

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

# **NOVEL FISH METHODS TO UNVEIL GENOME ARCHITECTURE**

Ana Mota



**Karolinska  
Institutet**

Stockholm 2021

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Universitetsservice US-AB, 2021

© Ana Mota, 2021

ISBN 978-91-8016-415-3

Cover illustration: The three blobs represent different chromatin states. The color gradient resembles the DNA density for each compartment.

# Novel FISH methods to unveil genome architecture

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Ana Mota**

The thesis will be defended in public at Scheelesalen, Karolinska Institutet, Scheeles väg 2, December 17<sup>th</sup> 2021

*Principal Supervisor:*

Assistant Prof. Magda Bienko  
Karolinska Institutet  
Department of Medical Biochemistry and  
Biophysics  
Division of Genome Biology

*Co-supervisor(s):*

Professor Camilla Björkegren  
Karolinska Institutet  
Department of Cell and Molecular Biology  
Department of Biosciences and Nutrition

Associate Prof. Sebastian Deindl  
Uppsala Universitet  
Department of Cell and Molecular Biology  
Division of Molecular Systems Biology

*Opponent:*

Professor Beat Fierz  
EPFL  
Department of Basic Sciences  
Division of Institute of Chemical Sciences and  
Engineering

*Examination Board:*

Associate Prof. Anna Klemm  
Uppsala University  
Department of Information Technology  
Division of Visual Information and Interaction

Professor Marcus Wilhelmsson  
Chalmers University of Technology  
Department of Chemical and Biological  
Engineering  
Division of Chemistry and Biochemistry

Dr. Antoine Coulon  
Institut Curie  
Department of Multiscale Physics-Biology-  
Chemistry and cancer  
Division of Physico-Chimie Curie Lab





To everyone that supported me and made this thesis exist.



## ABSTRACT

The genome consists of incredibly long DNA strands that encode all the vital information for the cell to function. The DNA inclusion in the very tight nuclear space and, simultaneously, the establishment of a hierarchical organization of the chromatin that favors transcription of certain genes over others, sparked a long-lasting quest to understand the design principles that govern genome architecture. The lack of proper technologies to study nanoscale structures made the progress slow until the advent of high precision microscopy techniques, next-generation sequencing technologies and in situ methodologies, which revolutionized the field. Not only did the new methodologies unveil which chromatin regions have preferred interactions with other regions, they also revealed that individual DNA sequences have preferential locations inside the nucleus, such as being either closer to the center of the nucleus or to the nuclear periphery. When these methods were applied to different cell and tissue types, genes were found to have different nuclear localization and compaction and to be surrounded with distinctive constellations of genes, depending on the tissue type. Those observations point to a strong interplay between genome architecture and chromatin activity that results in variable cellular functions. Gene activity is modulated by epigenetic modifications of chromatin, which vary depending on nuclear environment and affect gene accessibility to transcriptional machinery. Thenceforth, increased gene accessibility promotes a higher association with the transcription machinery that ultimately transcribe a set of accessible genes.

The following thesis papers follow the concept of first developing the needed technology to then be able to address biological questions that could otherwise not be tackled. Henceforth, the insights into the chromatin structure presented here are debated in the light of the new methodology.

**Paper I** aims to empower genome organization studies with a larger and more accessible repertoire of DNA FISH probes, which allow us to visualize small (10 kb) DNA regions using fluorescence microscopy. The paper describes the iFISH technique that reduces the probe production cost by pooling thousands of oligos that belong to many different probes in one single tube. The probes are selectively amplified thanks to a combination of barcodes that can include a color barcode that will hybridize with fluorescently labelled oligos. The probes are very specific and produce high signal-to-noise ratio signals in 6 colors. The ability to image 6 colors simultaneously, together with nucleus staining, permitted a thorough analysis of chromosome intermingling in both embryonic and differentiated cells.

**Paper II** describes a new method, GPSeq, which measures the radial position of DNA regions through genome-wide sequencing. The radial position is computed from samples that go through different, increasing, digestion times with restriction enzymes in fixed cells. The restriction enzymes gradually progress into the center of the nucleus homogeneously, forming concentric layers of digested chromatin. The longer the digestion time, the deeper the area cut by the restriction enzymes towards the nuclear center, which increases the number of different DNA sequences being detected through sequencing. To validate GPSeq in an independent

manner, extensive DNA FISH was required to target a vast number of loci scattered across various chromosomes, to compare radiality estimates coming from GPSeq with the gold-standard DNA FISH approach.

**Paper III** demonstrates how to enhance the imaging power of FISH by acquiring more loci simultaneously through multi-color probes. This expansion of iFISH into miFISH increases the number of probes that are individually detected, which empowers higher-throughout chromosome organization studies. The combining of different loci provided 120 pairwise distance measurements that were used to study megabase-scale models for chromosomal arrangement in space. This paper provides publicly available and open-access images that are coupled with detailed information of the multi-color probes datasets. Additionally, the iFISH image datasets used for the method validation are also available, together with miFISH datasets of one probe per channel, which helped to setup the method.

**Paper IV** introduces a new FISH technique, FRET-FISH, for measuring the condensation level of a target locus. FRET is employed to quantify the DNA proximity within a locus of interest using Förster resonance energy transference between dye pairs. In case the dye pairs are within a short distance, the dye of higher quantum excitation energy transfers the energy to the lower energetic dye that in turn emits the photons. Thus, the overall proximity between oligos labelled by either donor or acceptor dyes translates into a local compaction estimate, a metric that has been lacking thus far. For the technique optimization, FISH oligos were modified to include an additional sequence to stabilize the fluorescent dye interactions. Furthermore, the oligos were placed at a far enough distance to avoid instantaneous FRET with neighbor oligos. The resulting FRET efficiencies were compared with ATAC-seq and Hi-C results and showed a good correlation. The radial localization in the nucleus affects the locus structure in such a way that a locus closer to the lamina is more compacted than a locus closer to the center of the nucleus. Lastly, the same genes were used to study the influence of drug-induced alterations in compaction, cell cycle and genome instability.

# LIST OF SCIENTIFIC PAPERS

## THESIS PUBLICATIONS

- I. Gelali E. \*, Girelli G. \*, Matsumoto M., Wernersson E., Custodio J., Mota A., Schweitzer M., Ferenc K., Li X., Mirzazadeh R., Agostini F., Schell J., Lanner, F., Crosetto N., Bienko M. (2019) iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* 10, 1-15.
- II. 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.
- III. Mota A., Schweitzer M., Bienko M. (2021) Simultaneous visualization of DNA loci in single cells by combinatorial multi-colour iFISH. in revision
- IV. Mota A., Wernersson E., Li-Wang X., Gradin K., Crosetto N., Bienko M. (2021) FRET-FISH probes chromatin density at individual genomic loci in single cells. in preparation

## SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

- I. Li D., Kular L., Vij M., Herter E. K., Li X., Wang A., Chu T., Toma M., Zhang L., Liapi E., Mota A., Blomqvist L., Sérézal I. G., Rollman O., Wikstrom J. D., Bienko M., Berglund D., Ståhle M., Sommar P., Jagodic M., Landén N. X. (2019) Human skin long noncoding RNA WAKMAR1 regulates wound healing by enhancing keratinocyte migration. *PNAS* 116, 116.

\* These authors contributed equally to the work.



# CONTENTS

1	GENOME ARCHITECTURE IN EUKARYOTES.....	1
1.1	Genome organization advances.....	1
1.2	Microscopy approaches to study genome architecture .....	6
2	CHROMATIN COMPACTION IN EUKARYOTES .....	9
2.1	Condensates as essential structures of interphase chromatin.....	9
2.2	Methods to study condensation of chromatin .....	11
2.2.1	Sequencing.....	12
2.2.2	Microscopy .....	14
2.2.3	Live-cell imaging and chromatin dynamics.....	17
3	RESEARCH AIMS .....	19
4	MATERIALS AND METHODS.....	21
4.1	Probe design.....	21
4.1.1	Applied in iFISH and GPSeq .....	21
4.1.2	Applied in miFISH .....	21
4.2	Probe selective amplification (TDA).....	22
4.3	Cell culture .....	22
4.4	Cell fixation and permeabilization .....	23
4.4.1	Applied in iFISH and GPSeq .....	23
4.4.2	Applied in miFISH and FRET-FISH .....	24
4.5	Probe <i>in situ</i> hybridization in genomic DNA (DNA FISH) .....	24
4.5.1	Applied in iFISH and GPSeq .....	24
4.5.2	Applied in miFISH and FRET-FISH .....	25
4.6	DNA and RNA FISH.....	25
4.7	Immunofluorescence staining with H3K27me3 and H3K27ac .....	25
4.8	Nascent RNA detection kit .....	26
4.9	Image acquisition .....	26
4.10	Image processing and FISH dots identification .....	27
4.10.1	Applied in iFISH, GPSeq and miFISH .....	27
4.10.2	Applied in FRET-FISH .....	27
4.11	Signal decoding for miFISH.....	28
4.12	Analysis of chromosome topologies for miFISH.....	29
4.13	FRET efficiency calculation.....	29
4.14	Workflow for GPSeq .....	30
4.15	YFISH and imaging .....	30
5	RESULTS.....	31
5.1	Paper I: iFISH is a free resource enabling versatile DNA FISH to study genome architecture in single cells.....	31
5.1.1	iFISH probe design.....	32
5.1.2	Validation of the iFISH probes .....	33
5.1.3	Chromosome territories and intermingling .....	34

5.2	Paper II: GPSeq reveals the radial organization of chromatin in the cell nucleus .....	38
5.2.1	GPSeq implementation.....	39
5.2.2	GPSeq cross-validation with independent techniques.....	40
5.2.3	Validation of radial genome organization models.....	41
5.2.4	3D genome modeling .....	45
5.3	Paper III: Simultaneous visualization of DNA loci in single cells by combinatorial multi-color iFISH .....	47
5.3.1	miFISH probe design .....	48
5.3.2	Validation of miFISH pairwise distances with Hi-C.....	50
5.3.3	Lamina distances .....	51
5.3.4	Pairwise distances and higher-order chromosomal structures.....	52
5.3.5	Global chromosome shapes.....	53
5.4	Paper IV: FRET-FISH probes chromatin density at individual genomic loci in single cells .....	55
5.4.1	FRET-FISH probe design .....	56
5.4.2	Optimization of FRET-FISH probe design to study local DNA density.....	58
5.4.3	Validation of FRET-FISH to chromatin changes upon drug treatment .....	59
5.4.4	Prolonged cell culturing leads to chromatin condensation.....	61
5.4.5	Chromatin density changes during the cell cycle .....	63
5.4.6	Validation of the FRET efficiencies with ATAC-seq and Hi-C published data.....	63
5.4.7	Radial localization in the nucleus influences chromatin condensation .....	64
6	DISCUSSION AND CONCLUSIONS .....	67
7	ACKNOWLEDGEMENTS.....	73
8	REFERENCES.....	75



# LIST OF ABBREVIATIONS

AFM	Atomic Force Microscopy
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
BLISS	Breaks Labeling In Situ and Sequencing
ChIP-seq	Chromatin immunoprecipitation sequencing
ChromEMT	Chromatin Electron Microscopy Tomography
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC-binding Factor
DamID	DNA adenine methyltransferase IDentification
DNA	DeoxyriboNucleic Acid
DSB	Double-Strand Break
EMT	Electron Microscopy Tomography
ESC	Embryonic Stem Cell
ESI	Electron Spectroscopy Imaging
FISH	Fluorescence In Situ Hybridization
FFPE	Formalin-Fixed Paraffin-Embedded
FLIM	Fluorescent Lifetime Imaging Microscopy
FRAP	Fluorescence Recovery After Photobleaching
FRET	Förster Resonance Energy Transfer
GC	Guanine and Cytosine
GPSeq	Genomic loci Positioning by Sequencing
HB	Hybridization Buffer
Hi-C	High-throughput chromosome conformation capture
HP1	Heterochromatin Protein 1
IF	ImmunoFluorescence
IVT	In Vitro Transcription

LAD	Lamina-Associated Domain
MEF	Mouse Embryonic Fibroblast
MERFISH	Multiplexed Error-Robust Fluorescence In Situ Hybridization
NAD	Nucleolus-Associated Domain
NGS	Next-Generation Sequencing
NIH3T3	NIH Swiss murine fibroblast cell
PALM	Photo-Activated Localization Microscopy
PCC	Pearson's Correlation Coefficient
PCR	Polymerase Chain Reaction
PFA	Para-formaldehyde
PHB	Pre-Hybridization Buffer
PGS	Population-based Genome Structure
PRC	Polycomb repressive complex
ORCA	Optical Reconstruction of Chromatin Architecture
RNA	RiboNucleic Acid
RPE	Retinal Pigment Epithelial
RT	RetroTranscription
SCC	Spearman Correlation Coefficient
SMC	Structural Maintenance of Chromosomes
SMLM	Single-Molecule Localization Microscopy
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SNR	Signal to Noise Ratio
STORM	STochastic Optical Reconstruction Microscopy
TAD	Topologically Associating Domain
UMI	Unique Molecular Modifier
XCI	X Chromosome Inactivation

# 1 GENOME ARCHITECTURE IN EUKARYOTES

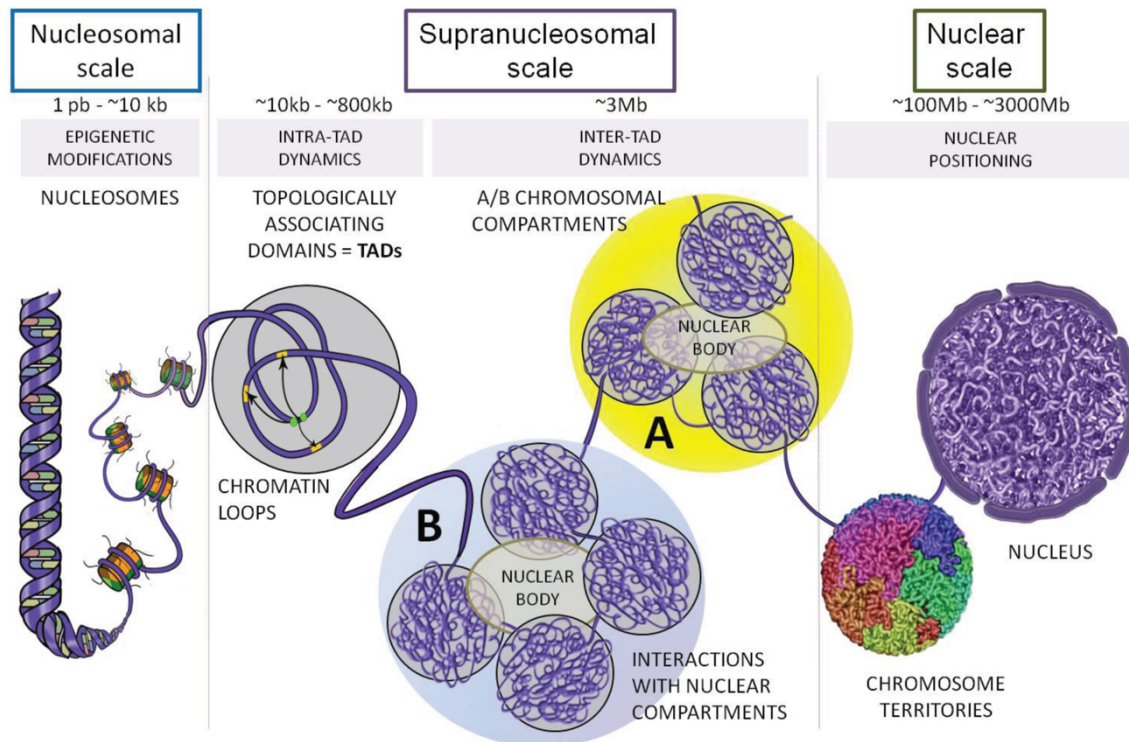
## 1.1 Genome organization advances

The cell nucleus was the first cell organelle to be discovered in the 17<sup>th</sup> century due to its opaqueness when observed through microscopy. This opaqueness was thereafter associated with a dense concentration of nucleic acids, which compose the genomic DNA. This organelle is responsible for protection and secure transmission of genetic information that encodes all cell processes. The densely packed DNA in the nucleus is organized in a heterogenous manner, forming highly packed regions called heterochromatin and less packed regions called euchromatin. This high compaction of heterochromatin is linked to its relatively low level of activity and relatively low gene density, and conversely, the openness of euchromatin reflects its high level of expression and higher density of genes. The higher-order spatial organization of chromatin seems not to be a random, given the preferential location of heterochromatin at the periphery of the nucleus in close contact to the nuclear lamina. This unavoidably places euchromatin at a more central location in the nucleus.<sup>1,2</sup> This is in line with the typical positioning of individual chromosomes, which carry variable ratio of hetero- / euchromatin. Typically, gene-rich chromosomes that are rich in euchromatin are positioned in the nuclear center (in human these are chromosomes 1, 16, 17, 19, and 22).<sup>3</sup> These observations strongly point to a link between gene location and its activity and indeed, it is currently believed that nuclear position and local environment of a gene have very important implications for gene expression regulation. Together, this makes the field of genome organization very relevant for our understanding of the basis of transcriptional regulation, emergence of variable cellular phenotypes and cell adaptability.<sup>4</sup>

One important example of the link between cellular function and genome organization, and at the same time an exception to the typical arrangement of chromatin in the nucleus, is the chromatin arrangement in the rod cells of the retina of nocturnal animals.<sup>5,6</sup> In those nuclei gene-rich chromosomes, which are typically more internally positioned in conventional nuclei, are located close to the nuclear lamina. Such an arrangement of chromatin makes these nuclei act as efficient lenses able to collect as much of light as possible for those animals to be able to see well in darkness. The discovery of the inverted type of chromatin prompted researchers to question whether this cell type is an exception to the rule or whether more unconventional arrangements of chromatin exist, and ever since, a few more have been discovered.

