPSeq supports the idea of compartmentalization of the human genome

Next, we looked at Hi-C data coming from HAP1 cells, and compared it with the GPSeq score. We used A and B compartments coordinates obtained with Hi-C [33] and plotted them in a radial manner based on their GPSeq score. In line with previous reports (summarized in the introduction chapter), we identified regions that correspond to A and B compartments, with A compartments localized in the nuclear interior and B compartments positioned at the periphery of the nucleus with some cases of B compartments being present in the central regions.

GPSeq reveals radial positioning of certain genetic and epigenetic features

We then looked closer at the radial distribution of the A/B subcompartments (A1-2, B1-3) [18] and found that the most central B2 subcompartment is highly enriched in H3K9me3, which is a heterochromatin mark. This observation could explain that in order to resist the internal active environment, central heterochromatin needs to be densely decorated with repressive marks. We found that in general, marks of active chromatin (H3K27ac, H3K4me3), as well as gene density and expression increased towards the nuclear interior. On the other hand, H3K9me3, decreased towards the nuclear center. Moreover, when we further explored subcompartments, we found that DNA accessibility increased towards the center (**fig. 10A.**). We then wondered if there is any radial arrangement of DNA replication and using Repli-seq data, we noticed a gradual

progression of this process from the nuclear center towards the periphery (**fig. 10B.**). With these observations we demonstrated that radiality is a key aspect of the higher-order structure of the human genome.

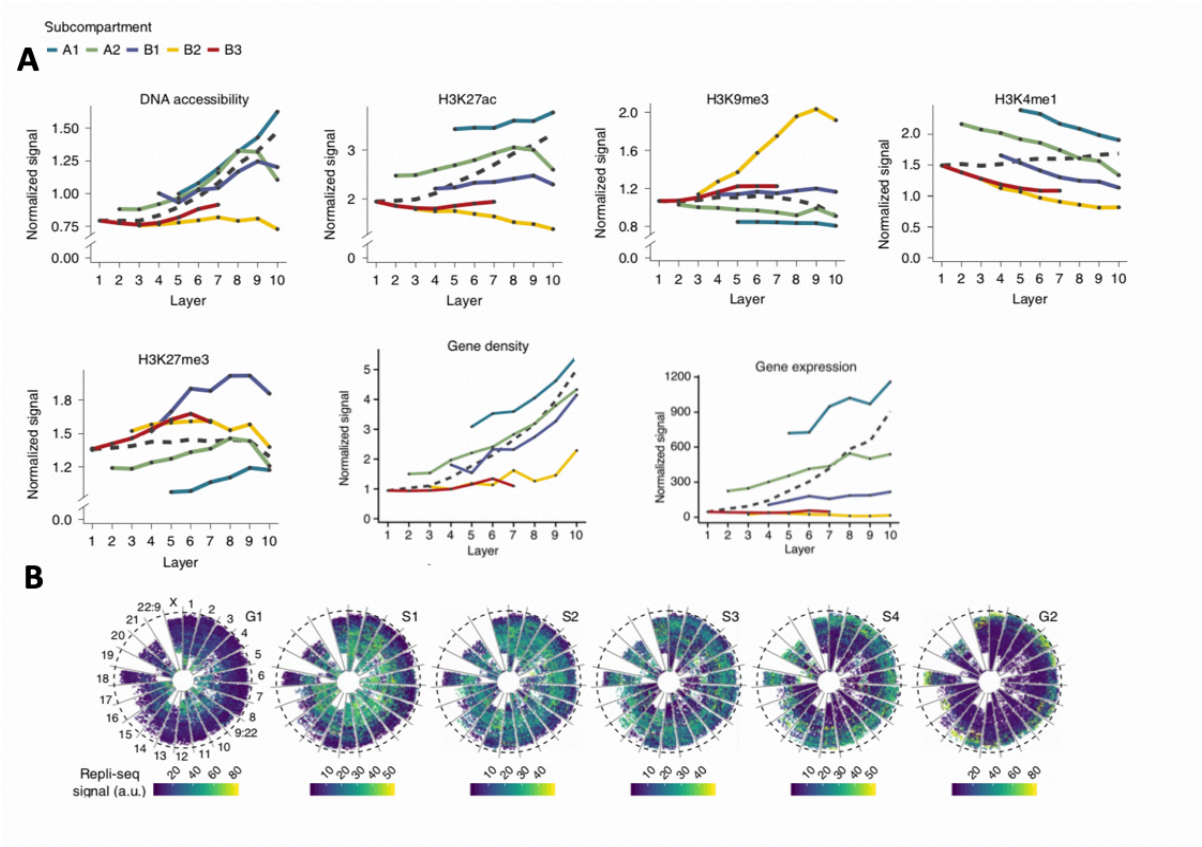

Fig. 10. (**A**) Radial arrangements of DNA accessibility, selected histone marks and gene sets in five A and B subcompartments. Radial patterns without dividing by subcompartment are depicted by the dashed grey lines. The nucleus was computationally divided into ten concentric layers based on a GPSeq score. 1- nuclear periphery, 10- nuclear interior. (**B**) Radial arrangements of the Repli-seq signal. Adapted from **paper I**.

Chromosome size, gene density, gene expression and GC content as contributors to the chromosomal positioning

Next, we were curious what contributes to the radial position of genomic loci. As discussed in in the introduction chapter, there are currently two models that try to explain this. The first model states that chromosomes are positioned based on their size, with big chromosomes occupying peripheral regions of the nucleus, whereas small ones being located in the nuclear interior [47–49]. The second model proposes that the radiality is the outcome of the gene density, with gene poor chromosomes being localized at the outer regions of the nucleus, and the center being mostly occupied by gene-rich chromosomes [50,51]. We observed that on a chromosome level, GC content together with chromosome size were the highest predictors of the radiality, with no added value of using either gene expression or gene density ($R^2 = 0.939$). However, at the sub-chromosomal level (1 Mb resolution), the most accurate model of the radial positing

of genomic loci included GC content, gene density, gene expression, and chromosome size (R^2 = 0.741). Taking all these studies together, we identified the predictors of the radial positing of genomic regions, both at a whole chromosome level as well as at 1 Mb resolution. These observations provide an evidence that the genome organization cannot be directly explained by either chromosome size or gene density, as it was speculated before, but by a combination of both of these features together with gene expression and GC content.

Reconstruction of 3D genome

To further strengthen the power and potential of our technique together with our will to discover new biological insights, we investigated whether GPSeq combined with Hi-C could improve whole-genome reconstructions, previously obtained with Hi-C [21]. For this purpose, we developed an algorithm called *chromflock* that generated 10,000 single-cell structures at 1 Mb resolution, either using Hi-C (H) only or Hi-C together with GPSeq (HG). We demonstrated that the correlation of the genome structures with DNA FISH was substantially higher for HG structures, compared to H structures**.** Furthermore, we applied *chromflock* to look at the radial organization of A/B subcompartments and observed that in HG- but not in H-structures, A1 and B1 subcompartments were positioned as the most central, followed by A2 and B2, whereas B3 was found to occupy nuclear periphery (**fig. 11A-B.)**. These observations provide an evidence that GPSeq significantly improves the information provided by Hi-C only, therefore allowing for a more precise reconstruction of the 3D genome in human cells.