To get a thorough understanding of genome organization, highly advanced techniques for large-scale data acquisition and processing are needed. One of the technological

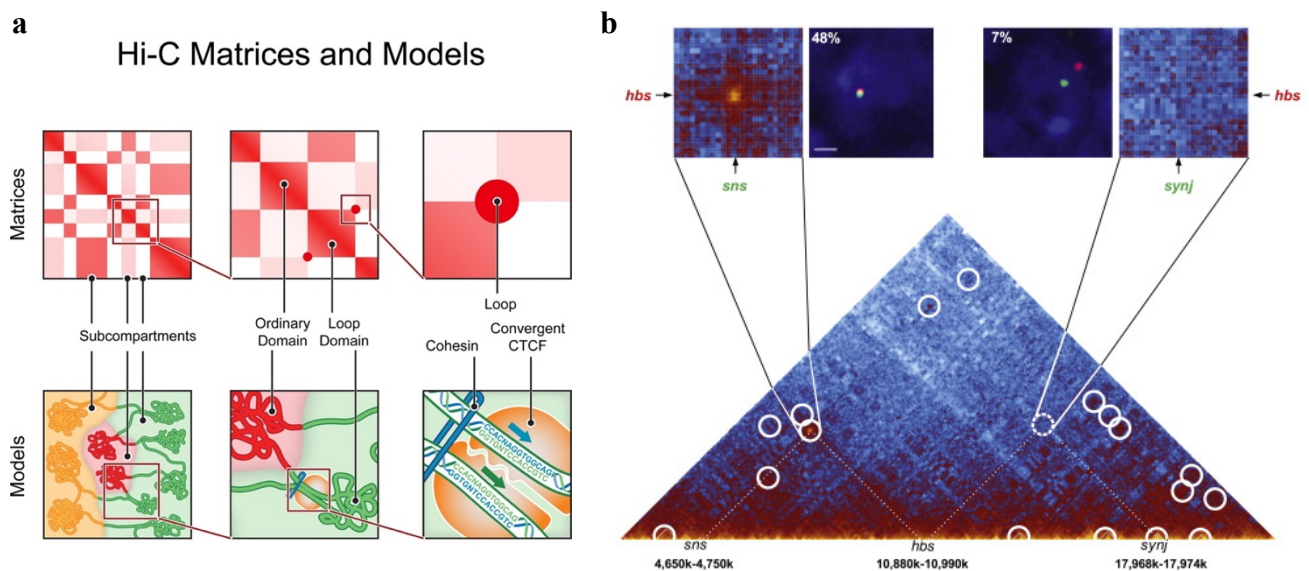
breakthroughs was the development of the high-throughput chromosome conformation capture (Hi-C) technique, which allowed identification of structural units of the genome (**Figure 1**).<sup>7</sup> Hi-C is a sequencing-based technique that quantifies inter-locus interactions by measuring proximities between DNA sequences that are close in 3D space. The resulting Hi-C map is a matrix of contact probabilities of all possible combinations of DNA loci, as shown in **Figure 2a**. In Hi-C, two DNA loci can be found in proximity even if located many nucleotides apart in the linear genome, possibly revealing functional interactions.



**Figure 1** – Schematic representation of the current model of mammalian genome organization in 3D, from the nucleosomal to the nuclear scale. From the bottom up, the DNA helix is coiled around histones to form nucleosomes, which on the supranucleosomal scale can form TADs. Distinct TADs interact with each other, generating two distinct compartments that differ in their histone marks and transcriptional activity (A and B). At the nuclear scale, the chromosomes form discrete territories. Ea *et al.*, *Genes*, 2015.

By mapping proximity between loci genome-wide, Hi-C revealed structural units called topologically associating domains (TADs)<sup>8</sup>, which are structural domains ranging in size from thousands to millions of DNA bases that show relatively high contact probability for DNA loci residing within its boundaries. TAD boundaries are marked by the CTCF motif – a binding site for the transcription factor CTCF (**Figure 2**). The formation of a TAD has been hypothesized to be a product of the loop extrusion process, which leads to the creation of a DNA loop whose

stems are formed by CTCF sites. Regions within such a loop are thought to contact each other more frequently than other neighboring regions. In the loop extrusion model, the process of extruding a loop is thought to start at a random site and then be blocked upon encountering the CTCF protein bound to its recognition site. The protein complex responsible for the extrusion is thought to be the cohesin complex, given that cohesin has the ability to move along DNA<sup>9</sup> also in association with other proteins<sup>10,11</sup> and colocalizes with CTCF<sup>12–14</sup>. However, loop extrusion remains a highly controversial model as it does not explain how cohesin would overcome obstacles such as the transcription machinery or even nucleosomes while extruding a loop. Hence, the nature of the formation of TADs and in general genome condensation is still a subject of intense debate.



**Figure 2** – Examples of Hi-C matrices and corresponding models. (a) The top left matrix has several TADs represented by dark red squares with symmetry at the diagonal. The matrix shows the contact probability between two genomic regions in x and y with the darker red color representing more frequent contacts, while the white color represents more random contacts. The first model shows the presence of compartments indicated by the checkerboard pattern. The second model zooms into smaller structural domains, either in a looped configuration or not. The third model shows a loop domain stabilized by cohesion, which serves as a ring encircling DNA, with CTCF as anchoring point. Rao *et al.*, *Cell*, 2014. (b) The triangle situated at the top of the diagonal shows the level of interaction between different genomic regions. Two sub-triangles are shown that represent domains with and without an anchor point, as highlighted by the circles zooming in the region. Sexton *et al.*, *Cell*, 2012.

Besides TADs, higher-scale chromatin arrangements have been predicted from Hi-C maps. The mammalian genome seems to be divided into spatially segregated A and B compartments, which represent interactive associations of chromatin regions within each

compartment and a depletion of such contacts between loci belonging to the two compartments. Shorter or longer genomic windows belonging to one or the other compartment alternate along every chromosome. This is evident when looking at **Figure 2a**, where 3 arrows are pointing at 3 different regions away from the diagonal, indicating contacts between non-consecutive compartments. These regions show distinct color gradients, revealing variable interaction strength. Importantly, distant regions can be marked in dark red indicating that genomic windows located far away along the linear genome can be found in proximity very often. When plotting all long-distant interactions, two very distinct values of strength or interaction probability appear, in line with two different types of compartments.<sup>7</sup> The compartment highlighted in the model with yellow and with dark red in the matrix has a more intense interaction and is arbitrarily known as compartment B. The other compartment has lighter red color in the matrix and red in the model and is called A. The white region in the matrix resembles random contacts that are established at the borders of the compartments. The two compartments correlate with decorations with different histone marks, with compartment B associating with inactive and repressive histone marks like H3K9me3, and compartment A associating with histone marks indicative of active transcription such as H3K4me3. The formation of the two compartments might be driven by homotypic interactions between the DNA-binding proteins, in line with the recently revisited phase separation theory.<sup>15</sup> Given that the B compartment is marked by inactive histone marks it represents the well-known heterochromatin. One proposed way of heterochromatin formation is through histone tail-driven liquid-liquid phase separation, in particular by the protruding tail of the H1 linker histone.<sup>16,17</sup> Additionally, the heterochromatin structure could also arise through the action of HP1 $\alpha$ , a protein that binds to chromatin and creates nucleation sites for phase separation that expand to neighbouring chromatin, forming a dense mesh of DNA and proteins blocking other proteins from entering and activating genes.<sup>18–20</sup> Another observation of liquid-like condensates, this time within the euchromatic or A compartment, concerns super-enhancers, where the BRD4 and Mediator proteins drive phase separation and thereby maintain expression of key genes via compartmentalization of the transcription apparatus.<sup>21</sup>

Finally, Hi-C-based studies confirmed previous findings from DNA FISH that individual chromosomes occupy exclusive volumes called chromosomal territories (**Figure 1**).<sup>22</sup> In Hi-C, the captured cross-linked associations are predominantly in *cis* rather than in *trans*, which means that DNA loci belonging to the same chromosome interact much more with each other than with any other chromosome. There is a remarkable exception to this, in the case of undifferentiated cells where the genome is kept more active, adaptable and dynamic within the limited space of the nuclei. The differentiation process can quickly change these

properties and create isolated chromosome territories that are more condensed with better-defined intra-nuclear positions.<sup>23,24</sup>

Inspired by the discovery of chromosome territories, Cremer *et al.* proposed that these territories are separated by an interchromatin space where transcriptional machinery, among other proteins, is free to circulate.<sup>25</sup> This hypothesis is currently debatable, since many studies demonstrated some level of intermingling between chromosomes.<sup>23,26,27</sup> One example of intermingling comes from human lymphocytes, particularly when the cells were exposed to ionizing radiation.<sup>27</sup> Another example was reported by Maharana *et al.*, in which the extent of intermingling was compared between undifferentiated embryonic stem cells and differentiated cells.<sup>23</sup> Chromosome intermingling in differentiated cells was found to be less transient, more stable and to involve more significant volume of certain chromosomes. Additionally, the same group also noticed that the intermingling sites displayed higher transcription activity than the remaining regions of the chromosome. Hence, the well-separated chromosome territories might be the dominant scenario, with the intermingling an exceptional event needed for gene activation or recombinational repair, or the interchromatin space may represent a rather transient formation. These findings of the relation between genome structure and transcription activity led to the long-standing question of whether the transcriptional status affects genome organization or, whether genome architecture influences gene activation, for example by favoring interactions needed for transcription activation, such as enhancer-promoter interactions. Likely, both scenarios exist in parallel at different sites in the nucleus.

Intriguingly, a recent study has shown that chromosome territories were established as hierarchical structure only in mammalian cells, while chromosome territories were found absent in non-mammalian species.<sup>28</sup> In the same study, other features of the genome were found to be conserved along different species, like TADs that are positionally correlated with cohesin in most species.<sup>8,29</sup> Moreover, the lack of chromosome territories in certain species was associated with the absence of condensin II subunits and the study demonstrated that depleting condensin II in species that have chromosome territories leads to longer and less compacted chromosomal arms in mitosis, and stronger interactions between the centromeres. The less compacted arms intermingle more extensively with other chromosomes, while TADs are kept unchanged, evidencing that TADs and chromosome territories likely have independent formation pathways. Interestingly, the proteins underlying TAD and chromosome territory formation are, respectively, cohesin and condensin, which have very similar molecular structures and both belong to the family of structural maintenance of chromosomes (SMC) complexes. However, their activity seems to be regulated by different pathways since they preferentially bind to chromatin at different cell cycle stages.<sup>30</sup> Even though both proteins were

hypothesized and demonstrated to form loops upon binding to DNA *in vitro*, the directionality of the loop extrusion differs as cohesin extrudes DNA in both directions<sup>31</sup> while condensin only does so in one direction.<sup>32</sup> In this way, the TAD structure can always be reproduced after every cell division because CTCF sites halt the cohesin-mediated loop extrusion proceeding in both directions, while condensin-mediated condensation during mitosis is associated with loss of TADs since the loop extrusion is not halted by any specific sequence.<sup>33,34</sup>

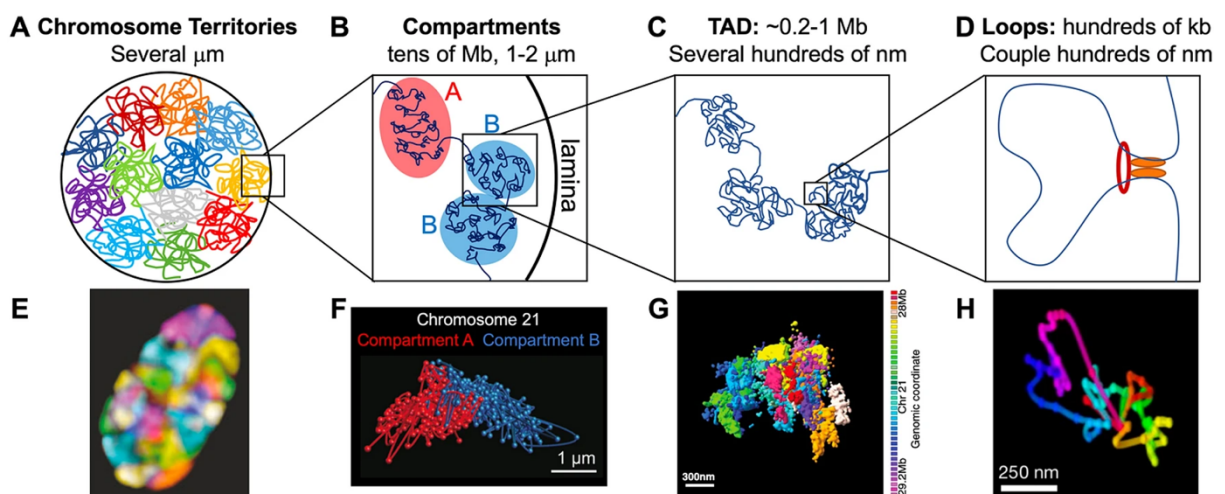
## 1.2 Microscopy approaches to study genome architecture

To confirm observations from various sequencing approaches, orthogonal ways needed to be developed. To this aim, the DNA fluorescence *in situ* hybridization (FISH) technology has witnessed an incredible revival. New multiplexed DNA FISH techniques have been implemented and used to re-create similar contact matrices as those generated by Hi-C after imaging thousands of cells and averaging contact frequencies from all. Importantly, microscopy offers the ability to visualize a genomic structure in an authentic environment and allows addressing: 1) which loci are in close proximity in how many cells; 2) how the genome is arranged in the region where specific contacts happen; 3) where the proximity occurs inside the nucleus. Of note, also transcription can be assessed simultaneously by performing RNA FISH or immunochemistry target assays. Until recently, microscopy also had the advantage of inherently being a single-cell technique. However, protocols for single-cell Hi-C (scHi-C) have now been established to produce single-cell contact maps and infer single-cell genome structures<sup>33,35</sup>. Nevertheless, the technical noise for single-cell sequencing techniques is high, which lowers the resolution of the assay from the typical 10 kb (for a bulk Hi-C) to 100 kb (besides the lack of higher-order spatial organization information). The FISH method has been considerably improved over the past 10 years by bringing together different approaches, including increasing the number of dyes used simultaneously, using super-resolution microscopy (as exemplified by oligo stochastic optical reconstruction microscopy (oligoSTORM)<sup>36</sup>, and the use of sequential rounds of hybridization and stripping combined with microfluidics<sup>37</sup>, as pioneered by the Zhuang, Boettiger and Wu labs.<sup>38–40</sup> Benefiting from these technical advances, the Cai and Yuan groups targeted 3,660 loci, 70 RNA species and 17 antibodies using sequential FISH and immunofluorescence in 446 cells.<sup>41</sup>

The above-mentioned progress in the FISH methodology makes it now possible to envision FISH as a standalone assay powerful enough to reveal novel aspects of genome architecture, and not as a mere validation tool as typically has been done in the past 10 years. Until recently, due to the limited number of target loci that could be visualized in one single FISH experiment, one typically relied on predictions from Hi-C when designing FISH probes.