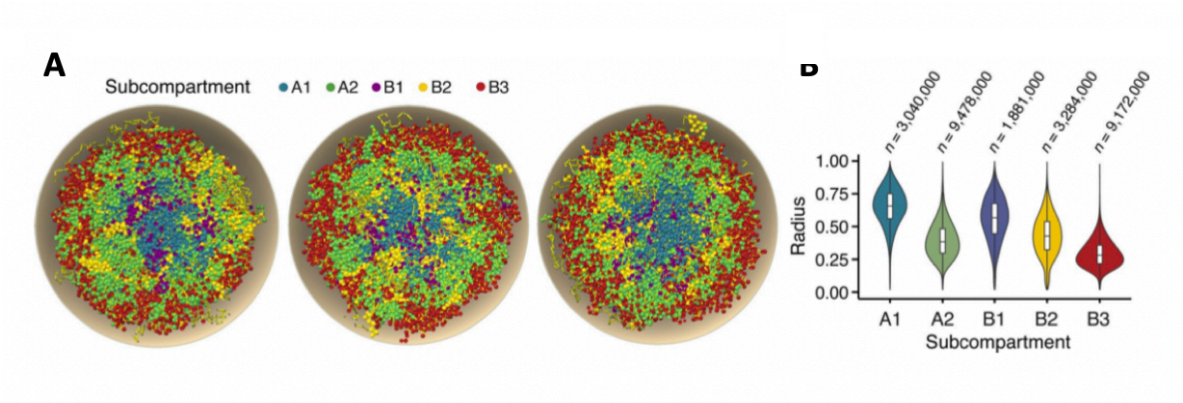

Fig. 11. Radial distribution of A and B subcompartments in *chromflock* structures. (**A**) 3 out of 1,000 HG structures generated by combining Hi-C together with GPSeq. Each bead represents a 100 kb genomic window. Nuclear envelope is depicted in grey. (**B**) Radial arrangement of A and B subcompartments in 1,000 HG structures. 0- indicates nuclear periphery, 1- nuclear center. *n*, number of beads corresponding to each A and B subcompartment. Adapted from **paper I**.

GPSeq reveals genome fragility

With the proposed model in 1975 by Dr. Hsu that heterochromatin serves as a 'bodyguard'- a protector of the euchromatin from the DNA damage [8], we applied GPSeq to investigate the radial patterns of DNA breaks and various mutations (**fig. 12.)** . For this purpose, we used the BLISS technique that we previously developed to precisely map genome-wide landscape of DSBs [140] (**paper II**). We demonstrated that DSBs frequency increased towards the center of a nucleus for both genic and intergenic regions (**fig. 12A-B.)**. When we looked at the regions that are frequently translocated in cancers, we found that they are mostly central and are characterized by the higher accumulation of DSBs when compared to untranslocated regions (**fig. 12C.)**. We further took advantage of our *chromflock* and investigated further HG structures. We demonstrated that parts of the genome that intermingle the most, are the ones that carry the highest number of genes involved in gene fusions (**fig. 12D.)**, which was not observed in structures generated with Hi-C only, therefore highlighting the sensitivity of GPSeq.

In addition to DSBs and gene fusions, we also investigated the radial arrangement of both single-nucleotide polymorphisms (SNPs) and cancer-associated single-nucleotide variants (SNVs) (**fig. 12E-F.)**. Previous studies reported that the frequency of these two mutation types are higher in heterochromatin [141]. We found that SNVs frequency was higher at peripheral regions (**fig. 12E.)**, supporting the 'bodyguard' hypothesis, whereas SNPs increased towards the nuclear interior (**fig. 12F.)**, which might correspond to the heterochromatic segments found in the small chromosomes that are located in the nuclear interior.

Trying to reveal as many new aspects of the genome fragility as possible, we looked at the radial arrangement of genes that are involved in the DNA damage process (**fig. 12G.)**. We showed that genes that are either upregulated upon UV damage or are involved in the repair of DSBs, are enriched in the nuclear interior. On the other hand, genes that are downregulated upon UV damage, were found to be enriched at the nuclear periphery.

Thanks to GPSeq, these observations add spatial information to the genomic landscape of DSBs that we previously obtained with BLISS (**paper II**).

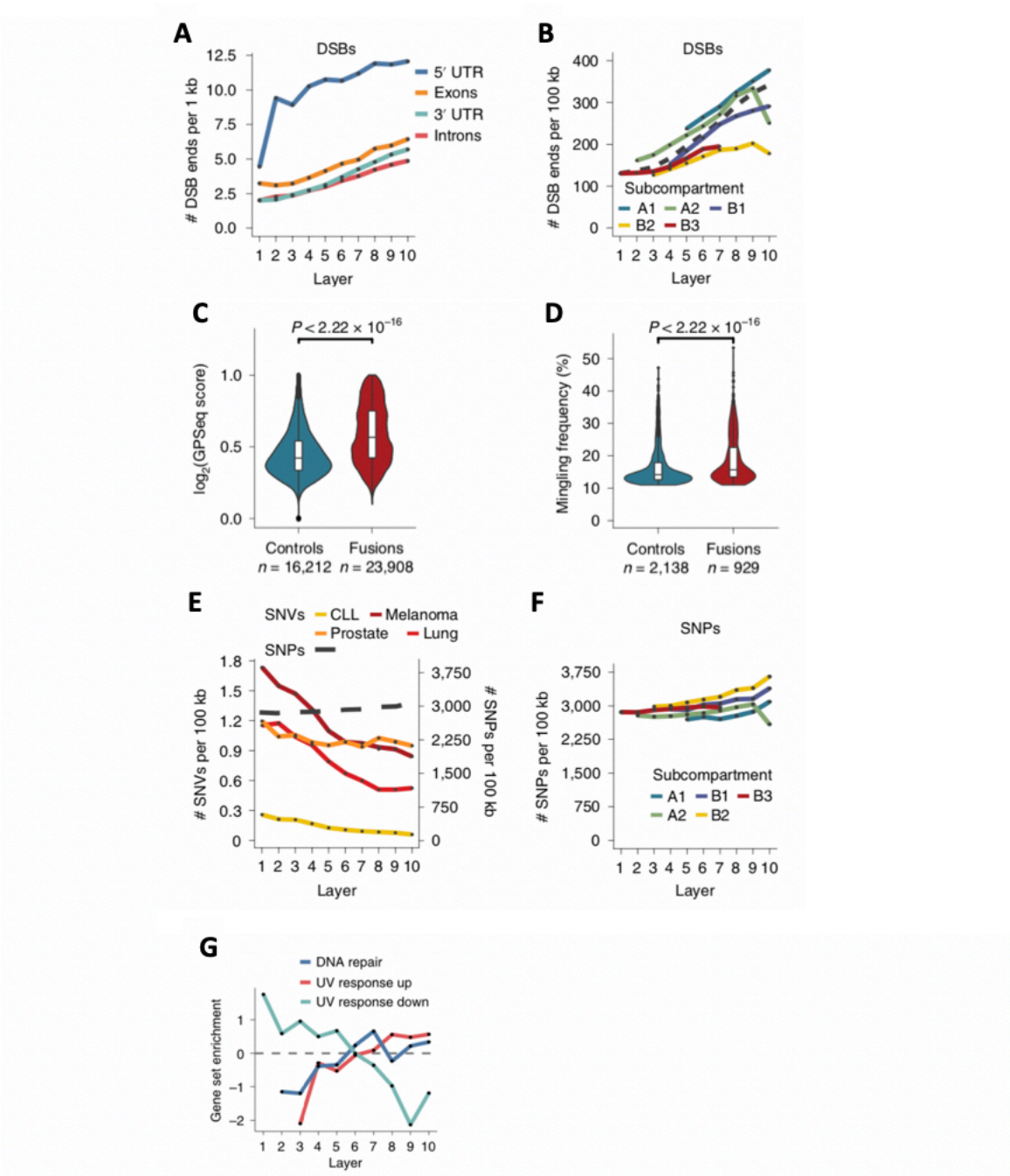

Fig. 12. Radial organization of mutations, DSBs and selected genes. (**A**) Radial arrangement of DSBs for genic and intergenic regions. (**B**) Radial patterns of DSBs in A and B subcompartments. Dashed line represents the arrangement of DSBs without dividing by subcompartment. (**C**) Radial arrangement of genomic loci that are either involved in cancer-associated fusions or not (controls). n indicates the number of analyzed genomic windows, at 100 kb resolution. (**D**) The frequency of intermingling across genes that either fused or not. n means the number of analyzed beads in the HG *chromflock* structures. (**E**) Radial arrangements of SNVs in different cancer types and of SNPs. (**F**) SNPs radially distributed across A and B subcompartments. (**G**) Radial distribution of selected genes involved in the DNA damage process. Adapted from **paper II**.