Indeed, the existence of the compartments revealed using Hi-C data was confirmed by DNA FISH. However, the structure of a TAD, predicted by Hi-C to be a single globule, looks rather different under a microscope when regions encompassing Hi-C-predicted TADs are visualized by FISH. Rather than one large globule, oftentimes single cells have two or more smaller globular structures in that location.<sup>42</sup> This indicates that while high-throughput population approaches are extremely valuable, the correspondence of structures identified in bulk sequencing assays to those present in individual single cells is limited. Luckily, the astonishing improvements of super-resolution microscopy and FISH allow us now to obtain detailed images of relatively large genomic regions at high resolution from thousands of cells. In one such study, a 700 kb region was imaged at a resolution of 2 kb using 70 hybridization/stripping rounds<sup>39</sup>, while in another study a 30 kb resolution was achieved for a 2.5 Mb region using 65 hybridization/stripping rounds (**Figure 3**)<sup>40</sup>. Of note however, such experiments take several weeks to be performed and require a very complex setup. Consequently, to achieve the same resolution for the whole genome would require many months of non-stop imaging and an immense storage capacity that is indispensable for each imaging session. Therefore, this type of imaging has high costs associated and prohibitive time that typical labs cannot afford.



**Figure 3** – Hierarchical genome organization and respective examples of microscopic images. A - Schematics of chromosome territories several micrometers in length. B – Schematics of compartments tens of megabases in size. C - Topologically associating domain (TAD) at the sub-megabase scale. D - Loops bringing distant genomic regions in proximity. E - Each chromosome is hybridized by DNA FISH oligos and imaged with wide-field fluorescence microscopy showing chromosome territories. F - Super-resolution microscopy of A and B compartments in a cell using FISH oligos with A and B color code based on Hi-C maps. G - A TAD-like globular structure imaged by sequential OligoSTORM. Each color represents a single hybridization step that starts at the genomic coordinate of 28 Mb until 29.2 Mb in chr21. H - A 330-kb region imaged by ORCA and each color represents a hybridization step. Shim, *Genes & Genomics*, 2021.

To alleviate costs associated with high-throughput FISH, our lab developed pipelines for a more affordable production of FISH probes for smaller genomic windows that can be used, apart from in basic research, by translational research and diagnostics labs and that are compatible with protocols that preserve the 3D structure of cells.<sup>43</sup> The oligonucleotide databases and ready-to-use probes are publicly available and we keep on expanding our probes repository in order to cover as much of the human and mouse genome as possible by ready-to-use probes. In parallel, new databases have been designed with unique oligonucleotides targeting genomes of model organisms that are less frequently used. iFISH probes have been used to study various aspects of the spatial arrangement of chromosomes in human cells, i.e. chromosomal intermingling, and found for example a larger extent of mingling in human embryonic stem cells in comparison to human fibroblasts. We look forward to multiplex DNA FISH experiments through color combinations and as such expand the number of individual DNA loci that can be simultaneously imaged.<sup>44</sup>

## 2 CHROMATIN COMPACTION IN EUKARYOTES

### 2.1 Condensates as essential structures of interphase chromatin

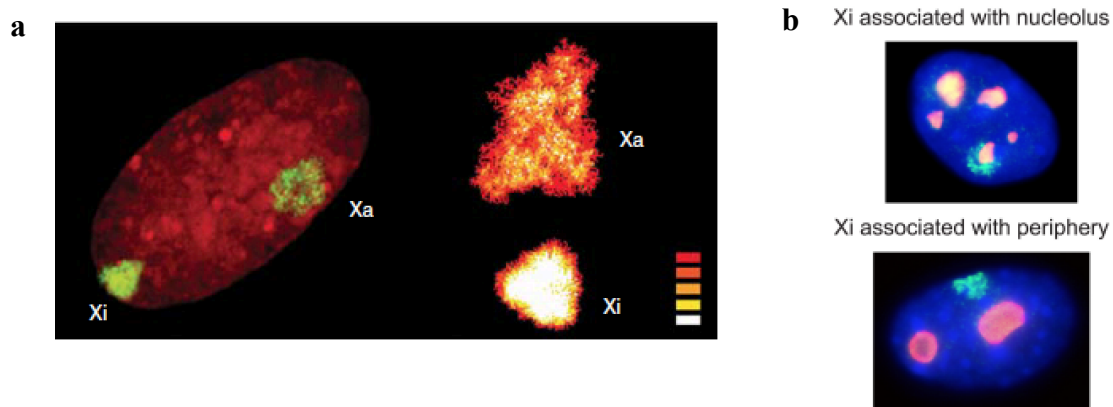
The genome architecture field has been driven by the curiosity of understanding how a long DNA strand encoding the whole complex structure and functioning of a cell could fit in a small nucleus. To add to the problem, an individual organism has billions of cells with the same (or highly similar) genotype but that express diverse phenotypes depending on how a given cell type ‘reads’ its own genetic information. In fact, the transcriptional activity of a pluripotent embryonic stem cell is very broad involving most of the genome at some level.<sup>24</sup> Changes to transcriptional activity during differentiation reduce the active portion of the genome in a cell lineage-specific manner. Additionally, during differentiation the histone modification repertoire becomes more complex and modifies transcriptional output and most likely genome structure, for instance by silencing pluripotency genes, and many genomic elements are repressed and become part of the heterochromatin. Of note, heterochromatin, apart from being a silencing compartment, has many structural roles like the formation of telomeres, which prevent chromosomes from deterioration, or centromeres, which ensure correct separation of sister chromatids.<sup>45,46</sup>

The key to the condensation of a negatively charged DNA polymer and the regulation of its accessibility are the positively charged histones. Histones reduce the electrostatic repulsion from the phosphate backbone, which allows DNA to be easily wrapped around them and in this way form so-called nucleosomes, which represent the first step in the compaction of DNA. This first layer of DNA packaging already obstructs access to DNA for various proteins including RNA polymerase II and can at the same time redirect various proteins to appropriate sites that are left unoccupied. In this way, the interaction between various proteins (most notably enzymes) and the DNA is not random and results in differential accessibility, not only between different genes but also along a gene body. For example, transcription start sites of active genes are more accessible than the rest of the gene body in order to properly initiate transcription.<sup>47</sup> Nonetheless, nucleosomes are highly dynamic and responsive to external signals to regulate DNA occupation thanks to various post-translational modifications of the amino (N)-terminal tails of histones by various remodeling enzymes. Modifications that alter overall histone charge, like acetylation, have the potential to disrupt the histone cores forming nucleosomes. When the nucleosome structure is disturbed, their contact with DNA is affected in consequence, for example in acetylation, the DNA will be more accessible for other proteins to bind.<sup>48</sup> On the other way, histone modifications can promote direct binding of

histones tail with other proteins. One example is heterochromatin protein 1 (HP1) that binds to methylated histones through its chromodomain and is responsible for the spread of condensation by interacting with other HP1 bound to DNA through liquid-liquid phase separation.<sup>49</sup>

One of the most prominent condensation events is X chromosome inactivation (XCI) in female cells, in which one of the two allele becomes highly condense and therefore mostly inactive (**Figure 4a**). XCI is believed to be necessary for maintaining the same gene dosage between males and females.<sup>50</sup> Nevertheless, the silencing pathways seem to have diverged during evolution and differ between different species. The XCI process is best understood in mouse, where the inactivation occurs during early embryonic developmental stages. XCI starts from the X inactivation center (Xic) that is hyper-transcribed from the future inactive X chromosome, producing *Xist*, a long non-coding RNA.<sup>51</sup> *Xist* coats chromosome X in order to attract the Polycomb silencing complexes PRC1 and PRC2. The loss of acetylation on the H3K27 residue caused by PRC1 is followed by the ubiquitination of H2K119 by the same complex, and eventually by the methylation of H3K27 mediated by PRC2, which in turn leads to condensation and inactivation.<sup>51–54</sup> *Xist* is also responsible for stabilizing the methylated histones close to the Xic locus and anchoring the repressed chromosome X to the cell lamina or nucleolus forming the Barr body (**Figure 4b**). Furthermore, *Xist* expression initiation and the corresponding X chromosome coating leads to further chromatin modifications that include: depletion of RNA polymerase II; enrichment in histone repressive marks such as H3K9me and H4K20me<sub>3</sub>; macroH2A deposition; DNA hypermethylation at CpG islands; and loss of TADs in the reshaped Xi structure.<sup>55,56</sup>

XCI is achieved through the use of a large molecular machinery that sits on the genome and is responsible for establishment of a complex epigenetic landscape. Several research groups hypothesize that the Xi does not engage in inter-chromosomal interactions due to the local saturation of molecules on its surface, which lead to liquid-liquid phase separation.<sup>57</sup> Contrary to the expectation, proteins that are constituents of the heterochromatin of the Xi do not form solid-like structures that were found in chromocenters or in bacterial nucleoli.<sup>58</sup> The liquid-liquid phase separation of the Xi is accomplished through proteins that directly interact with *Xist*, such as Spen, Ptbp1, HnrnpK, PRC1 and PRC2, which multimerize.<sup>57</sup> *Xist* contains nucleotide repeats that are present in scaffold RNAs and promote protein sequestration. More recently, Polycomb proteins were found to assemble through phase separation due to the intrinsically disordered region of the CBX2 protein.<sup>59,60</sup>



**Figure 4** – The imaged nucleus represents differentiated female mouse cells in which one allele of chromosome X is silenced, Xi, and Xa is the active allele. Both images show a preferential position of Xi at the periphery of the nucleus or nucleolus. (a) On the left: DNA FISH using paint probes against chrX in human fibroblasts; green: FISH signal, red: DNA stain. On the right: Xi and Xa territories are compared in terms of shape; volume and signal intensity profiles varied due to different DNA densities. Respectively, the white, yellow and red colors are high, medium and low intensities. The Xi allele is expected to be more condensed than Xa and thus shows higher intensity due to higher density of DNA. Cremer & Cremer, CSHLP, 2010. (b) Xi is inferred from RNA FISH against *Xist* (green), in red: immunostaining against the nucleophosmin protein that is found abundantly in the nucleolus, and in blue: DNA stain. Fang *et al.*, *Front. Cell Dev. Biol.*, 2019.

The study of the differences between the active and the inactive alleles on the two chromosomes X copies has been possible thanks to the use of mouse cell lines derived from cross-bred mouse species such as the *Mus spretus* and the *Mus musculus*. The spectrum of single nucleotide polymorphisms (SNPs) between these two species is large and allows for the distinction between the maternal and the paternal X alleles<sup>61</sup>. Thanks to RNA-seq performed on such cell lines, 3-7 % of Xi genes were found to escape silencing and were named escapees.<sup>62</sup> The set of genes escaping silencing seems constant among cells of the same tissue type, while variations of expression can be observed across different tissues.<sup>63</sup> Of note, silencing of escapees is responsible for severe health conditions, such as the autism-like Turner syndrome.<sup>64</sup>

## 2.2 Methods to study condensation of chromatin

The last decade witnessed a considerable progress in the field thanks to the development of new methods that always precede new knowledge. For example, major discoveries that helped establish chromatin structure models were possible thanks to new sequencing-based methods,

most notably Hi-C. Despite the great progress, current methods, whether sequencing- or microscopy-based, do not directly probe for a highly relevant parameter in the study of chromatin architecture, being chromatin compaction. As a consequence, condensation models are built from data obtained by methods that measure properties such as inter-locus interactions, DNA accessibility, or epigenetic marks. These properties are then used as a proxy for DNA compaction. For example, highly interacting chromatin regions are considered to represent highly condensed regions, and yet they could instead represent highly dynamic regions.<sup>65</sup> Similarly, lowly accessible regions are considered to be tightly packed DNA sequences, and yet they could represent rather extended regions that are frequently coated by proteins. The following paragraphs introduce the state-of-the-art methods that helped shape the current models of chromatin compaction and discuss their limitations.

### 2.2.1 Sequencing

**Hi-C** (high-throughput chromosome conformation capture) captures proximities between genomic loci positioned physically close in nuclear space by cross-linking the chromatin during the fixation step<sup>7</sup>. Contacts captured by Hi-C are believed to be positioned as far apart as 1  $\mu\text{m}$ <sup>66</sup>. The cross-linked DNA is fragmented with a restriction enzyme and the sticky ends are filled in in a way that introduces biotin at the digested site. The ends are then allowed to ligate to other neighboring ends oftentimes forming hybrid DNA fragments that do not follow the linear genomic sequence but represent spatial contacts between DNA fragments from different genomic sites. Streptavidin purification then allows for selecting DNA that has undergone digestion and ligation. After high-throughput sequencing, the final output allows the construction of a genome-wide map of chromatin contacts, both intra- and inter-chromosomal. Hi-C maps obtained in bulk are responsible for much of the advances that have been recently witnessed in the field of genome organization, as explained in the previous chapter. To transpose those findings into single cells proved to be challenging, as chromatin behavior seems to be very heterogenous across single cells, which can make TADs or compartments to look smaller and more disorganized structures.<sup>67–69</sup>

**ChIP-seq** (chromatin immunoprecipitation sequencing) maps genomic regions to which a protein of interest is bound.<sup>70</sup> ChIP-seq is particularly useful for identifying the typical epigenetic environment of a certain DNA sequence. Various chromatin states (e.g., promoter, transcribed, repressed) were defined thanks to ChIP-seq data by analyzing the enrichment of specific histone marks at distinct genomic regions.<sup>71</sup> The ChIP-seq protocol starts with crosslinking of DNA to proteins bound to it, after which chromatin is purified and fragmented, and incubated with an antibody targeting the protein of interest. The antibody-protein-DNA

complex is then immunoprecipitated, and the DNA is amplified by PCR and finally sequenced using high-throughput platforms. The epigenetic landscape is widely used as a predictor of the genome architecture, for example, a region with high deposition of a certain histone mark is expected to be associated either with gene activation or silencing. This however does not mean that the chromatin 3D structure can be inferred from such data. Moreover, each experiment is typically performed with only one antibody which limits more complex studies like studying co-localization with other proteins to speculate about a certain underlying DNA structure.

**ATAC-seq** (assay of transposase accessible chromatin sequencing) brings us closest to the study of DNA compaction, though still not in a direct manner.<sup>72</sup> ATAC-seq probes for chromatin accessibility by measuring which DNA fragments can be modified by the Tn5 transposase. The least accessible genomic regions are shielded by proteins that sit on the DNA (such as histones or transcription factors), and are detected as relatively long fragments in the sequencing library. Instead, the most accessible chromatin regions are exposed to the enzymatic activity of the Tn5 and thus seen as smaller sequences. To validate the relation between histone occupancy and DNA accessibility, ATAC-seq data were shown to anti-correlate with ChIP-seq against histones<sup>72</sup>. The transposase not only cuts the open chromatin but also inserts adapters which are then used for PCR amplification and subsequent high-throughput sequencing library preparation. Of note, it is not clear how different biophysical properties of different chromatin types might affect the action of the transposase and therefore it remains unclear how biased/unbiased this approach is. For example, phase separation, which produces membrane-less structures, can theoretically affect the efficiency of the transposase by creating local environments (liquid droplets) that can differentially influence its action.

These biochemical methods have undeniably taken the genome organization field to the scientific spotlight, where chromatin 3D structure proved to have a strong link with gene expression profiles<sup>73–75</sup>. However, all sequencing methods carry various, often common limitations. Typically, sequencing data come from thousands to millions of pooled cells and therefore yield average profiles. It is now possible to switch to single-cell profiling, however those approaches are technically and computationally (due to high noise levels) very challenging.<sup>71</sup> Moreover, translation of the models generated from bulk experiments to single cell observations has been tricky given that single cell genome structures are highly heterogeneous.<sup>67,68</sup> Lastly, most of the works using these approaches rely on isolated cells from blood and cell culture, while the most relevant samples are cells isolated directly from tissues and the methods described here have a very limited success in this type of material.<sup>76</sup>

### 2.2.2 Microscopy

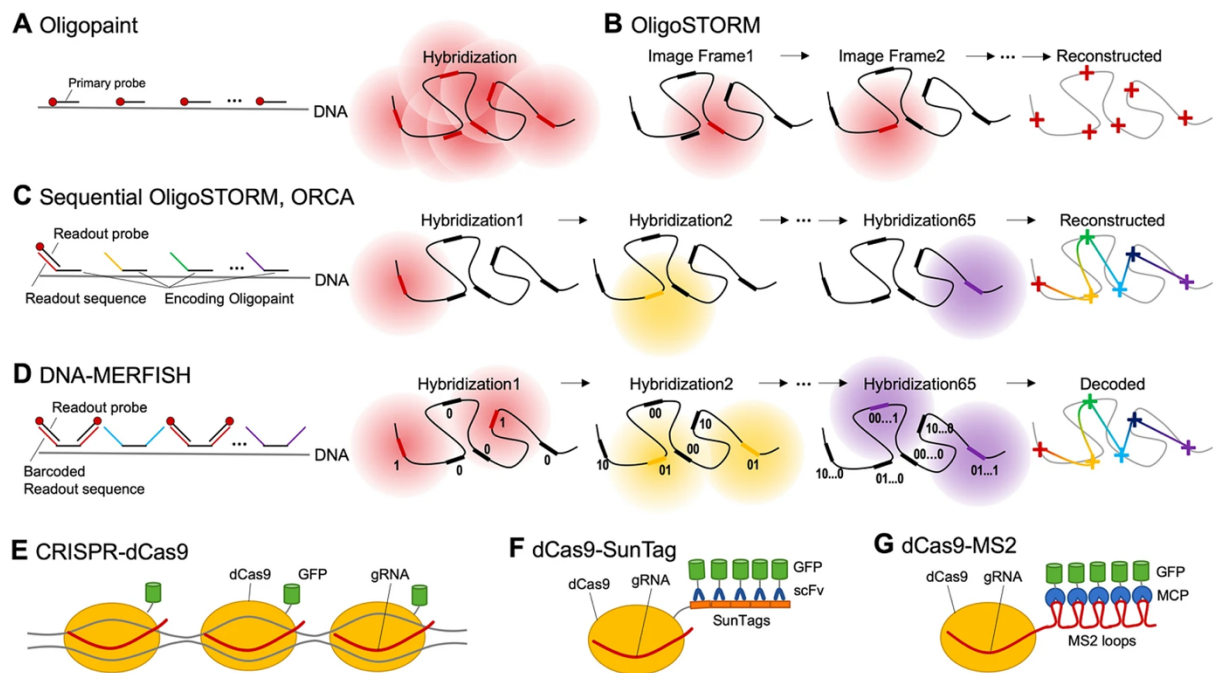
Formidable advances have been achieved using microscopy to study genome architecture, which greatly complement sequencing approaches by allowing direct visualization of the genome in the nuclear space. In order to acquire spatial information on the genome, **DNA FISH** has been the most common microscopy-based technique. In DNA FISH, fluorescently labelled oligos are hybridized onto a specific genomic region of interest in fixed cells, after the sample has been heated in a formamide solution for dsDNA denaturation (**Figure 5A**). Of note, FISH techniques can be applied even to FFPE samples and therefore, this technique is widely used in clinical diagnostics to study, for example, chromosomal rearrangements in cancer patients.<sup>77</sup>

A disadvantage of the FISH method is the high probe cost since each oligo is manufactured for a specific target, and a single target locus requires at least one hundred (often up to thousands) oligos to provide bright signals over nuclear background. To overcome such expenses, FISH oligos can be produced as an oligopool when including a unique combination of primers at the oligo ends to allow for high-throughput selective amplification of each probe.<sup>37</sup> Moreover, the adoption of microfluidics in the FISH field allowed for multiplexed visualization of many different loci by introducing a repetitive step of hybridization that incorporates fluorescently labelled oligos sequentially into the oligos already bound to the region of interest.<sup>78</sup> Ultimately, FISH is a very versatile technique where one can adapt the structure of the oligos to the desired application. For example, the overhangs added to the target-specific oligos can provide a template for rolling circle amplification, thus increasing the number of fluorophores per target oligo and thereby dramatically increasing the signal brightness.<sup>79,80</sup>

The implementation of super-resolution imaging in microscopy expanded the spatial resolution that was previously limited to 250 nm in the x,y down to 10-30 nm, breaking the fundamental limits of diffraction.<sup>81,82</sup> Single-molecule localization microscopy (**SMLM**) describes a group of powerful super-resolution techniques that improved the spatial resolution and that have been implemented in genome organization studies.<sup>83</sup> The most applicable method seems to be stochastic optical reconstruction microscopy (**STORM**), in which photoswitchable molecules are stochastically switched on to stimulate photon emission before they enter a dark state or photobleach.<sup>84</sup> This method suffers from long acquisition times since it requires several steps of photoactivation but it can accurately and precisely detect localizations of fluorescently-labelled molecules of interest. STORM revealed the structure of nucleosomes and described them as heterogeneous objects of hundreds of nanometers, as opposed to the previously postulated 30 nm fiber structure.<sup>85</sup> Importantly, a combination of Oligopaint DNA FISH probes



and STORM (named as **OligoSTORM**) (**Figure 5B**) allowed for imaging of small regions of 30 kb where TADs and sub-TADs were observed as globular structures in population-average level in line with Hi-C.<sup>85</sup> The resolution of genome imaging was then pushed further down to 2 kb using an approach called optical reconstruction of chromatin architecture (**ORCA**) by hybridizing fluorescent oligos sequentially along the linear genome tiling in total a 2 Mb region (**Figure 5C**).<sup>39</sup> **DNA-MERFISH** enables the highest throughput to date by encoding each region with barcodes (fluorescently labelled oligos with an unique sequence) that are decoded after all the imaging steps are completed (**Figure 5D**).<sup>42</sup>



**Figure 5** - Labeling methods for visualizing specific genomic loci in intact nuclei. A – Oligopaint is a DNA FISH technique where synthetic oligonucleotides hybridize to a denatured DNA strand at a specific sequence. The signal is detected from several oligos that densely cover a region of interest. Conventional DNA FISH is imaged with wide-field microscopy, which is diffraction-limited, making the detection of fine chromatin structures hard to achieve. B - Oligopaint combined with STORM (OligoSTORM) detects individual fluorescent molecules by stochastic switching of single-molecule fluorescence signal which reconstructs structures with finer detail. C - Sequential hybridization and imaging of the OligoSTORM method leads to optical reconstruction of chromatin architecture (ORCA). D – The probes for each region carry a unique barcode region enabling higher reading throughput (DNA-MERFISH). E - CRISPR-dCas9 labeling system for live-cell labeling of a specific genomic locus. The enzymatically deactivated version of Cas9 (dCas9) is fused to a fluorescent protein (FP) and recruited to a specific locus by a guide RNA (gRNA) sequence. Several FPs are required for signal detection which makes repetitive regions an easy target. F – The fluorescent signal is improved by fusing

a SunTag to dCas9. G - dCas9 with engineered gRNA that has an array of the MS2 loops which recruit FP-fused MCP protein. Adapted from Shim, *Genes & Genomics*, 2021.

The development of super-resolution microscopy came relatively late in comparison to electron microscopy (EM), and EM still achieves the highest resolutions. However, application of electron microscopy to biological samples remains challenging because the image quality depends on the contrast of the structures, while the sample preparation procedure aimed at increasing this contrast introduces artefacts. For example, early electron spectroscopy imaging (**ESI**) increased the contrast of chromatin over the remaining parts of the nuclei by covering it with heavy atoms such as uranium and lead salts.<sup>86</sup> However, this severely disrupts the chromatin structure, and therefore a new approach consisting of adding phosphorous ions to the imaging solution was developed.<sup>87</sup> Additionally, cryo-EM uses phase contrast between the atoms of the molecules and the vitreous ice.<sup>88</sup> Both techniques provide a very weak contrast of DNA and the resulting structures do not support hierarchical chromatin folding. The most recent advancement is the **ChromEMT** method where cells are fixed with glutaraldehyde and DNA is stained with the DRAQ5 dye, which facilitates DAB polymerization on DNA yielding a dark precipitate appearance.<sup>89</sup> According to the ChromEMT data, chromatin is a disordered chain with a diameter ranging from 5 to 24 nm and with different levels of chromatin densities in interphase nuclei.

In order to obtain more information about a locus environment, DNA FISH can be coupled with immunofluorescence (**IF**) that targets specific epigenetic marks on histones.<sup>90</sup> IF uses antibodies to target proteins and, similar to ChIP-seq, informs on the relative density of certain proteins around the region of interest. A combination of IF and RNA FISH was used to observe phase separation at super-enhancers by targeting intronic sequences of *Esrrb* loci with RNA FISH and simultaneously using IF to visualize the Mediator protein that regulates gene expression.<sup>21</sup> The epigenetic state of a locus can also be assessed by measuring chromatin accessibility, similar to ATAC-seq. **ATAC-see** and **ATAC-PALM** have been used to detect open chromatin in single cells using microscopy, involving a transposase that introduces fluorescently-labelled adapters into accessible DNA<sup>91,92</sup>. Lastly, condensation of chromatin in the nucleus can also be addressed by mechanical measurements such as atomic force microscopy (**AFM**), micropipette and squeezing.<sup>93,94</sup> For example, AFM detected a more elastic chromatin behavior in cells that undergo trichostatin A (TSA) treatment (which leads to chromatin decompaction through the interference with histone acetylation) and subsequently a more loose interaction of DNA with proteins.<sup>95</sup> Nonetheless, only DNA FISH-based methods can specifically target DNA sequences in fixed cells.

### 2.2.3 Live-cell imaging and chromatin dynamics

DNA FISH and sequencing methods produce a snapshot of the spatial arrangement of chromatin, since its dynamic behavior is unattainable from fixed cells. Chromatin dynamic studies should capture in a more realistic manner how condensates establish, how a given DNA locus interacts with other chromatin elements, and how dynamic such events are.<sup>96</sup>

To track down specific sequences in time, the bacterial **CRISPR/Cas9** system was engineered to bring EGFP-tagged endonuclease-deficient Cas9 protein to a region of interest in the genome. However, like in DNA FISH, the signal detection is possible only if multiple fluorescent molecules locate on adjacent sequences. Hence, multiple sgRNAs, targeting multiple regions in a tiled manner, are needed to detect a locus of interest using this approach, which limits its application to repetitive regions like telomeres (**Figure 5E**).<sup>97</sup> In a more recent CRISPR/Cas9 approach, the fluorescent brightness detected from one single targeted site was increased by fusing Cas9 to a SunTag to which 24 GFP molecules can bind (**Figure 5F**).<sup>98</sup> Further signal enhancement in such approaches was achieved by extending the sgRNA with 14 copies of the MS2 site that is recognized by the MCP protein fused to GFP (**Figure 5G**).<sup>99</sup> Despite such a signal amplification, the labeling efficiency at both alleles was as low as 5%. On a side note, the CRISPR/Cas9 system has also helped to develop a new DNA FISH approach, though not a live-cell one, by facilitating opening of the target region through consecutive use of CRISPR/Cas9 followed by a helicase to avoid heat denaturation of DNA.<sup>100</sup>

Another approach to study chromatin dynamics is fluorescence recovery after photobleaching (**FRAP**), which is widely used in live-cell experiments. First, the molecule of interest is labelled with a fluorescence molecule. Next, a region of the nucleus in which this protein resides is photo-bleached and if there is molecule mobility, the neighbor molecules diffuse to the bleached area and over time the fluorescence of the bleached area is recovered. For example, FRAP was applied to mouse zygotes expressing eGFP-H2B, where the bleached chromatin region experienced only a small signal recovery at the interface with the non-bleached region.<sup>101</sup> The signal recovery was associated to the chromatin looseness, which indicates the degree of freedom that a chromatin region has to move and the eventuality it has of interacting with different chromatin regions. In a different study, FRAP was performed on fibroblast cells in which fluorescently-labelled nucleotides were incorporated at early or at late replication stage in order to distinguish euchromatin and heterochromatin.<sup>58</sup> The results suggested that euchromatin and heterochromatin do not mix, even when decondensation is induced through chromatin hyperacetylation, implying a rather solid-like behavior in contrast to the liquid-liquid phase separation model. The solid-like state was also observed in

nucleosomal arrays *in vitro*, suggesting that the behavior is innate to the chromatin fiber and is independent on non-histone protein presence.

In another example, single particle tracking was used together with super-resolution techniques such as photoactivated localization microscopy (**PALM**) to investigate chromatin dynamics at nanoscale resolution.<sup>19</sup> The DNA-binding proteins were fused to fluorescently-labelled molecules that have photoactivable properties such as Dendra, PAmCherry or mEos2.<sup>102</sup> In line with the previous chromatin dynamics model, nucleosomes were found to form compacted nanodomains with low mobility in heterochromatin. Besides, this mobility was reduced during cell differentiation.<sup>103</sup>

A simple method to measure chromosome condensation *in vivo* consists of staining DNA with DAPI or Hoechst 33342 dyes followed by the measurement of the fluorescent lifetime using fluorescence lifetime imaging microscopy (FLIM)<sup>104</sup>. FLIM detects the time that a fluorophore takes to decay its photon contrary to the methods presented previously that detect the number of photons emitted. The photon lifetime depends on many environmental conditions, such as: (1) salt concentrations that affect the dielectric constant of the medium, and (2) proximity to other dyes and proteins that might quench and reduce the decay time.<sup>105</sup> **DAPI FLIM** informs on the formation of condensates by detecting regions with low photon lifetime caused by close interactions between DAPI molecules. Following the same principle of molecule proximity, Förster resonance energy transfer (FRET) takes two or more fluorescent molecules with overlapping spectrums of emission (in the case of the dye with the most energetic excitation state) and of excitation (in the case of the lowest energetic excitation state) that transfer the energy through non-radiative dipole–dipole coupling.<sup>106</sup> FRET efficiency is inversely proportional to intermolecular distance and this property allowed extensive applications as a molecular distance sensor.<sup>107–109</sup> **FLIM-FRET** aimed at studying chromatin condensation by monitoring FLIM from co-expressed histone H2B tagged to either Enhanced Green FP or the mCherry FP. This approach detected an increase of chromatin compaction in mitosis most likely due to nucleosomes being packed more closely, bringing the FRET pair molecules close-by translating this proximity proportionally to the lifetime decay.<sup>110,111</sup> However, neither DAPI FLIM nor FLIM-FRET target specific sequences of the DNA, limiting their application when one would want to target compaction of selected TADs or individual genes.