2.3..2 *Paper II: BLISS is a versatile and quantitative method to map DNA breaks in low-input cell and tissue samples*

To accurately measure the location and frequency of DSBs across the genome is fundamental to understanding the genome fragility, however the already existing methods are limited with regards to their sensitivity and applicability [142–144].

In BLISS, that stands for Breaks Labelling *In Situ* and Sequencing, in order to preserve the chromatin structure, cells or tissue sections are fixed with PFA (**fig. 13.**). Next DSBs are blunted with T4 DNA polymerase and then the blunted DSB ends are ligated with double-stranded oligonucleotide adapters, specially designed to allow linear amplification of DSB sites and their linear genomic neighborhood by IVT. The oligonucleotide adapters contain: 1) a sample-specific barcode sequence (one for each sample), enabling sequencing of multiple samples in the same sequencing run; 2) an 8-nucleotide random UMI sequence to quantify unique reads; 3) the Illumina RA5 adaptor sequence; and 4) the T7 promoter for T7 driven IVT. The linear amplification that is mediated by T7 polymerase reduces PCR amplification biases and allows only for the amplification of fragments where the DNA break appeared. After ligation, gDNA is extracted, purified and fragmented in order to prevent any bias that IVT might introduce due to different DNA fragments length and used as input for IVT. After IVT, the amplified RNA is then processed to prepare sequencing library separately for each time point, which is then sequenced using the Illumina platform (NextSeq 500). The obtained sequencing data are processed through a custom-built computational pipeline in which demultiplexed sequencing reads are checked for read quality. Only reads that contain the proper adapter sequence are aligned against the reference genome. Next, a UMI filtering step is performed in order to remove the PCR duplicates. The output of this analysis is a series of 'bed' files, one for each sample, which are then used to estimate the genomic location and frequency of DSBs.

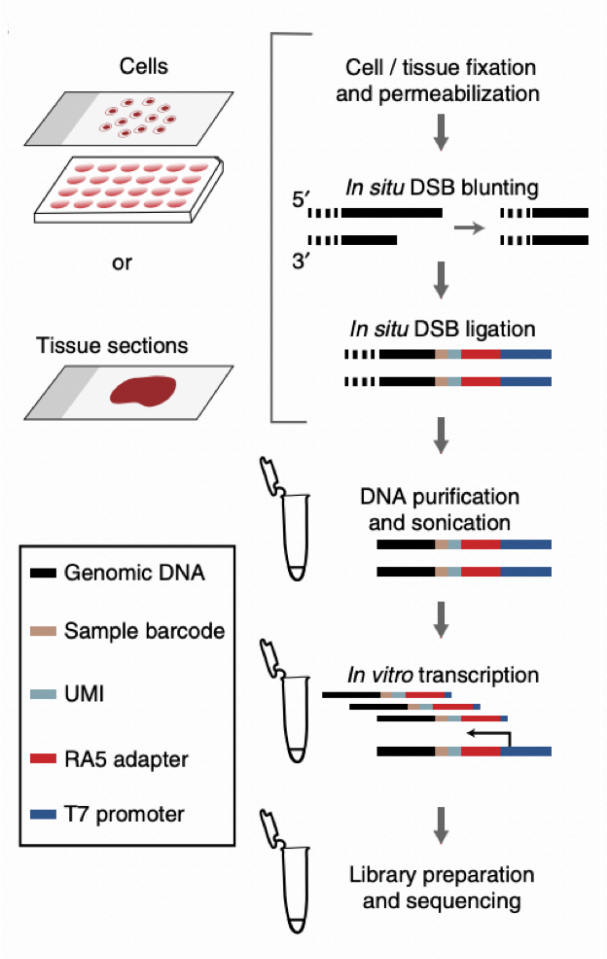

Fig. 13. Scheme of BLISS. Adapted from **paper II**.

In order to show the applicability of BLISS to different sample types, we applied this technique to human cell lines, mouse primary ESCs and tissue sections as well as purified nuclei obtained from mouse liver biopsies.

Validation of BLISS

In order to validate the technique, we performed the immunofluorescent staining of γH2A.X in KBM7 cells and demonstrated that BLISS can accurately quantify endogenous DSBs (80-100) in low-input material (<5000 cells). BLISS detected 80-100 DSBs, given that the number of γH2A.X foci was on average 85.7 (**fig. 14A-C**).

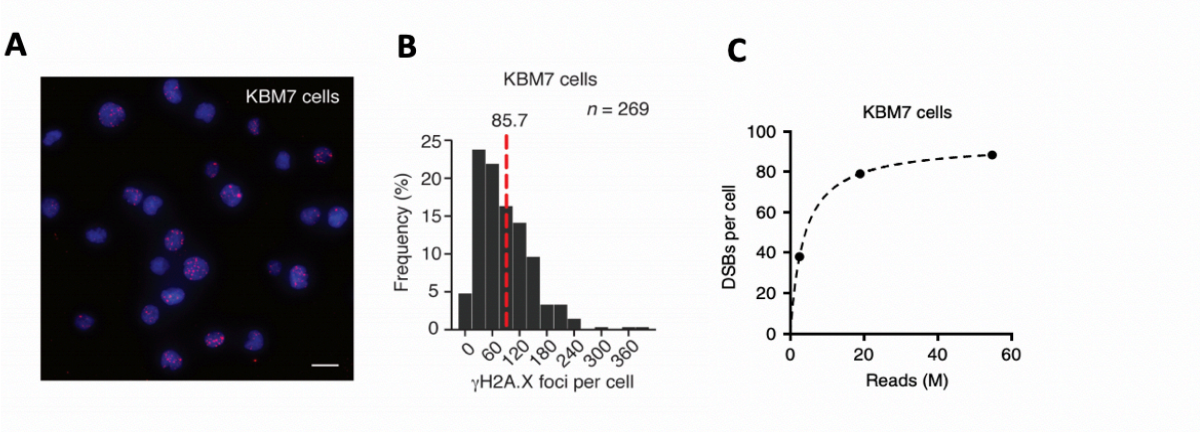

Fig. 14. Quantitative detection of DSBs in KBM7 cells. (**A**) Microscopy image of the immunofluorescence staining of γH2A.X foci (red). Blue- DNA stained with DAPI. Scale bar, 20 μm. (**B**) Number of γH2A.X foci per cell. n means number of analyzed cells. The number located above the red dashed line indicates the mean number of γH2A.X foci per cell. (C) Estimation of DSBs at increasing sequencing depth. Dashed line represents a hyperbolic interpolation. Adapted from **paper II**.

To further validate BLISS, we used UMIs to estimate DSBs induced with etoposide which is a topoisomerase inhibitor. We observed that upon etoposide treatment in U2OS cells, DSBs were enriched in the neighborhood of the TSS, which is in agreement with previous studies[145]. That accumulation of breaks is likely due to the role of TOP2 in resolving DNA entanglements in response to torsional stress generated by transcription, replication or chromatin compaction as described in the introduction chapter.

Applicability of BLISS

In order to show the applicability of BLISS to different sample types, we applied this technique to human cell lines, mouse primary ESCs and tissue sections as well as purified nuclei obtained from mouse liver biopsies. Using BLISS, we showed that the endogenous DSBs are enriched in both the neighborhood of TSS and in highly transcribed genes which is in line with others observations performed on cell lines [63,146,147] (**fig. 15A-B.**). However, using BLISS, for the first time ever, we directly delivered an evidence that in solid tissue- primary mouse liver tissue-

DSBs are enriched in the promoter region of highly expressed genes (**fig. 15C-D.**). Moreover, we found that genes with the highest enrichment of DSBs in liver tissues were involved in metabolic functions according to the Gene Ontology analysis. These observations indicate the power of BLISS to capture endogenous DSBs assocaited to tissue-specific processes and reveal the biological basis of the gene fragility.

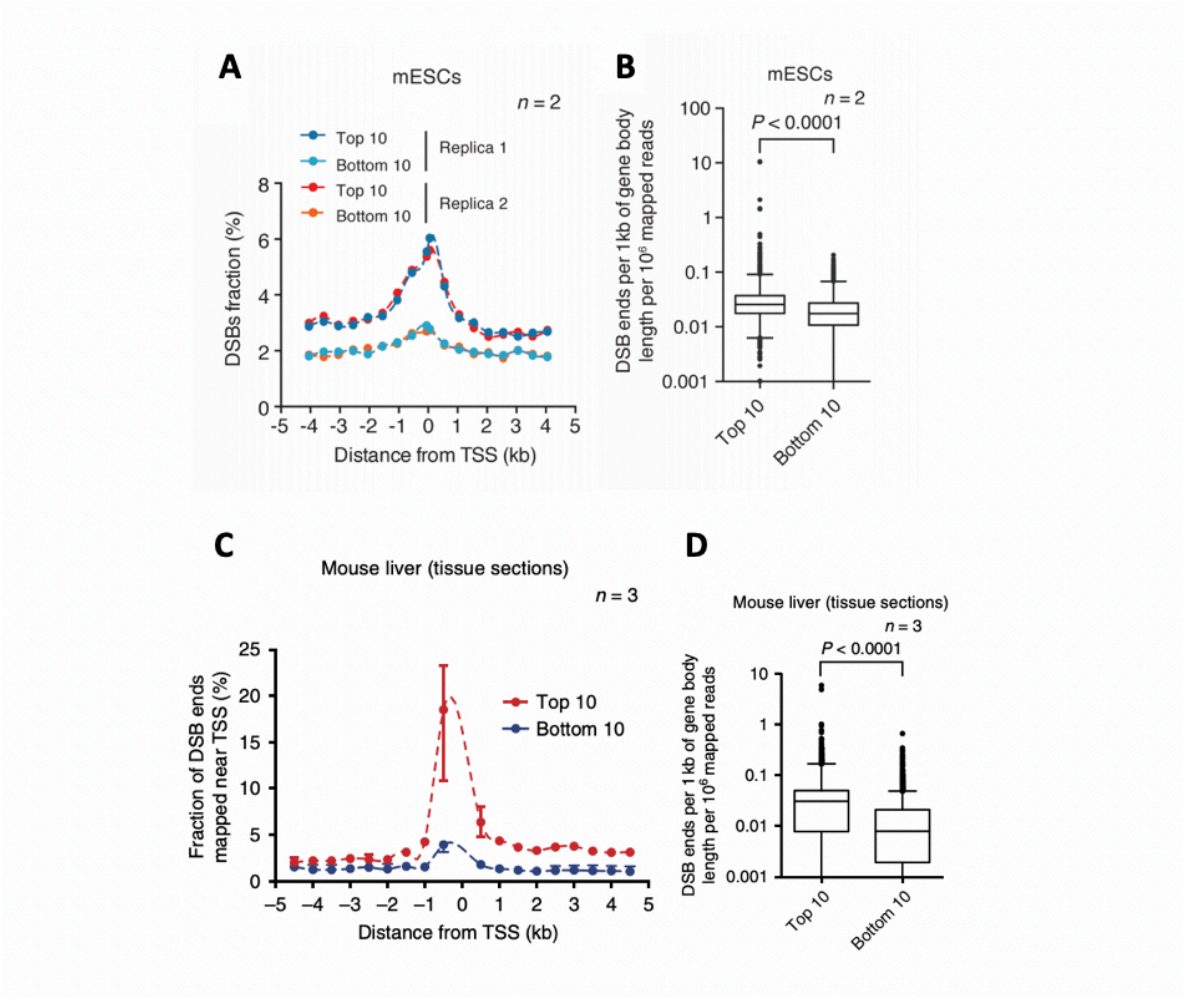

Fig. 15. Quantitative detection of endogenous DSBs by BLISS. (**A**) The percentage of the DSB ends that were mapped ±4.5 kb around the transcription start sites (TSS) of the top 10% and bottom 10% expressed genes of mESCs. (**B**) Number of DSB ends mapped per kb inside the gene body of the top 10% and bottom 10% expressed genes in mESCs. (**C**) Fraction of DSBs ends in mouse liver, mapped around TSS that were found in the top 10% (red) and bottom 10% (blue) of expressed genes. Dashed lines indicate the spline interpolation. (**D**) Same as (**B**) but in mouse liver. n represents the number of biological replicates. Adapted from **paper II.**

Benchmarking of BLISS for CRISPR/Cas9 and CRISPR/Cpf1 off-target detection

To further show the power and sensitivity of the technique, we used BLISS to characterize genome-wide the specificity of CRISPR (clustered regularly interspaced short palindromic repeats) nucleases that greatly improved genome editing field [148]. Since the off-target cleavage DNA activity can lead to some mutation events [149,150], it is important to thoroughly estimate the nucleases' activities before these enzymes can be safely applied to the clinical human genome editing. For this purpose, we used two CRISPR-associated RNA-guided

endonucleases, Cas9 [148] and Cpf1 [151]. First, we studied transfected HEK293 cells with *Streptococcus pyogenes* Cas9 (spCas9) and sgRNAs targeting EMX1 and VEGFA genes that were previously identified by other CRISPR off-target screening methods such as BLESS [142], GUIDEseq [143] or Digenome-seq [144]. We demonstrated that BLISS identified among many previously detected by BLESS off-targets, numerous of new off-target sites (**fig. 16A.**). When we compared side-by-side BLISS together with GUIDEseq and Digenome-seq, we revealed that even though there is a high level of agreement between these three technologies with regards to the identification of top off-targets, there are some differences at the level of low-frequency off-target sites, especially in the case of VEGFA gene (**fig. 16B-C.**). Next, we aimed to estimate the sensitivity of BLISS by identifying DSBs induced by both Cpf1 from *Acidaminococcus sp*. (AsCpf1) and *Lachnospiraceae bacterium* (LbCpf1), and showed that both nucleases are highly specific and that they outperform Cas9 when it comes to the off-target activity (**fig. 17.**), which is in line with the recent studies [152].

Overall, these results demonstrate that BLISS is a versatile and quantitative technique that allows for an efficient genome-wide mapping of DSBs profiles in low-input samples of both cells and tissue. The power of BLISS over other methods that are used in the genome engineering community to detect DSBs, lies in its applicability for tissue samples, making it valuable for studies about CRISPR nucleases in clinically relevant samples. The other advantage of BLISS over existing techniques is in the adapter design that among other features, includes UMIs that leads to the robust discrimination of DSBs. Due to its protocol, as described above, BLISS is cost-effective and less laborious or time-consuming than other techniques such as BLESS. Altogether, we are convinced that BLISS is a robust method for genome-wide DSB sequencing that can be beneficial for both DNA damage and genome editing fields.