### 3 RESEARCH AIMS

In the four papers described below, we aimed at developing new methods to study genome organization. The methods aim to deliver new perspectives of how the genome is packed inside the nucleus by enabling new parameters to be investigated, like radially or condensation. Besides, our methods provide open-resource image datasets, oligo database for DNA FISH and probe design pipelines. In this way, chromatin studies are more accessible to a wider variety of research groups and diagnostics labs that are not FISH specialists.

The specific goals for each paper are described below.

#### **Paper I:**

- To develop high-throughput production of probes, reduce probe costs and allow color-multiplexing with 6 different colors to increase the number of loci imaged simultaneously. Open-source oligo databases are included on the FISH4U.org website that assists with probe design. (iFISH)

#### **Paper II:**

- To develop a new sequencing method that measures the radial localization of the genome in the nucleus. (GPSeq)

#### **Paper III:**

- To develop a method that increases the number of loci measured in a single experiment by combining colors for each locus. The method aims to understand how chromosome is organized at the higher level. The image datasets are available for public usage. (miFISH)

#### **Paper IV:**

- To develop a microscopic technique that measures chromatin compaction at specific targets. (FRET-FISH)



## 4 MATERIALS AND METHODS

### 4.1 Probe design

#### 4.1.1 Applied in iFISH and GPSeq

For the iFISH (paper I) and GPSeq (paper II), only human cell lines were used and therefore the human genome was the reference (Grch37/hg19 GCA\_000001405.1) and all 40 nt sequences appearing only once were extracted. To avoid binding in other genomic regions, a filter was created to discard all 40-mers found in other sites with 7 nt or more bases consecutively, or with a GC-content outside of the 35–80% interval. Next, the average melting temperature was assessed. Afterwards, the sequences with homology of 70% or higher to more than one genomic location were removed. The energy necessary to disrupt secondary structures at 65 °C was calculated and discarded the 40-mers with delta energy values below zero. Additionally, only the 40-mers with a melting temperature in a range of 20 °C around the previously calculated average temperature remained. Lastly, overlapping 40-mers by starting from the first one and iterating through were discarded. After all the filters, the 40-mers are kept in a sqlite3 database for easy access and are freely used for probe design through the iFISH4U interface.

Each probe consists of 96 oligos, each 100 nucleotides (nt) long. The oligo sequence contains two 20 nt adapter sequences (F and R) for PCR amplification, a 40 nt sequence complementary to the DNA target (T), and a 20 nt color barcode (C) to which a detection fluorescently labelled oligo can be hybridized

#### 4.1.2 Applied in miFISH

The iFISH4U webtool ([www.ifish4u.org](http://www.ifish4u.org)) is where the 40-mers targeting the human genome were used to design the 16 miFISH probes targeting 16 loci on chr2. In the same way as before, each oligo of the probe is 100 nt long, however, due to the mixing of two colors per probe, the number of oligos increased to 700 in each probe. All the oligos were purchased as a single 12k oligopool (Twist Biosciences). The F and R adapter sequences of the oligos were combined in distinct combinations to make the probes distinguishable and amplified selectively. Moreover, 10 of the 16 probes were labelled with two alternating colors (out of 6 colors that can be resolved using our microscope), while the remaining 6 probes were labelled with a single color.

## **4.2 Probe selective amplification (TDA)**

To produce the FISH probes, the oligopool with all probes is selectively amplified by real-time PCR using the SYBR Select Master Mix (Thermo Fisher Scientific) and a probe-specific combination of primers complementary to the F and R adapter sequences were dispensed in the 96-well plates. The PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter) and the DNA concentration was measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Afterwards, the double stranded PCR product in each well (probe) is converted into RNA by the HiScribe T7 Quick High Yield RNA synthesis kit (New England Biolabs), a process called the in vitro transcription (IVT) which was carried out at 37 °C for 16 h. Each tube contained a final volume of 30 µL consisting of 1 µg of purified PCR product, 6.67 mM of dNTPs (Thermo Fisher Scientific), 2 units of RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific) and 2 µL of T7 RNA polymerase mix. Next, the resulting RNA was captured with Agencourt RNAClean XP beads (Beckman Coulter) for purification and the product has the concentration measured with the Qubit RNA BR assay (Thermo Fisher Scientific). Later, RNA is converted into cDNA by reverse transcription (RT) carried out at 50 °C for 1 h using the Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) and a primer carrying the C adapter sequence serving as docking site for the fluorescently labeled detection oligo. Each tube contained a final volume of 20 µL containing 15 µg of purified RNA, 1.5 mM of dNTPs, 20 µM of the corresponding primer, 1x reverse transcription buffer, 10 units of Maxima H Reverse Transcriptase and 2 units of RNaseOUT (Thermo Fisher Scientific). The enzymes were inactivated by heating the reaction to 85 °C for 5 min. Subsequently, the template RNA was removed with 20 µL of 0.5 M EDTA (Thermo Fisher Scientific) and 20 µL of 1 M NaOH (Sigma Aldrich) and incubated at 95 °C for 15 min. Finally, the resulting single-stranded DNA (ssDNA) is purified and concentrated by using Oligo Binding Buffer (Zymo Research) and Zymo-Spin IC columns (Zymo Research) which added 40 µL Nuclease-Free Water (Thermo Fisher Scientific) for each probe in the column. The DNA concentration was measured with the Qubit ssDNA Assay Kit (Thermo Fisher Scientific) and the probes were stored at -20 °C.

## **4.3 Cell culture**

A549 lung carcinoma cells, HME human mammary epithelial cells, IMR90 fetal lung fibroblasts, MEF embryo fibroblast and NIH3T3 embryo fibroblast were purchased from ATCC (cat. no. CCL-186, PCS-600-010, CCL-185, SCRC-1040 and CRL-1658, respectively) and HAP1 chronic myeloid leukemia cells from Horizon Discovery (cat. no. C859). Human



embryonic stem cells (HS975 40) were derived and used following the donor's written consent and approval from Regional Ethical Review Board in Stockholm (2011/745–31/3).

Specifically: A549 cells were cultured in Ham's F-12K (Kaighn's) Medium (Thermo Fisher Scientific) supplemented with 10% FBS; HME cells in Medium 171 (Thermo Fisher Scientific) supplemented with Mammary Epithelial Growth Supplement (Thermo Fisher Scientific); IMR90 cells in Minimum Essential Medium (MEM, Merck) supplemented with 10% FBS, 1% non-essential amino acids (Thermo Fisher Scientific) and 1% L-glutamine (Thermo Fisher Scientific); MEF cells in Dulbecco's Modified Eagle's Medium (DMEM, Sigma) supplemented with 15% FBS (Sigma); NIH3T3 cells in DMEM supplemented with 10% FBS; HAP1 cells in Iscove's Modified Dulbecco's Medium (IMDM, Sigma) supplemented with 10% FBS and primed hESCs were cultured in NutriStem hPSC XF Medium containing bFGF and TGF57 (Biological industries) on coverslips pre-coated with 10 µg/ml Human recombinant laminin-521 (BioLamina). The cells were incubated at 37 °C in 5% O<sub>2</sub> and 5% CO<sub>2</sub>. When the MEF cells reached 80% confluency were dispersed on coverslips (thickness 1.5 µm, 22x22 mm, VWR) and HAP1 cells on 9-well chambered coverslips (custom-made by Grace Bio-Labs), and the remaining cells were immobilized in both type of supports.

#### **4.4 Cell fixation and permeabilization**

##### **4.4.1 Applied in iFISH and GPSeq**

Unless otherwise specified, the incubations are performed at room temperature. Briefly, the cells are fixed in 1x PBS (Thermo Fisher Scientific)/4% formaldehyde (EMS) for 10 min, followed by quenching of the unreacted formaldehyde in 1x PBS/125 mM glycine (Sigma Aldrich) for 5 min. Subsequently, the coverslips are washed three times, 5 min each with 1x PBS (Thermo Fisher Scientific)/0.05% Triton X-100 (Promega) and permeabilized the cells in 1x PBS/0.5% Triton X-100 for 20 min, then washed twice for 5 min with 1x PBS/0.05% Triton X-100. The coverslips were incubated overnight with 1x PBS/20% glycerol and then, subjected to five cycles of freeze and-thaw in liquid nitrogen (30 s in liquid nitrogen, thawing in ambient air, 2–3 min in 1x PBS/20% glycerol at room temperature). Following, another two washes with 1x PBS/0.05% Triton X-100 for 5 min. Afterwards, the coverslips were incubated in 0.1 N HCl for 5 min and quickly rinsed twice in 1x PBS/0.05% Triton X-100. Lastly, the coverslips were exchanged to 2x SSC buffer (Thermo Fisher Scientific) and incubated overnight in 2x SSC/50% formamide (Thermo Fisher Scientific)/50 mM sodium phosphate (Sigma Aldrich). The following day, the coverslips were exchanged to 2x SSC and stored at +4 °C.

#### 4.4.2 Applied in miFISH and FRET-FISH

The glycerol and liquid nitrogen step were demonstrated that were not responsible for keeping the 3D structure and, additionally, increased the nucleus background. Similar results were obtained by removing those steps from the protocol. Therefore, for miFISH and FRET-FISH the new protocol without those steps were applied.

### 4.5 **Probe *in situ* hybridization in genomic DNA (DNA FISH)**

#### 4.5.1 Applied in iFISH and GPSeq

The coverslip is immersed in a pre-hybridization buffer (PHB) containing 2x SSC/5x Denhardt's solution (Thermo Fisher Scientific)/50 mM sodium phosphate buffer/1 mM EDTA/100 ng/μL ssDNA (Invitrogen)/50% formamide, pH 7.5–8.0, and incubated them for 1 h at 37 °C in a humidity chamber. The DNA FISH probe at the final concentration of 0.5 nM is mixed with the first hybridization mix (HM-1) at 1:9 vol/vol ratio with 1.1x first hybridization buffer (HB-1) containing 2.2x SSC/5.5x Denhardt's solution/55 mM sodium phosphate buffer/1.1 mM EDTA/111 ng/μL ssDNA/55% formamide/11% dextran sulfate (Sigma Aldrich) pH 7.5-8. The coverslip is removed from PHB and then, immersed on top of 30 μL of HM-1 deposited onto a microscope slide. The coverslip is sealed with fixogum (MP Biomedical) and placed on a heating block for 3 min at 75 °C in order to denature the DNA in the nucleus, followed by a 15-18 h incubation at 37 °C. The next day, we washed the coverslip three times, 10 min each at 37°C in 2× SSC/0.2% Tween, while shaking, followed by two washes, 7 min each at 58°C in 0.2× SSC/0.2% Tween pre-warmed at 58°C, inside a water bath, a brief wash in 4× SSC/0.2% Tween at RT, two brief washes in 2× SSC and one final short wash in 2× SSC/25% formamide. Next, the samples were washed, and the second hybridization was performed by incubating the samples in a humidity chamber with 100 μL of the second hybridization mix (HM-2: containing fluorescently-labeled oligonucleotides, each at a final 20 nM concentration, in 2xSSC/25% formamide/10% Dextran sulfate/1 mg/mL E.coli tRNA/0.02% bovine serum albumin/10 mM Vanadyl-ribonucleoside complex) at 30 °C for 3 h. Afterward, the samples were washed and stained for DNA (30 min at 30 °C in 1.23 ng/mL Hoechst 33342 in 2x SSC/25% formamide). Lastly, the coverslip is rinsed twice in 2x SSC, before mounting them in GLOX containing 2x SSC/10mM Tris-HCl pH 7.5 (Sigma)/0.4% Glucose (Sigma)/10mM TROLOX (Sigma)/37 ng/μL Glucose Oxidase (Sigma)/32mM Catalase (Sigma).

#### 4.5.2 Applied in miFISH and FRET-FISH

Same protocol as in iFISH but with the following changes: (1) the DNA denaturation takes 2min and 30sec to reduce the nucleus damage; (2) after first hybridization, 2× SSC/0.2% Tween wash is removed; (3) instead of 7 min each at 58 °C in 0.2× SSC/0.2% Tween wash, the time is reduced to 5 min and the temperature increased to 65 °C, the background noise is reduced; (4) the second hybridization was prolonged for 24 h.

#### 4.6 **DNA and RNA FISH**

The protocol follows a workflow similar with DNA FISH but, in order to preserve RNA, RNases inhibitors were included in most of the buffers, after the fixation in 1x PBS/ 4% paraformaldehyde when the nucleus starts to permeabilize and RNases might diffuse. For the sample preparation, all the buffers except for HCl were heated at 60 °C for 10 min after adding Ribonucleoside Vanadyl Complex (RVC, New England Biolabs) at a final concentration of 10 mM. For the FISH experiment, PHB step is excluded as a source of RNases so the coverslips are incubated directly with HM-1 (probe at 1:9 v/v ratio, 1.1x HB-1, 1U/μL Protector (Roche)) and heated at 75 °C for 2 min 30 sec on a heating block for DNA denaturation. The coverslip is kept in the humidity chamber for 15–18 h at 37 °C. Afterwards, the coverslip is washed two times, 5 min each at 65 °C in 0.2x SSC/0.2% Tween/10mM RVC pre-warmed at 65 °C, inside a water bath, a quick wash in 4x SSC/0.2% Tween/10mM RVC at RT, another wash in 2x SSC/10mM RVC and a final wash in 2x SSC/25% formamide/10mM RVC. At this point, the second hybridization mixture is prepared with 5 ng/μL transcript probe in HB-2 with 1U/μL Protector and incubated with the coverslip for 24 h at 30 °C. Afterwards, the coverslip is washed for 1 h at 30 °C in 2x SSC/25% formamide, followed by the third hybridization mix containing the secondary fluorescently labelled oligonucleotides at a final 20 nM concentration in HB-2 for 3 h at 30 °C. Then, the coverslip is washed for 1 h at 30 °C in 2x SSC/25% formamide and the nucleus is stained with 1.23 ng/mL Hoechst 33342 in 2x SSC/ 25% formamide at 30 °C for 30 min. At last, the coverslip is rinsed twice in 2x SSC before mounting with GLOX for imaging.

#### 4.7 **Immunofluorescence staining with H3K27me3 and H3K27ac**

The coverslips were washed two times with 1xPBS briefly at room temperature. Then, the coverslip was placed on a drop of 300 μl of the Blocking solution (5% BSA/1xPBS/0.1% Tween20) for overnight at 4°C in a humid environment. Thereafter, the coverslip was lift up with 1 ml of the Blocking solution, and place directly onto a drop of 300 μl of primary antibody

solution (abcam, H3K27me3 and H3K27ac antibodies come from Rabbit) and was incubated for 1h in a humid environment at room temperature. Afterwards, the coverslip was lift up with 1 ml of the Wash solution (0.1% Tween20/1xPBS). Then, the coverslip was transferred to a petri dish for three washes in the Wash solution for 10 min each at room temperature on a shaker. Subsequently, the coverslip was placed on a drop of 300  $\mu$ l of secondary antibody solution (abcam, H3K27me3 and H3K27ac antibodies come from Rabbit) for 1h at room temperature in a humid environment. Later, the washes with the Wash solution were repeated. Finally, the coverslip was stained with 1xPBS/1 ng/ $\mu$ l Hoechst for 5 min at room temperature and then, mounted with GLOX.

#### **4.8 Nascent RNA detection kit**

The protocol is provided by Invitrogen for Click-iT<sup>®</sup> RNA Imaging Kits. The cells are plated at 40–50% confluency in 6-well plates, and fed the following day with 1 mM of 5-ethynyl uridine for 1 hour. Afterwards, each well is immersed in 1 mL of 3.7% formaldehyde/1x PBS and incubated for 15 minutes at room temperature. Then, each well is washed once with 1x PBS. The coverslips are permeabilized with 0.5% Triton X-100/1x PBS for 15 min at room temperature and then washed with 1x PBS. The wash solution is removed and added 500  $\mu$ L of Click-iT<sup>®</sup> reaction cocktail (a mixture of Click-iT<sup>®</sup> RNA reaction buffer, CuSO<sub>4</sub>, Alexa Fluor<sup>®</sup> azide, Click-iT<sup>®</sup> reaction buffer additive) prepared in the same day and incubated for 30 minutes at room temperature, protected from light. The cells are washed with 1 mL of Click-iT<sup>®</sup> reaction rinse buffer. The nucleus is stained with 1.23 ng/mL Hoechst 33342 in 1x PBS at 30 °C for 15 min. At last, the coverslip is rinsed twice in 1x PBS before mounting with GLOX for imaging.

#### **4.9 Image acquisition**

All coverslips were imaged using 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 EMCCD camera (Andor Technology). Depending on the purpose of the experiment, each sample is acquired in multiple image stacks which have 0.2  $\mu$ m of spacing in between 81-95 focal planes when studying precise spatial locations, or 0.3  $\mu$ m in between 49-60 focal planes for more general use such as FRET-FISH. To correct for chromatic aberrations and shifts between the channels, TetraSpeck Microspheres (0.1  $\mu$ m, fluorescent blue/green/orange and dark red, Thermo Fisher Scientific) were imaged before each imaging session. The DNA stain channel was used as the reference channel and determined the location of the beads by fitting a 2D Gaussian profile in  $x$ ,  $y$  and a 1D Gaussian in  $z$ .

## 4.10 Image processing and FISH dots identification

### 4.10.1 Applied in iFISH, GPSeq and miFISH

The images after chromatic aberration correction went through 3D automatic segmentation of cell nuclei. First, the DNA staining channel with Hoechst 33342 was deconvolved with the Huygens Professional Software (Scientific Volume Imaging, v17.04) with the following parameters: CMLE algorithm, null background, signal-to-noise ratio equal to 7, and 50 iterations. Then, the 3D segmentation is performed with `tiff_auto3dseg` script in the `pygpseq` Python3 package. For FISH dot picking, DOTTER is a custom designed software written in MATLAB (MATLAB and Statistics Toolbox Release R2020a) and C99 with GSL (<https://www.gnu.org/software/gsl/>) that identifies FISH signals as dots, due to the diffraction limit inherit to the microscopic resolution, based on their intensities and the full width half maximum (FWHM) value for each dot. For each color channel, the field of views were inspected in order to adjust settings to capture the FISH dots. To distinguish between different alleles and analyzed them separately, k-means clustering is applied for each nucleus according with the radius distance set by the user, partitioning all the dots into two clusters in G1 cells. The dots are divided in two different clusters according to the nearest mean (cluster centroid), and the process is iterated until the mean positions with the smallest dot distances within the cluster are found. The  $x$ ,  $y$ ,  $z$  coordinates of each dot were extracted in a .csv file and selected G1 cells based on the Hoechst 33342 signal intensity in the nucleus. For radial position localization of the FISH dots, a Python script, `gpseq_fromfish`, was written to locate the FISH dot in the nucleus according with their distance to the nuclear edge and to the center defined by the DNA staining. The dots distance was normalized by dividing with the nucleus radius, which are different depending on which nuclear edge is defined since the cells are ellipsoidal. Therefore, the voxel numbers of the closest nuclear edge to the FISH dot until the nucleus center is measured.

### 4.10.2 Applied in FRET-FISH

A similar approach for DOTTER picking of FISH dots was pursued for the study of Xi and Xa differences. Briefly, multiple fields of view and each channel by eye were inspected manually by selecting a threshold that seemed to best capture true dots. After, the same threshold was applied to the remaining fields of view. DOTTER allows to threshold the intensity per pixel for each color and as well as to threshold the full width half maximum (FWHM) value for each dot. This semi-automatic method attempts to detect dots in the most unbiased manner. To

reduce the number of false positive and false negative dots, the signals were then inspected by channel in each nucleus, for all processed datasets. For allelic distinction, the dots were labelled with different labeling according to their proximity to the Xist cloud. Then, the x, y, z coordinates of each dot were extracted in a .csv file and selected G1 cells based on the Hoechst 33342 signal intensity in the nucleus.

On the other hand, the stabilization sequence tests and drug treatment relied on a more automatic pipeline that detects the two most intensive signals per G1 nucleus after segmentation. Since the donor, acceptor and FRET signals are expected to be positioned in the same x, y, z coordinates, the co-localization of the 3 different channels provide higher confidence in the FRET signals detection. The chromatic aberrations and mechanical shifts can lead to some position deviation, nevertheless, the FRET channel signals are always co-localized with the acceptor channel because emission settings are shared. The pipeline discards G2 cells and exports all FISH dots x, y, z coordinates and the corresponding signal intensities.

#### **4.11 Signal decoding for miFISH**

The analysis was restricted to G1 cells containing two clearly identifiable dot clusters, corresponding to two chromosome homologues, and computed the 3D coordinates of all the dots in each cluster. For each cluster in each G1 cell, 10 signals were expected to be detected for each fluorescent dye used except for the Alexa Fluor 790 (AF790) channel, where were expected 2 dots, since this channel was not included in any dual-color probe. Hence, for all the channels except AF790, the coordinates of the 10 brightest dots in each cluster were extracted, whereas for AF790 the coordinates of the 2 brightest dots in each cluster were extracted. To match signals coming from the same dual-color probe, all pairwise distances were computed in between all the signals for all the channels. Ideally, the color signals coming from same probe should have a distance close to zero. However, a shift is expected due to chromatic aberrations between signals that in theory should be perfectly aligned. Two additional set of experiments aimed to measure this technical error and identify a cut-off for co-localization calling, so only one dual-color probe was included in the experiment alongside two singly labelled probes targeting different loci on chr2. In order to distinguish the chromatic aberration from mechanical shifts, the two experiments had the dual-color probes either the dichroic mirror positioned in the same multi-band cube or in different multi-band cubes, then requiring change of the cube holding during imaging. The median distance between two signals coming from the same probe labelled with ATTO 542 (AT542) and ATTO 647N (AT647N)—for which dichroic mirrors are placed in the same cube—was lower than 0.25  $\mu\text{m}$ . In contrast, the median distance between two signals coming from the same probe labelled with Alexa Fluor

488 (AF488) and Alexa Fluor 594 (AF549)—for which dichroic mirrors are placed in different cubes and hence more mechanical movement is needed to image both—increased to 0.55  $\mu\text{m}$ . Therefore, in further analyses, only signals with pairwise distance lower than 0.55  $\mu\text{m}$  were retained for experiments in which dichroic mirrors were placed in different cubes, while for experiments in which dichroic mirrors shared the same cube, the threshold was set to 0.25  $\mu\text{m}$ .

Next, the dual-color probes were associated and cut-offs were implemented for dots that split into two by removing dots within a threshold. Any dot positioned more than 5  $\mu\text{m}$  away from all the other dots in the same cluster was considered as an outlier. The dots without dual-color probe assignment were considered as single-color probes and the brightest dots of the cluster that had not yet been selected was assigned to them. The single-color probes have twice the number of oligos labelled with the same dye in the same loci, therefore brighter signals are expected. Altogether, this approach yielded the expected dot counts per nucleus, with signals following the expected FWHM distribution.

#### 4.12 Analysis of chromosome topologies for miFISH

To reconstruct individual chromosome topologies, for each cluster in which all the 16 probes were identified, the dots were interpolated by fitting a cubic spline curve through the dot  $x$ ,  $y$ ,  $z$  coordinates using the `cscvn` command in MATLAB. There is a substantial cluster-to-cluster variability in the shape of the resulting curves. Due to cluster-to-cluster variability in the shape, potential repeating patterns in the splines were inspected such as curvature that was calculated using the **Eq. 1**.

$$k = \frac{\sqrt{(z''y' - y''z')^2 + (x''z' - z''x')^2 + (y''x' - x''y')^2}}{(x'^2 + y'^2 + z'^2)^{\frac{3}{2}}} \quad (1)$$

Then, the curvature values were normalized by identifying the maximum curvature value for each cluster and dividing each curvature value of a given spline by that maximum curvature value of the same spline, so that the curvature values would range from 0 to 1. Only the curvature peaks above 0.1 were considered for the following analysis.

#### 4.13 FRET efficiency calculation

The FRET efficiency is calculated as an emission sensitized assay, that measures the FRET channel and donor channel intensity and calculates the FRET efficiency as in **Eq. 2**.

$$FRET\ efficiency = \frac{FRET\ Inten.}{FRET\ Inten. + Donor\ Inten.} \quad (2)$$

Because the FRET acquisition signal is influenced by other effects such as cross-excitation, bleed-through and cross-talk, controls were introduced for FRET measurements which

consisted in samples or wells that had gone through the same preparation but only contains either the donor or the acceptor probe. Then, the ratio was calculated in between the FRET and donor/acceptor intensities.

#### **4.14 Workflow for GPSeq**

In “Nature Protocol Exchange” (<https://doi.org/10.21203/rs.3.pex-570/v1>) there is a more detailed protocol. In brief, GPSeq workflows starts by subjecting fixed cells to restriction digestion at 37 °C according to the defined time points for each enzyme, ranging from 1 min to 6 h for HindIII (NEB) and 1 to 30 min for MboI (NEB). 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. The phosphorylation step was conducted with calf intestinal alkaline phosphatase (Promega) for 2 h at 37 °C in order to proceed to the ligation step with YFISH or GPSeq adapters and T4 DNA ligase reaction mix (Thermo Fisher Scientific) for 18 h at 16 °C. The unused adapters are washed away with the high-salt buffer (10 mM Tris-HCl/1M NaCl /0.5% Triton X-100 pH 8). Next, the cells were scrapped off from the coverslips and after digested with Proteinase K (NEB) for 18 h at 56 °C. Afterwards, the genomic DNA was extracted, sonicated and *in vitro* transcribed (IVT) using T7 RNA polymerase (Thermo Fisher Scientific) for 14 h at 37 °C. The RNA product of the IVT reaction went through the library preparation with modified TruSeq Small RNA Library Preparation kit (Illumina) and sequenced on the NextSeq 500 platform (Illumina).

#### **4.15 YFISH and imaging**

The previous section explained the digestion and ligation step, then the ligated samples were hybridized with 200 nM of labelled oligonucleotides in a hybridization buffer containing 2x SSC/25% formamide/10% dextran sulphate/1 mg ml E. coli tRNA/0.02% bovine serum albumin. The hybridization reaction was incubated in a humidity chamber for 18 h at 30 °C with the cells placed onto the hybridization mix droplet on a piece of parafilm. Afterwards, the samples are washed with 2x SSC/25% formamide for 1 h at 30 °C. Then, the nuclei were stained with 2x SSC/25% formamide/0.1 ng/μl Hoechst 33342 for 30 min at 30 °C and later rinsed with 2x SSC and mounted in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). The wide-field imaging used the same setup described for DNA FISH experiments.



## 5 RESULTS

### 5.1 Paper I: iFISH is a free resource enabling versatile DNA FISH to study genome architecture in single cells

The increased interest in using DNA FISH to study genome organization has exposed the limitations of the FISH method, such as: high probe production cost, lengthy protocols, compromised probe specificity, low resolution and low portfolio of ready-to-use probes. Until 10 years ago, typical DNA FISH probes were made using bacterial artificial chromosomes (BACs) that were used as a template in a PCR where a set of random primers would amplify long DNA fragments. Those would subsequently be fragmented by DNase I, yielding dsDNA probes ranging in size from 200nt to 500nt. BAC-based probes however suffered from the limitations mentioned above. A next step was the design of a database of unique 200-mers ideally suited for DNA FISH and an accompanying database of primer pairs to amplify them from the human genome. This approach, called HD-FISH<sup>112</sup>, allowed for the generation of highly specific probes that allowed specific and sensitive detection of DNA loci as small as 3 kb. However, it was the switch from such dsDNA to oligo-based ssDNA probes that revolutionized FISH. This was possible thanks to the oligopool technology, which allows for production of complex oligo libraries in a cost-efficient manner, as well as computational approaches that allow for identification of DNA sequences that would perform well as oligo probes.<sup>113</sup> An oligopool can contain thousands of different oligos belonging to different probes to which a specific barcode sequence is added to allow selective amplification of each oligo. The inclusion of the T7 promoter in such oligos allows their in vitro transcription, which if followed by the reverse transcription step and RNA digestion, do eventually yield ssDNA probes. These ssDNA probes represent the original sequences present in the oligopool and are in high concentration suitable for DNA FISH. Thanks to their superior specificity, probes produced this way can target much smaller genomic windows, thereby increasing the resolution of DNA FISH. Since the development of the Oligopaint approach described above there have been numerous further improvements to this technology, always to increase the production yield, increase multiplexing and lower the cost.<sup>114,115</sup> For example, MERFISH introduced a high-throughput RNA FISH (and later DNA FISH) method to produce thousands of probes, which are all simultaneously hybridized to the genome.<sup>37</sup> Afterwards, fluorescently labelled oligos that have unique sequences, are added sequentially to hybridize to the complementary sequences present in the target-specific oligos, in a two-step hybridization process<sup>42</sup>. Even though the state-of-the-art of DNA FISH has improved significantly over the past decade, DNA

FISH is still not accessible to many labs, probe design remains a challenge to most, let alone their production in-house. Until recently there were no databases with FISH-suitable DNA sequences and a complete lack of user-friendly interfaces that would assist non-experts in probe design. OligoArray<sup>116</sup> and PROBER<sup>117</sup> were the first software packages created to aid the design of FISH probes, however they required extensive bioinformatics expertise. OligoMiner<sup>118</sup> provided a freely accessible database of oligos for the human genome, however the genome was covered only sparsely. Therefore, we developed a new method named iFISH, which provides a 3D DNA FISH protocol, a freely accessible oligo database that contains 78% more oligos than OligoMiner, and an interface for probe design.