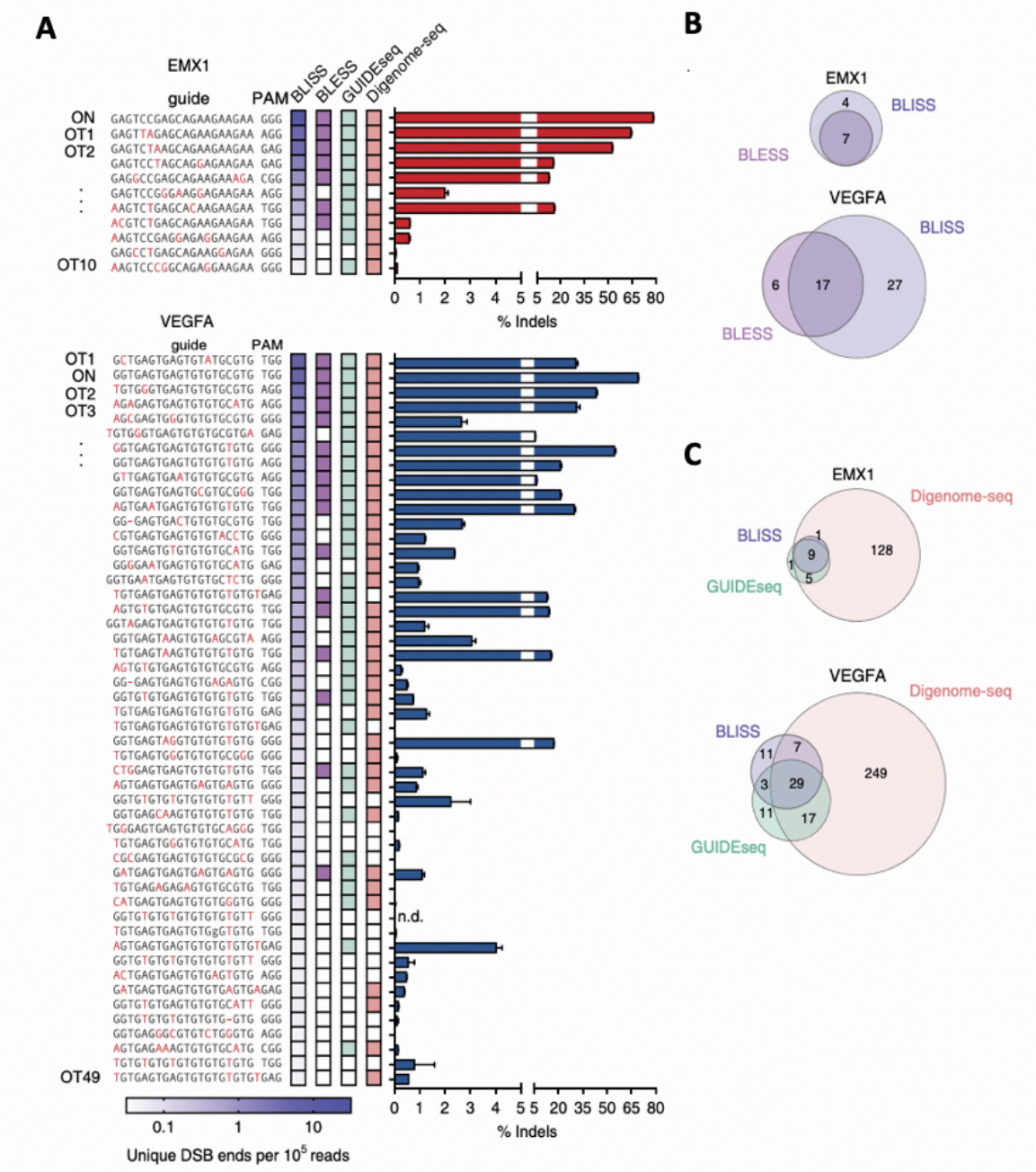

Fig. 16. Assessment of BLISS sensitivity performed through the genome-wide quantification of SpCas9 on- and off-target DSBs. (**A**) On- and (ON) off-target (OT) sites detected by BLISS, BLESS, GUIDEseq and Digenome-seq. Colored squares depict if the BLISS target was previously described by either of the above-mentioned methods. Each of the individual site was validated by using the targeted deep sequencing which resulted in the estimation (%) of reads including an insertion or deletion (indel). (**B**) Overlap between ON and OT sites found by BLISS versus BLESS. (**C**) Overlap between ON and OT sites detected by BLISS versus GUIDEseq and Digenome-seq. Adapted from **paper II.**

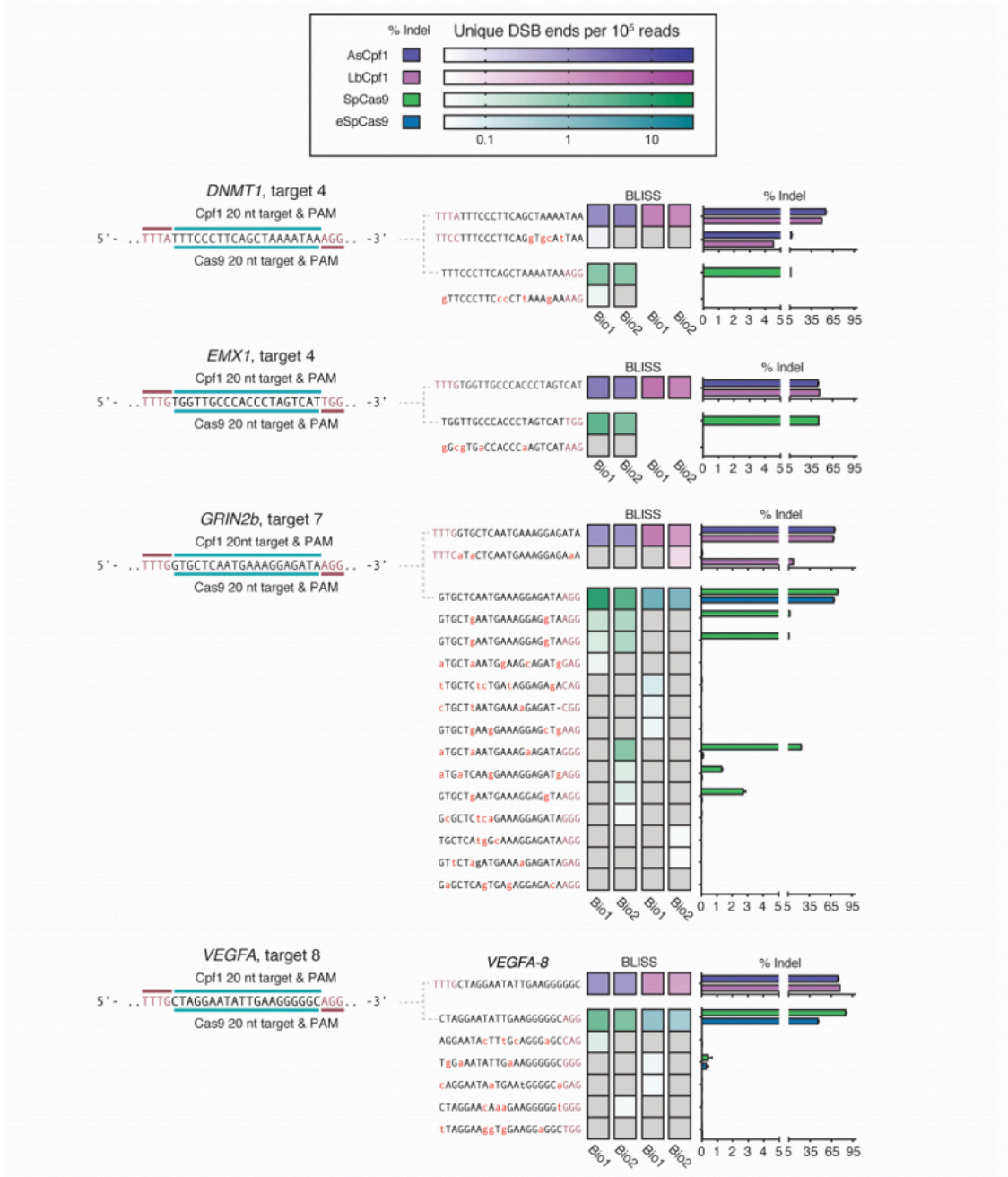

Fig. 17. Comparison of the results obtained by BLISS technique, with guides that were targeted by AsCpf1, LbCpf1, SpCas9, and eSpCas9. Each of the individual site was validated by using the targeted deep sequencing which resulted in the estimation (%) of reads including an insertion or deletion (indel). Adapted from **paper II.**

3 DISCUSSION AND CONCLUSIONS

The work included in this thesis reports the development of two novel genome-wide methods: GPSeq to map 3D genome organization in human cell nucleus (**paper I**) and BLISS, to profile the genomic landscape of DSBs in eukaryotic cells (**paper II**).