#### 5.1.1 iFISH probe design

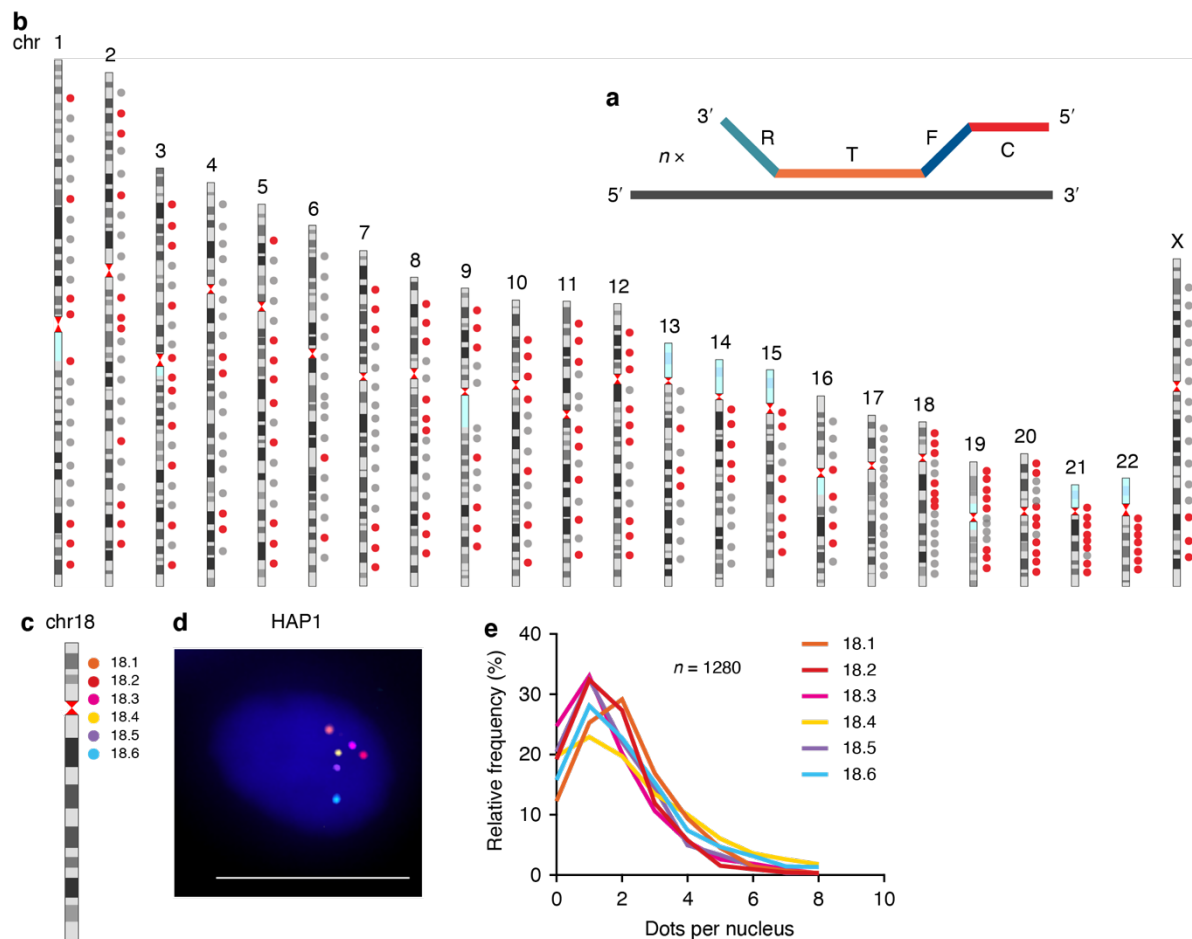
In iFISH each probe consists of 96 oligos that balance the signal-to-noise ratio (SNR) of the FISH signals and the resolution, i.e., the smallest possible genomic region to be detected. The median size of the 330 iFISH probes which were designed in this work, with many tested experimentally, is 7.6 kb. Each oligo is comprised of multiple sequences that make up the full oligo structure (**Figure 6a**). The target sequence (T) is the sequence present in the iFISH database that hybridizes to the genomic region of interest and is 40 nt long. Both 5' and 3' ends of the oligo have adapters, the forward (F) and reverse (R) primers that are 20 nt long. An extra sequence of 20 nt is added adjacent to the forward primer that is complementary to one of the multiple dye-labelled oligos and therefore this sequence is called the color barcode (C). None of the 20 nt sequences hybridize to any site of the human genome, in other words they are orthogonal to the human genome, and these were extracted from a previously available database of 240,000 25-mers<sup>119</sup> in a computationally-demanding process. Thanks to the similar amplification kinetics of the various oligos tested, the same PCR protocol can be used for all the probes.

Given the relatively low and heterogenous genome coverage of the OligoMiner database, the iFISH project also involved the creation of a new, more comprehensive FISH oligo database. To test whether the iFISH database indeed provides better results than the OligoMiner database, two different sets of DNA FISH experiments were performed with probes designed using one or the other database: (1) an OligoMiner probe with the same number of oligos as in iFISH (96) but targeting a larger target size, and (2) an OligoMiner probe targeting the same target span, but with a lower number of oligos than in the iFISH probe. As expected, in both cases the signal to noise ratio of the FISH signals was lower in the case of the OligoMiner probe. When looking at the dot counts in the first experiment, the OligoMiner probe with the same number of oligos as the iFISH probe showed the same

distribution of dots per cell as observed for the iFISH probe. However, in the second experiment, 35-63% of the cells hybridized with the OligoMiner probe did not show any detectable signal due to the lack of oligos providing signal. The OligoMiner probes were designed to have more strict GC content range, however, this property seems not to compensate for the lack of oligos spanning the region of interest.

#### 5.1.2 Validation of the iFISH probes

A large-scale validation effort was carried out in the iFISH project, aimed at analyzing 153 out of 330 probes produced in different colors (**Figure 6b**). In total, 162,305 FISH dots from 47,747 HAP1 haploid chronic myeloid leukemia cells were identified as true signals in a semi-automated manner. To increase the multiplexing and throughput of the assay, the microscopy setup was upgraded to discriminate between up to seven fluorescent dyes identical or equivalent to DAPI, Alexa Fluor 488, ATTO 542, Alexa Fluor 594, ATTO647N, ATTO700 and IR800CW dyes. In a proof-of-principle experiment, six iFISH probes targeting 6 consecutive DNA loci on chr18 were visualized in 6 different colors (**Figure 6c, d**) and 74% of the cells displayed 1-4 dots per nucleus per channel (**Figure 6e**). The experiment proved that it is possible to perform a 7 channel-imaging using the iFISH probes, opening up new opportunities for more multiplexed genome studies.

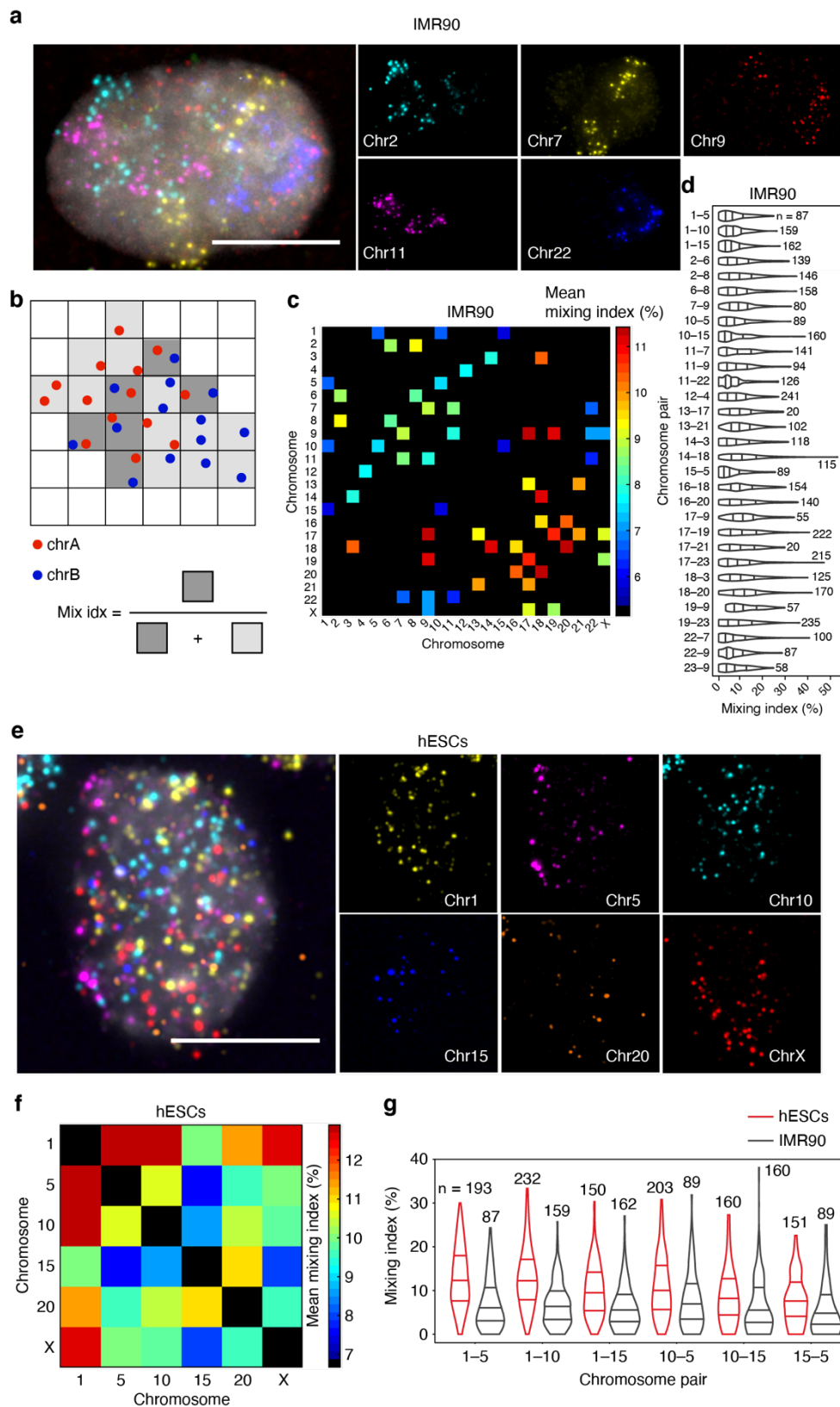


**Figure 6** - iFISH probe design implementation. (a) Scheme of oligos present at a certain number ( $n$ ) in the iFISH probe and each oligo has a different target (T) sequence. The different regions are described in the text above. (b) 330 iFISH probes targeting all autosomes and chrX. The red dots resemble probes that were individually analyzed, while probes labelled grey were included in the spotting probes together with the probes colored with red. (c) Ideogram of the 6 probes targeting chr18 as in (b, red dots) with different colors. (d) A representative HAP1 cell with the probes represented in (c) with the same color code. Blue, DNA, Scale bar, 10  $\mu$ m. (e) The probes numbered in (c) are shown with their relative dot frequencies per nuclei.

### 5.1.3 Chromosome territories and intermingling

iFISH probes were shown to improve the specificity and signal-to-noise ratio of DNA FISH signals, making them suitable for studying not only individual DNA loci but also chromosomal territories, as the lower risk of unspecific signals aids the proper identification of the boundaries of chromosomes. Additionally, the fact that with iFISH 6 colors can be used simultaneously (in addition to the DNA stain by DAPI or alike) allows for detection of 6 chromosomes in any given experiment (**Figure 7a**). In iFISH the detection of individual chromosomal territories is achieved by designing a set of iFISH probes targeting multiple small regions (< 10kb) spaced equidistantly along the chromosome. A set of iFISH probes that visualize a chromosomal

territory is called a “spotting probe”, and we designed spotting probes for all human chromosomes, with chromosomes 1 to chr16 probed every 10 Mb and all other chromosomes spotted every 5 Mb. In this way, a given chromosomal territory can be visualized as a cloud of discrete dots. This way of chromosome detection conceptualized a novel metric for quantitative assessment of inter-chromosomal intermingling, which had been challenging thus far. We measured the extent of mingling between different chromosomes as the extent of regional overlapping (mixing) of single dots coming from two different chromosomes. For this, 3D images were split into small cubes and for each chromosome pair, the ratio of cubes containing signals from both chromosomes was calculated over the sum of cubes containing any of the two chromosomes (**Figure 7b**). This analysis was applied to diploid human fibroblasts (IMR90) and to human embryonic stem cells (hESCs). 31 chromosome pairs in IMR90 cells were analyzed and the average mixing index between short chromosomes was found to be higher than between the longer chromosomes (**Figure 7c,d**) which is in line with Hi-C measurements.<sup>7</sup> In hESCs, chromosome territories were found to be more sparsely spread throughout the nucleus than in IMR90, and some chromosomes completely lacked a defined location in the nucleus (**Figure 7e**), which was to our knowledge the first example of cells without a clear territoriality of chromosomes. Consequently, chromosome intermingling was found to be considerably higher in hESC (**Figure 7f,g**). In conclusion, the chromosome spotting probes allowed for a more resolved study of chromosomal territories, including the intermingling properties.



**Figure 7** – Chromosome territories and intermingling for IMR90 and hESCs. (a) An example of an IMR90 cell with 5 chromosomes targeted with iFISH probes. Scale bar: 10  $\mu$ m. (b) A representation of how the mixing index was calculated between the spots coming from two different chromosomes represented in different colors. The calculation is done in 3D space. (c) A map of the average mixing index for 31 chromosome pairs in IMR90. (d) Distribution of the 31 chromosomes pairs in (c). (e) A

representative example of a hESC showing that most of the 6 chromosomes do not show territoriality.  
(f) Average mixing index for 6 chromosome pairs shown in (e). (g) Comparison of the mixing index of 6 chromosome pairs in IMR90 and hESC as violin plot.

## 5.2 Paper II: GPSeq reveals the radial organization of chromatin in the cell nucleus

Genomic loci Positioning by Sequencing (GPSeq) originated from the reasoning that there should be a way to control radial diffusion of proteins through the meshwork of crosslinked chromatin, thereby enabling the study of the radial organization of the genome. Restriction enzymes were chosen to digest the radial layers of chromatin, making the set-up amenable for sequencing-based read out. In GPSeq, digestion of chromatin occurs gradually from the periphery towards the center, which is achieved by halting restriction enzyme activity at gradually increasing times that allow the progression of the digestion to be stopped at successive positions relatively to the center. The fragmented DNA from each sample is then collected and used to prepare sequencing libraries.

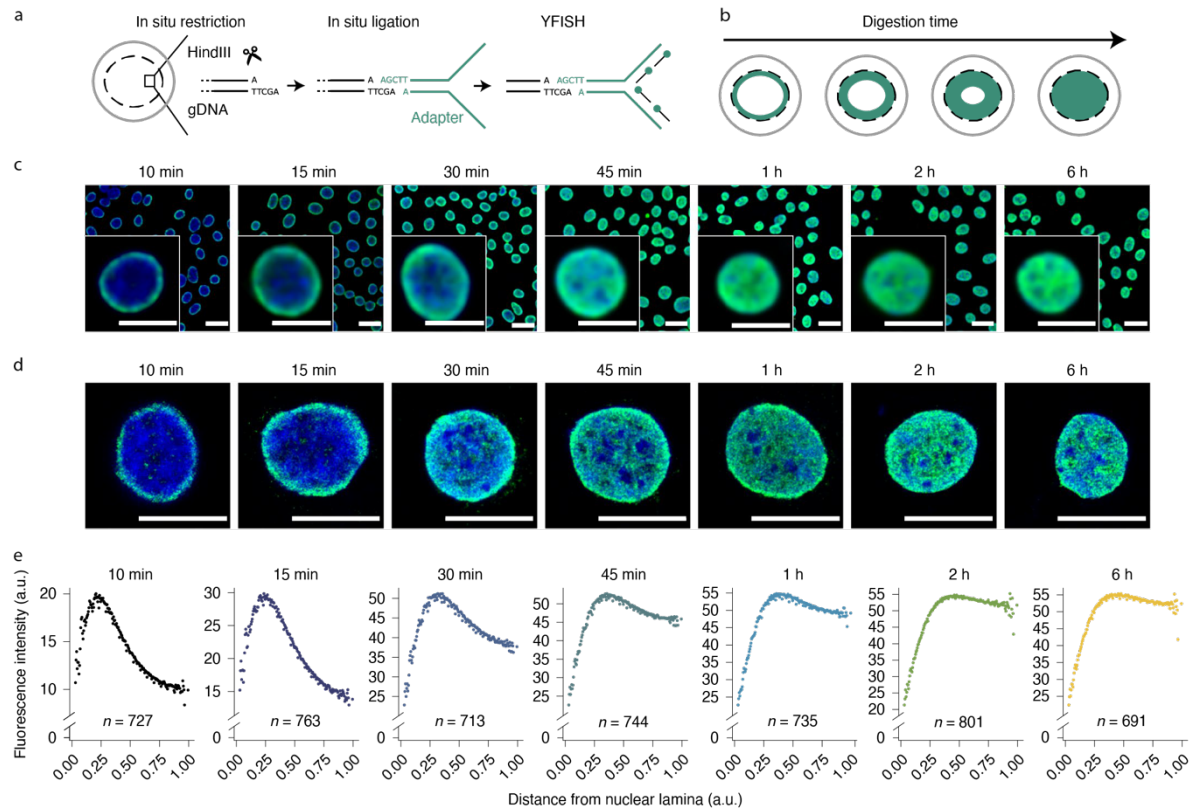
To develop the GPSeq method, HAP1 cells were chosen for their haploidy and circular shapes. Importantly, diploid cells can only have their chromosome position reliably measured with GPSeq if the homologous chromosomes can be distinguished using, for example, frequently occurring SNPs. Otherwise, the radiality score of a locus would be the average of the two homologous alleles. The reasoning behind choosing circular cells is that cells with more ellipsoidal shapes have different radius distances, which might cause the progression of the digestion in the areas with a shorter radius to be faster, yielding a mix of fragments containing both these regions and fragments from the more elongated sides of the cell nucleus.