In conclusion, we developed GPSeq, a novel sequencing-based technology to measure genome-wide radial position of genomic loci with regards to the nuclear periphery/center. As summarized in the introduction chapter, most of our knowledge about the radial arrangement of chromatin comes from the microscopy studies (DNA FISH) of a few selected loci, therefore technique like GPSeq adds important and missing information to the growing field of the 3D genome structure. Unlike GPSeq, that profiles genome radiality across the entire nuclear radius, prior sequencing-based techniques that probe radial aspects of genome organization, are limited to certain genomic regions. For instance, with 10 μm being an average diameter of the human nucleus, TSA-seq [107] is limited to 1.5 μm radius, similar to DamID [108–110] that can only accurately estimate the spatial position of chromatin when genomic loci are close to the nuclear lamina. The other advantage that GPSeq has over other methods is the amount of starting material needed for the experiment, being on average 500 thousands of cells, whereas for TSA-seq [107] and SPRITE [106] require many millions of cells. Unlike GPSeq, TSA-seq seems to be both costly and time-consuming, as it requires sequential antibody staining of different batches of cells for over 1 month. Another limitation that SPRITE possess is the length of the crosslinked fragments (up to 1 kb), that could lead to artifacts during library preparation, thus eventually lowering the representation of inter-chromosomal contacts. The radiality could be as well estimated using the GAM method [92], however its resolution depends on the number of cut nuclear slices making it unclear whether this technique could accurately assess the radiality at high resolution. Another limitation of GAM is the time that is needed to prepare the material, including cryosectioning of cells and nuclei laser microdissection, therefore limiting its application to a high number of cells. Another drawback is the fact that this technology requires that all cells have relatively consistent shape to avoid a scenario that some slices could have more and other less of the genome. The other technologies that can be used to model radial genome structures are Hi-C [21] and Dip-C [105] but these two methods compared to GPSeq, are costly, laborious and time-consuming. Taken together, GPSeq overcomes aforementioned limitations making it the first novel technique to profile genome radiality across the entire nuclear radius in a highly reproducible manner.

Using GPSeq, we supported the idea of compartmentalization of the human genome into A and B compartments and revealed relevant and novel biological insights such as the radial positioning of genomic and epigenomic features. We showed that in general, marks of active chromatin (H3K27ac, H3K4me3), DNA accessibility, as well as gene density and gene expression increase towards the nuclear interior. On the other hand, repressive mark H3K9me3 decreases towards the center. However, when we explored A/B subcompartments we demonstrated that the radial organization of various chromatin traits follows unique patterns,

for instance, the most central B2 subcompartment is highly enriched in H3K9me3 which could be due to the fact that in order to resist the internal active environment, central heterochromatin needs to be densely decorated with repressive marks. Interestingly, we showed a gradual progression of replication from the nuclear center towards the periphery.

Moreover, using our newly developed *chromflock* algorithm, we demonstrated that Hi-C when combined with GPSeq, allows for a more precise reconstruction of the 3D genome in human cells, compared to information provided by Hi-C only, strengthening further the power and sensitivity of our technique.

Using GPSeq, we identified the predictors of the radial positing of genomic regions, both at a whole and sub-chromosome level. We provided an evidence that the genome organization cannot be predicted by either chromosome size or gene density alone, as it was speculated before, but by a combination of both of these features together with gene expression and GC content.

Using BLISS, we were able to explore the genome fragility, showing that in cell lines as well as in tissue, TSS neighborhoods and highly transcribed genes are vulnerable to double-strand breaks. We showed the sensitivity of BLISS by estimating off-target activity of two nucleases-Cas9 and Cpf1 in CRISPR system and demonstrated that Cpf1 is more specific when compared to Cas9. In combination with GPSeq, we were able to achieve for the first time ever, a genome-wide map of the radial distribution of DSBs. We demonstrated that DSBs frequency increased towards the center of a nucleus. When we looked at regions that are frequently translocated in cancers, we found that they are mostly central and are characterized by higher accumulation of DSBs when compared to untranslocated regions. We found that SNVs frequency was higher at peripheral regions, supporting the 'bodyguard' hypothesis [8], whereas SNPs increased towards the nuclear interior, which might correspond to the heterochromatic segments found in the small chromosomes that are located in the nuclear interior. This might explain different tendencies of heterochromatin positioned at different radial locations to undergo various mutational processes. We showed that genes that are either upregulated upon UV damage or are involved in the repair of DSBs, are found at the nuclear interior.

In summary, this thesis highlighted the development and application of two novel genome-wide techniques that allows to explore both the genome organization and genome fragility in the 3D space of the nucleus. We envision that the GPSeq design could be applied to radially map proteins or chromosomal interactions, thus improving existing methods such as ChIP-seq or Hi-C. Since we showed that GPSeq can be applied to various cell types, we believe it could be applied to tissue, opening up the possibility of exploring the chromatin landscape in clinical samples. It would be interesting to apply GPSeq to biologically relevant cell models to monitor genome architecture changes, for example during aging or cell differentiation. On the other hand, the possible development of single-cell BLISS could advance our understanding of how DSBs are distributed on a single-cell level. We believe that further use of both GPSeq and BLISS by the scientific community, can broaden our understanding of genome instability, DNA repair choices and overall 3D genome landscape.

4 ACKNOWLEDGEMENTS

The time has come to publicly acknowledge everyone who supported me throughout this journey, despite my sarcastic nature ☺

First of all, I would like to thank **Magda Bienko**, my main supervisor, who believed in me and gave me a chance to taste the method development lab. Having spent so much time in your group, Magda, stimulated my creativity and curiosity on different levels, which eventually contributed to my professional and personal development. I really appreciate the time you spent on showing me how important it is to be detailed-orientated and to deliver the high-quality data and yet still having fun.

Nicola Crosetto, you have been always around and supported my dream to become a CEO of a sequencing-based company one day ☺ Magda's and yours approach to science is contagious and it has always impressed me how one can be so passionate about the research.

I would like to thank my dissertation committee members, **Dr. Marianne Farnebo**, **Dr. Theodoros Foukakis**, **Prof. Masood Kamali-Moghaddam**, my opponent **Dr. Andreas Lennartsson** and the chairperson, **Dr. Vicent Pelechano** for finding time to read my thesis and being here today to discuss my research with me. I would also like to thank my co-supervisor, **Prof. Björn Högberg**.

The Bienko and Crosetto group members:

Reza, you were the first one I met in the lab. Not necessarily the best choice to experience the first day but what to do ;) I cannot express how grateful I am to had a chance of interacting with you and call you my friend. Despite your dry-humor, your positive attitude, laughter and business-orientated mindset were a huge support when my experiments failed. **Nana**, I cannot believe I say it, but, I envy you! ☺

Quim, Quin, Michi's doppelganger, we have gone through a lot and only you know what it took to make GPSeq worked ☺ Hong-Kong, Berlin…only Tokyo left. All of these memories, ahh…Being called your minion was a true pleasure, I am getting emotional ☺ And on top of that, you got my sarcasm, what else to ask for, haha! Te amo!

Michi, the proudest Perugian that ever walked on this planet. I will never get how an Italian person has no food-standards, well, I get it's all about pasta but…. Your ciuffo and your special dance when you write your name in the air with your culo…these are the things that, I am afraid, will stick with me forever. I am really impressed that you survived with my sarcasm for such a long time and yet being helpful as hell! Me loves you mucho! ☺

Maud, min Chihuahua, you're truly something. It's interesting to see how an individual of your height, can make so much of noise. You have been always around in and outside of the

lab, listening to all of my complains and TMIs. You're a true friend and keep on drinking milk! ☺

Ning, crabs-crepes…who would even care? ☺ It was so nice to spend the time with you and laugh at all silly things and co-share the excitement of the libraries prep some late nights. Emily, one day I will speak to you in Swedish ☺

Xiaolu, the happiness you spread was amazing to experience. Knowing that you were always keen to help and trouble-shoot anything related to the library-prep, was just great, thank you for that! Good luck with your new lab, big boss!

GG, thanks for all your effort to make GPSeq work well from the computational thing of view and also thanks for introducing us to both, your home-made pizza and Fuori di Pizza!

Masa, the big boss of Sony corp., I wish that one day I would understand what you're trying to say without Quim's help, hah! ☺ All of these memories with you, **Emma**, **Meg** and **Anne**-dinners, bouldering, oh guys, my okane family. Thank you for that! ☺

Erik, thanks for introducing me to Surströmming and for all the nice chats in the lab!

Emma, both your sense of humor and your kindness were so needed on so many occasions, including all for the administration work! Thanks for always being here to listen to me! ☺

Ana, after so many years I've learnt how to spell your name properly ☺ Thanks for all of chats and laughter!

Eleni, Gelato, all of these nice moments we shared and laughter we had. Ti nafto… ☺

Merula, gosh, your 'disappointment o'clock', 'snack time', laughter we shared across so many occasions…those are the memories I will cherish forever! GPSeq would have not been the same without you! And you're a true friend! ☺

Britta, the avocado lover, thanks for all your help and cheering me up wherever it was needed!

Xinge thanks for all the chats and laughter!

Katta, the only thing I regret is that you arrived to our lab so late, by the end of my PhD ☹ But having a chance to experience your sense of humor, positiveness and an enormous support was truly amazing!

Ele-Elo, **Eleonora**, how you survived me is a mystery to me, haha. So happy to get a chance to meet you. There is a bright future ahead of you, I can tell!

Big thanks to everyone from the BiCro labs: **Silvano**, **Federico**, **Su**, **Luuk**, **Roberto**, **Thu**, **Solrun**, **Gustaw**, **Marcin**, **Carla**, **Xiaoze**… I hope I did not forget about anyone. If I did, well, that's life, haha!

Banu and **Alba**, words cannot express what role you play in my life. Being a part of the Bermuda triangle is a true privilege, love you so much! ☺

Angelo, Hercules, thanks for introducing me to Barry's and all efforts to put me in shape ☺

Dörte, thanks for all of parties- Halloween, Trädgarden etc. And big thanks to you and your parents for helping with dealing of Anna Netrebko's concert cancelation! ☺

Professor **Jing**, you're a nocturnal researcher that helped me to better deal with all of the late evenings spent in the lab. Our Max, Thai, Pizza dinners and laughter about us and our future was just a pure joy! Thank you for that! And I know that I'm hotter than Pawel, haha ;)

Ray, thanks for laughter! **Hannes**, thanks for all dinners and parties we experienced together! Big thanks to everyone from the Elsässer lab: **Simon**, **Birthe**, **Ruth**, **Lea**, **Philipp**, **Carmen**, **Anna-Maria**, **Kyle**, **Lorenzo**, **Rozina** (Cartier and diamonds), **Maria**, having you as my neighbors was a true fun!

Also big thanks to the other members of the Genome Biology division: **Dimitris**, **Jaime**, **Valeria**, having you around made everything easier! **Ann-Sofie**, **Maria**, **Pelle**, **Asimina**, **Louise**, **Li**, **Dani**, thanks for all chats! ☺

The KTH folks, **Aman**, **Sharath** and **Tharagan,** such a joy to had you around!

Big thanks to **Flor**, **Cam**, **Sabrina** for all your support! ☺

Joan, my best flatmate! For all of laughter and support, mucho thanks! ☺

My scientific journey would have not been possible if **Eva Klein** did not give me a chance to join hers and **George** group, which I will be forever thankful for! The way you look/ed at life change the way I perceive it, thank you for it! To everyone from the Kleins group: **Harsha**, **Twana**, **Mushtaq**, **Benedek**, **Dani**, **Tanya**, **Hyrettin**, **Barbro** and **Mia**, big thanks! ☺

Well, I guess to not make my family disappointed I would have to kindly thank them too, so here it is:

Wielkie dzieki za Wasze wsparcie, placze, smiechy i pozostale emocje, ktore towrzyszyly mi podczas mojego doktoratu! Babcio, Cociu Danusi i Janko, za Wasza ciagla wiare we mnie! Reszta, bezimmiena, wraz z Sitka, wiecie o kim mowa, tez w tych podziekowaniach jest uwzgledniona! Pawel, Twoja okladka jest najlepsza na KI! Mucho kocham wszystkich i besos! ☺

I guess there is many names that I might have forgotten to mention but hopefully, once the corona is gone, I will thank you in person!

5 REFERENCES

1. Schneider, R. & Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* **21**, 3027–3043 (2007).

2. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1897–1901 (1976).

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

4. Nishino, Y. *et al.* Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.* **31**, 1644–1653 (2012).

5. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, (2017).

6. Heitz, E. Das heterochromatin der moose. *Bornträger* (1928).

7. Dillon, N. & Festenstein, R. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet.* **18**, 252–258 (2002).

8. Hsu, T. C. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* **79 Suppl**, 137–150 (1975).

9. Gilbert, N. *et al.* Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555–566 (2004).

10. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3 (2015).

11. Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* **28**, 1–13 (2007).

12. Solovei, I. *et al.* LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598 (2013).

13. Solovei, I. *et al.* Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356–368 (2009).

14. Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).

15. Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y. & Malnic, B. Nuclear compartmentalization of odorant receptor genes. *Proc Natl Acad Sci U S A* **111**, 2782–2787 (2014).

16. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).

17. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

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

19. Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat. Rev. Cancer* **14**, 389–393 (2014).

20. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).

21. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

22. Zhang, H. *et al.* Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* **576**, 158–162 (2019).

23. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).

24. Drissen, R. *et al.* The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* **18**, 2485–2490 (2004).

25. Ethier, S. D., Miura, H. & Dostie, J. Discovering genome regulation with 3C and 3C-related technologies. *Biochim. Biophys. Acta* **1819**, 401–410 (2012).

26. Deng, W. & Blobel, G. A. Do chromatin loops provide epigenetic gene expression states? *Curr. Opin. Genet. Dev.* **20**, 548–554 (2010).

27. Dean, A. In the loop: long range chromatin interactions and gene regulation. *Brief. Funct. Genomics* **10**, 3–10 (2011).

28. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).

29. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

30. Bouwman, B. A. M. & de Laat, W. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol.* **16**, 154 (2015).

31. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* **60**, 676–684 (2015).

32. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).

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

34. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).

35. Rabl, C. Uber zelltheilung. *Morphol. Jahrb.* **10**, 214–330 (1885).

36. Comings, D. E. Arrangement of chromatin in the nucleus. *Hum. Genet.* **53**, 131–143 (1980).

37. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).

38. Cremer, T. *et al.* Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum. Genet.* **62**, 201–209 (1982).

39. Chandley, A. C., Speed, R. M. & Leitch, A. R. Different distributions of homologous chromosomes in adult human Sertoli cells and in lymphocytes signify nuclear differentiation. *J. Cell Sci.* **109 (Pt 4)**, 773–776 (1996).

40. Manuelidis, L. Individual interphase chromosome domains revealed by in situ hybridization. *Hum. Genet.* **71**, 288–293 (1985).

41. Pinkel, D., Straume, T. & Gray, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2934–2938 (1986).

42. Pinkel, D. *et al.* Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9138–9142 (1988).

43. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D. C. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80**, 224–234 (1988).

44. Cremer, T., Lichter, P., Borden, J., Ward, D. C. & Manuelidis, L. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Hum. Genet.* **80**, 235–246 (1988).

45. Boyle, S. *et al.* The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**, 211–219 (2001).

46. Nagele, R., Freeman, T., McMorrow, L. & Lee, H. Y. Precise spatial positioning of chromosomes during prometaphase: evidence for chromosomal order. *Science* **270**, 1831–1835 (1995).

47. Hoo, J. J. & Cramer, H. On the position of chromosomes in prepared mitosis figures of human fibroblasts. *Humangenetik* **13**, 166–170 (1971).

48. Sun, H. B., Shen, J. & Yokota, H. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* **79**, 184–190 (2000).

49. MILLER, O. J., MUKHERJEE, B. B., BREG, W. R. & GAMBLE, A. V. NON-RANDOM DISTRIBUTION OF CHROMOSOMES IN METAPHASE FIGURES FROM CULTURED HUMAN LEUCOCYTES. I. THE PERIPHERAL LOCATION OF THE Y CHROMOSOME. *Cytogenetics* **2**, 1–14 (1963).

50. Croft, J. A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).

51. Cremer, M. *et al.* Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosom. Res. an Int. J. Mol. Supramol. Evol. Asp. Chromosom. Biol.* **9**, 541–567 (2001).

52. Wiblin, A. E., Cui, W., Clark, A. J. & Bickmore, W. A. Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J.*

Cell Sci. **118**, 3861–3868 (2005).

53. Gilchrist, S., Gilbert, N., Perry, P. & Bickmore, W. A. Nuclear organization of centromeric domains is not perturbed by inhibition of histone deacetylases. *Chromosom. Res. an Int. J. Mol. Supramol. Evol. Asp. Chromosom. Biol.* **12**, 505–516 (2004).

54. Weierich, C. *et al.* Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosom. Res. an Int. J. Mol. Supramol. Evol. Asp. Chromosom. Biol.* **11**, 485–502 (2003).

55. Tjong, H. *et al.* Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1663-72 (2016).

56. Ferguson, M. & Ward, D. C. Cell cycle dependent chromosomal movement in pre-mitotic human T-lymphocyte nuclei. *Chromosoma* **101**, 557–565 (1992).

57. Bridger, J. M., Boyle, S., Kill, I. R. & Bickmore, W. A. Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr. Biol.* **10**, 149–152 (2000).

58. Walter, J., Schermelleh, L., Cremer, M., Tashiro, S. & Cremer, T. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol.* **160**, 685–697 (2003).

59. Canela, A. *et al.* Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-521.e18 (2017).

60. Pommier, Y., Sun, Y., Huang, S.-Y. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703–721 (2016).

61. Baranello, L. *et al.* RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357–371 (2016).

62. Baranello, L. *et al.* DNA break mapping reveals topoisomerase II activity genome-wide. *Int J Mol Sci* **15**, 13111–13122 (2014).

63. Madabhushi, R. *et al.* Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592–1605 (2015).

64. Yusufzai, T. M., Tagami, H., Nakatani, Y. & Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**, 291–298 (2004).

65. Uusküla-Reimand, L. *et al.* Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182 (2016).

66. Canela, A. *et al.* Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Mol. Cell* **75**, 252-266.e8 (2019).

67. Cowell, I. G. *et al.* Model for MLL translocations in therapy-related leukemia involving topoisomerase IIβ-mediated DNA strand breaks and gene proximity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8989–8994 (2012).

68. Zhang, Y. & Rowley, J. D. Chromatin structural elements and chromosomal

translocations in leukemia. *DNA Repair (Amst).* **5**, 1282–1297 (2006).

69. Herbert, S. *et al.* Chromatin stiffening underlies enhanced locus mobility after DNA damage in budding yeast. *EMBO J.* **36**, 2595–2608 (2017).

70. Ström, L., Lindroos, H. B., Shirahige, K. & Sjögren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* **16**, 1003–1015 (2004).

71. Bot, C. *et al.* Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage. *J. Cell Sci.* **130**, 1134–1146 (2017).

72. Oka, Y., Suzuki, K., Yamauchi, M., Mitsutake, N. & Yamashita, S. Recruitment of the cohesin loading factor NIPBL to DNA double-strand breaks depends on MDC1, RNF168 and HP1γ in human cells. *Biochem. Biophys. Res. Commun.* **411**, 762–767 (2011).

73. Caron, P. *et al.* Cohesin protects genes against γH2AX Induced by DNA double-strand breaks. *PLoS Genet.* **8**, e1002460 (2012).

74. Lang, F. *et al.* CTCF prevents genomic instability by promoting homologous recombination-directed DNA double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10912–10917 (2017).

75. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).

76. Janssen, A. *et al.* A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev.* **30**, 1645–1657 (2016).

77. Chiolo, I. *et al.* Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**, 732–744 (2011).

78. Kowalczykowski, S. C. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

79. Peng, J. C. & Karpen, G. H. Epigenetic regulation of heterochromatic DNA stability. *Curr. Opin. Genet. Dev.* **18**, 204–211 (2008).

80. Ryu, T. *et al.* Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat. Cell Biol.* **17**, 1401–1411 (2015).

81. Ziv, Y. *et al.* Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nat. Cell Biol.* **8**, 870–876 (2006).

82. Costes, S. V *et al.* Image-based modeling reveals dynamic redistribution of DNA damage into nuclear sub-domains. *PLoS Comput. Biol.* **3**, e155 (2007).

83. Leatherbarrow, E. L., Harper, J. V, Cucinotta, F. A. & O'Neill, P. Induction and quantification of gamma-H2AX foci following low and high LET-irradiation. *Int. J. Radiat. Biol.* **82**, 111–118 (2006).

84. Natale, F. *et al.* Identification of the elementary structural units of the DNA damage response. *Nat. Commun.* **8**, 15760 (2017).

85. Jakob, B. *et al.* DNA double-strand breaks in heterochromatin elicit fast repair protein

recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res.* **39**, 6489–6499 (2011).

86. Noon, A. T. *et al.* 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nat. Cell Biol.* **12**, 177–184 (2010).

87. Ayoub, N., Jeyasekharan, A. D., Bernal, J. A. & Venkitaraman, A. R. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. *Nature* **453**, 682–686 (2008).

88. Gall, J. G. & Pardue, M. L. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. U. S. A.* **63**, 378–383 (1969).

89. Barbieri, M. *et al.* Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat Struct Mol Biol* **24**, 515–524 (2017).

90. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

91. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e10 (2019).

92. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

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

94. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).

95. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).

96. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).

97. Markaki, Y. *et al.* The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *Bioessays* **34**, 412–426 (2012).

98. Gelali, E. *et al.* iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* **10**, 1636 (2019).

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

100. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901 (2007).

101. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).

102. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539-553.e8 (2020).

103. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).

104. Xu, H., Zhang, S., Yi, X., Plewczynski, D. & Li, M. J. Exploring 3D chromatin contacts in gene regulation: The evolution of approaches for the identification of functional enhancer-promoter interaction. *Comput. Struct. Biotechnol. J.* **18**, 558–570 (2020).

105. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).

106. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).

107. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).

108. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).

109. Marshall, O. J., Southall, T. D., Cheetham, S. W. & Brand, A. H. Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586–1598 (2016).

110. Vogel, M. J. *et al.* Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* **16**, 1493–1504 (2006).

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

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

113. Redolfi, J. *et al.* DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019).

114. Baranello, L., Kouzine, F., Sanford, S. & Levens, D. ChIP bias as a function of cross-linking time. *Chromosom. Res* **24**, 175–181 (2016).

115. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

116. Zhang, G. *et al.* N6-methyladenine DNA modification in Drosophila. *Cell* **161**, 893–906 (2015).

117. Greer, E. L. *et al.* DNA Methylation on N6-Adenine in C. elegans. *Cell* **161**, 868–878 (2015).

118. Finn, E. H. & Misteli, T. Genome Architecture from a Different Angle. *Dev. Cell* **41**, 3–4 (2017).

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

120. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5384 (2010).

121. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).

122. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).

123. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

124. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

125. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

126. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

127. O'Geen, H., Echipare, L. & Farnham, P. J. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol. Biol.* **791**, 265–286 (2011).

128. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

129. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

130. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).

131. Iacovoni, J. S. *et al.* High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29**, 1446–1457 (2010).

132. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

133. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).

134. DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* **23**, 1224–1234 (2013).

135. Bailey, T. *et al.* Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput Biol* **9**, e1003326 (2013).

136. Teytelman, L. *et al.* Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One* **4**, e6700 (2009).

137. Williamson, I. *et al.* Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **28**, 2778–2791 (2014).

138. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* **38**, 1184–1193 (2020).

139. Neiman, M. *et al.* Library preparation and multiplex capture for massive parallel sequencing applications made efficient and easy. *PLoS One* **7**, e48616 (2012).

140. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).

141. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).

142. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).

143. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

144. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–43, 1 p following 243 (2015).

145. Yang, F., Kemp, C. J. & Henikoff, S. Anthracyclines induce double-strand DNA breaks at active gene promoters. *Mutat. Res.* **773**, 9–15 (2015).

146. Lensing, S. V *et al.* DSBCapture: in situ capture and sequencing of DNA breaks. *Nat. Methods* **13**, 855–857 (2016).

147. Schwer, B. *et al.* Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2258–2263 (2016).

148. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).

149. Aryal, N. K., Wasylishen, A. R. & Lozano, G. CRISPR/Cas9 can mediate high-efficiency off-target mutations in mice in vivo. *Cell Death Dis* **9**, 1099 (2018).

150. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).

151. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).

152. Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).