To visualize the diffusion of the restriction enzyme and thus the assay the digestion across the nucleus, a new fluorescence *in situ* hybridization assay was developed and named YFISH. After digesting the samples with restriction enzymes for increasing periods, the generated restriction overhangs are ligated to the compatible overhang of a Y-shaped adapter (**Figure 8a**). The non-complementary ends of the adapter create an opening, giving the Y shape, and the single stranded end allows the fluorescently labelled oligos to hybridize to them. In this way, the digestion progression can be easily monitored under a fluorescence microscope, which is essential for protocol optimization and in particular to optimize the different digestion times (**Figure 8b**). The imaging of the YFISH samples was mostly done with wide-field microscopy in order to acquire higher number of cells (**Figure 8c**), while the more detailed imaging was done using stimulated emission depletion (STED) (**Figure 8d**). The successfully and homogeneously digested samples from different time points then are then proceeded to the sequencing step.

The enzymatic restriction reaction was initially carried out with a 6-base cutter – HindIII - and comprised 6 conditions in HAP1 cells: 10 min, 15 min, 30 min, 1 h, 2 h and 6 h. For 200 randomly selected nuclei, we measured the radial signal intensity and the distributions demonstrated that the radial spread is homogenous in every nucleus (**Figure 8e**). As the HindIII



enzyme is GC-biased and as such has a bias for cutting genomic regions rich in GC bases, the second restriction enzyme selected for replicating the results was the 4-base cutter MboI, which has an opposite bias. Since the second enzyme produces different restriction ends, new YFISH and sequencing adapters were designed to allow binding to the sticky ends created by the MboI enzyme. Additionally, the digestion protocol was optimized for shorter time points: 1 min, 5 min, 10 min, and 30 min. The MboI enzyme reproduced the GPSeq scores from HindIII (PCC: 0.68 at 100 kb resolution and PCC: 0.91 at 1 Mb resolution).



**Figure 8** – GPSeq implementation scheme. (a) YFISH scheme for in situ restriction with HindIII enzyme and ligation with the YFISH adapter. (b) Radial progression of the digestion towards the nuclear center at increasing digestion times. (c) Progression of the digestion ring in YFISH images generated with epifluorescence microscope. (d) Same as in (c) with STED imaging. (e) Signal intensity progression along the digestion times.

### 5.2.1 GPSeq implementation

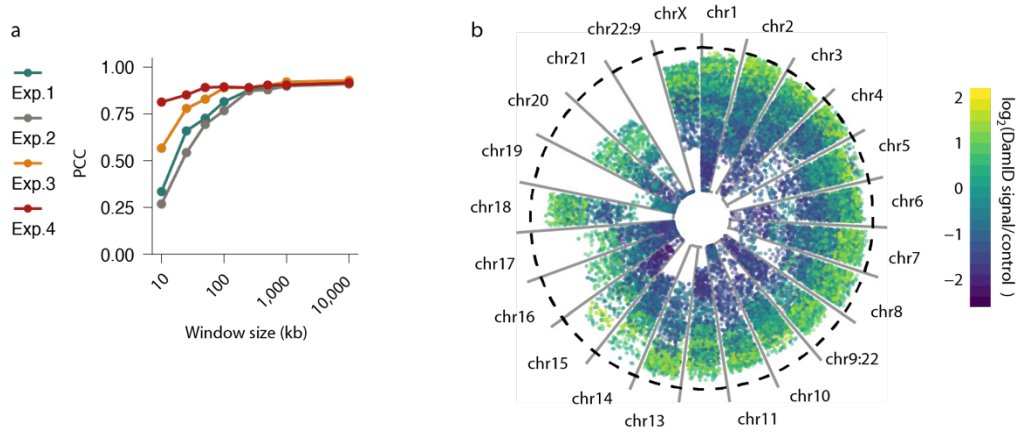
After the YFISH test results gave us “the green light” to continue, the development phase proceeded to the sequencing step. In order to prepare the samples for Next-Generation Sequencing (NGS), the digested strands are ligated with NGS-compatible adapters that consist of five distinct parts: (1) the overhang sequence compatible to the restriction overhang created by the restriction enzyme, (2) a barcode specific for each time point of digestion, (3) a unique molecular identifier (UMI) that is specific for (virtually) each ligation event and that is used to

remove amplification bias introduced in the PCR step, (4) an Illumina RA5 adapter, and (5) a T7 promoter that works as a primer for the IVT step. After ligating the adapter, the genomic DNA is extracted, fragmented through sonication, and linearly amplified in the IVT step. The IVT amplification step uses the T7 bacteriophage polymerase and yields a linear amplification of all fragments, thereby substantially increasing the library complexity. Finally, NGS libraries are prepared and loaded at equimolar concentrations onto the sequencer. During data analysis, the read counts representing fragment ends sequenced across the genome are analyzed for each timepoint of digestion. Then, the centrality of a given genomic window is defined based on the normalized number of sequenced restriction fragment ends mapping to that window at the defined timepoints. For example, when a genomic window displays high read counts in the shortest timepoints, this reflects a peripheral localization of the region.

### 5.2.2 GPSeq cross-validation with independent techniques

For GPSeq validation, the iFISH probes from paper I were used to study the radial localization of several selected DNA loci, after which the results from DNA FISH and GPSeq were compared. The FISH scores for radially were calculated by taking the distance of the signal to the nuclear center, divided by the normalized nuclear radius as explained in the methods section. All FISH distance measurements per locus from all cells analyzed were then averaged and compared with the GPSeq score for that locus. The FISH and GPSeq scores correlated highly, even at 100 kb resolution (**Figure 9a**).

Another method that we used to validate GPSeq is DNA adenine methyltransferase identification (DamID), which maps chromatin association with intra-nuclear structures such as the nuclear lamina.<sup>1</sup> Similarly to ChIP-seq, DamID identifies genome-wide interactions between DNA and a protein of interest. The difference with ChIP-seq is that DamID does not use antibodies to isolate the protein-DNA complex. Instead, a *E. coli* DNA adenine methyltransferase (Dam) is fused to the DNA binding domain of a protein of interest which promotes adenine methylation present in the vicinity of the protein of interest. DamID was implemented to identify genomic associations with nuclear lamina proteins such as Lamin B, which led to the identification of genomic regions frequently associating with the nuclear lamina, termed Lamina Associated Domains (LADs).<sup>120</sup> Here, LADs identified in Lamin B-DamID experiments on HAP1 cells were employed to validate the radial GPSeq score. As expected, the regions that are enriched at the nuclear periphery, as identified by DamID, agree with the GPSeq scores that predicted those domains to be closer to the nuclear lamina (**Figure 9b**).



**Figure 9** – Validation of GPSeq with independent techniques. (a) Pearson’s correlation obtained between DNA FISH and GPSeq data at different genomic window sizes. Exp. 1 and Exp. 2 are two GPSeq replicates with HAP1 cells and HindIII restriction enzyme. Exp. 3 and Exp. 4 are two GPSeq replicates with HAP1 cells and MboI restriction enzyme. (b) Pizza plot with radial representation of DamID signal for each chromosome, with dots representing 1 Mb genomic windows radially arranged according to their GPSeq score.

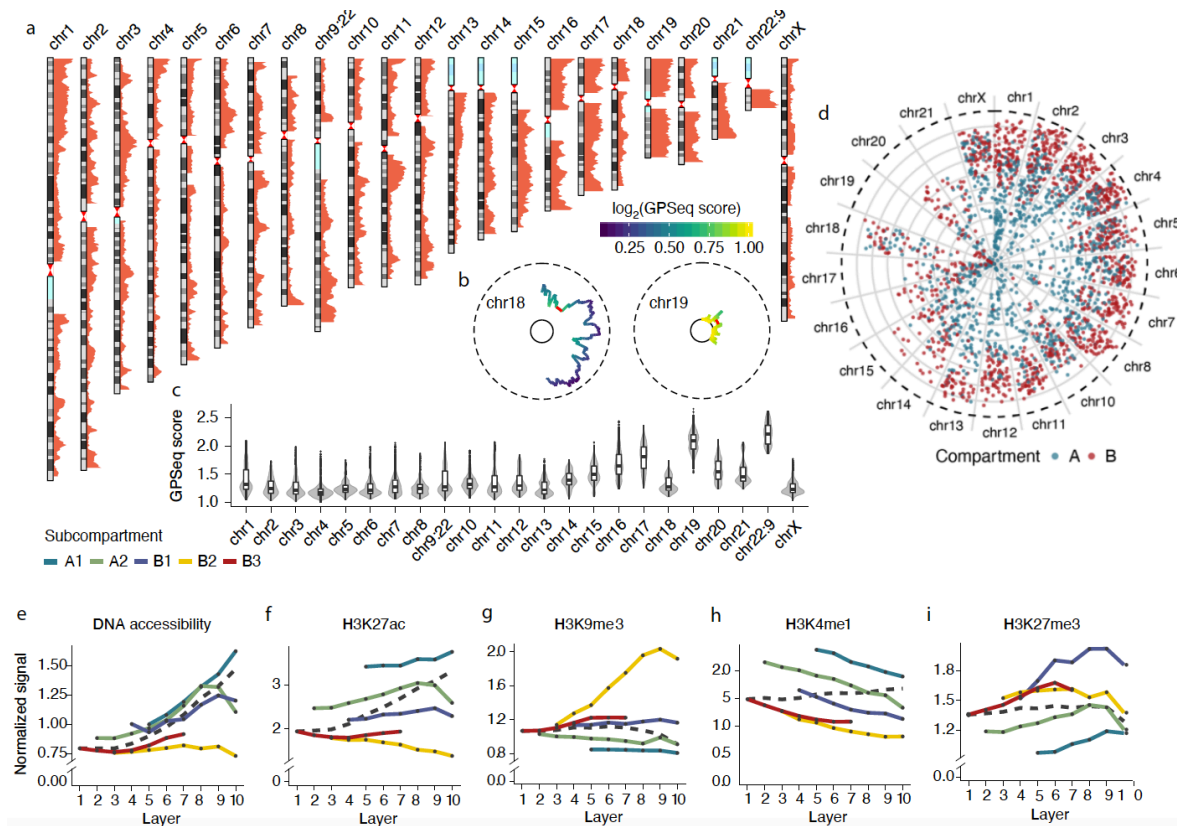
### 5.2.3 Validation of radial genome organization models

GPSeq revealed that chromosomes display a heterogeneous radial distribution along their length (**Figure 10a**). However, the smaller chromosomes (chr15 to chr22) are almost entirely positioned at the center, except for chr18 (**Figure 10b**). Therefore, the relative radial positioning of the chromosomes seems to favor that small chromosomes are packed in the center and absent from the outer layers, while longer chromosomes stretch from periphery to center at variable levels. However, the correlation between chromosome size and radially is modest (PCC: -0.725), indicating that the size on its own cannot be used as an accurate predictor of radially (**Figure 10c**).

To inspect other centrality predictors, chromosome regions were split into two different categories according to their gene density. Likewise, this feature showed a modest correlation with the GPSeq score (PCC: 0.710). Gene expression was also contemplated as a possible predictor, showing a similar modest correlation (PCC: 0.692). Lastly, GC content was considered, which is typically high in protein coding regions, and showed the highest correlation with the GPSeq score at both 1 Mb and at 100 kb (PCC: 0.813 and 0.666, respectively). Given that not a single feature, whether genetic or epigenetic, showed a high predictive power with regards to chromosome radially, a multivariate model was built to incorporate all of the features. This multivariate model was shown to have a great predictability power at the scale of entire chromosomes ( $R^2 = 0.939$ ; prediction error (PE) = 0.073), where using information only about chromosome size and GC content was enough (with extra

features not adding to the predictive power). At 1 Mb the model included information about chromosome size, gene density, gene expression and GC content, which resulted in a PE value of 0.12 with an  $R^2$  of 0.741.

As explained in the introduction, Hi-C maps divide the genome into A and B compartments that resemble euchromatin and heterochromatin, respectively. Here, we leveraged publicly available Hi-C data from HAP1 cells and plotted the genomic windows assigned to the A and B compartments in a radial manner according to their corresponding GPSeq score. As expected, regions belonging to the A compartments were shown to be present in the most central layers while regions assigned to B compartments were positioned in the outmost layers (**Figure 10d**).



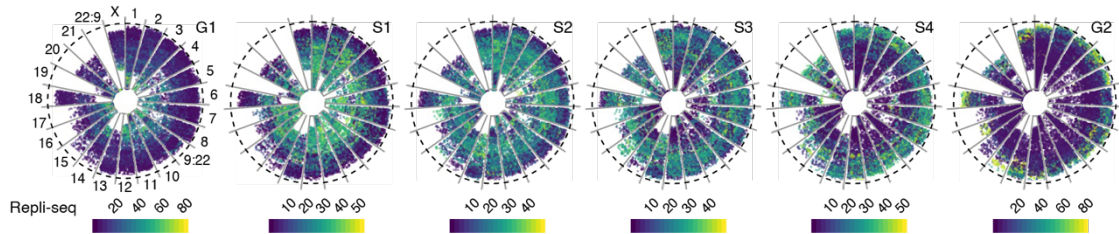
**Figure 10** – Radial representation for each chromosome. (a) GPSeq score profile for each chromosome with higher peaks representing central location and lower peaks represent peripheral location. (b) Radial representation of two small chromosomes, chr18 and 19, with opposite behaviors. (c) Violin plot for all the windows' GPSeq score for each chromosome. (d) Preferential location of each 1 Mb genomic windows (dots) by layer, and the dots are colored as blue (compartment A) or red (compartment B). (e–i) Each subcompartment (A1, A2, B1, B2 and B3) shows the normalized signal of DNA accessibility and histone marks for each radial layer in 100 kb resolution. The overall distribution without subcompartment stratification is represented in the dashed line.

The A and B compartments defined based on Hi-C data have been further subdivided into sub-compartments (A1-2, B1-3) in accordance with the epigenetic marks associated with distinct chromatin states, as described in Rao *et al.*<sup>121</sup>. Briefly, the differences between A1 and A2 are as follows: the A2 subcompartment has later replication time, higher H3K9me3 presence, lower GC content, and contains longer genes than A1. With respect to the heterochromatin sub-compartments, B1 is enriched in the H3K27me3 mark but depleted of the H3K36me3 mark, replicated in mid S phase, and considered as facultative or repressed heterochromatin. Genomic windows belonging to the B2 and B3 sub-compartments are associated with the nuclear lamina, with B2 additionally found at pericentromeric regions and containing nucleolus-associated domains (NADs).

Here, we used this subdivision of chromatin to study the radial distribution of 5 different epigenetic features, as well as DNA accessibility described for those sub-compartments. This was achieved by intersecting various ChIP-seq datasets as well as ATAC-seq data with our GPSeq results. Firstly, DNA accessibility was shown to increase towards the center in all the sub-compartments, except for the constitutive heterochromatin B2 and B3, in which the accessibility seemed unchanged along the nuclear radius. Interestingly, genomic windows of the B1 type, which are present in the very center of the nucleus, show higher accessibility than those belonging to the most transcriptionally active sub-compartment – A1 – if the A1 windows are more peripherally located (**Figure 10e**). Secondly, H3K27ac, which marks transcriptionally active chromatin, seems to be increasing towards the center genome-wide, but depleted in the B2 sub-compartments of the center (**Figure 10f**). Thirdly, H3K9me3, which marks heterochromatin, goes down very mildly towards the center genome-wide and yet the B2 sub-compartments in the center are highly enriched in it. A possible explanation for these observations is that B2 sub-compartment located at the very center may be protected from being activated by this higher deposition of silencing marks and depletion of activating marks, to maintain its heterochromatin state inside the most active nuclear area – the center (**Figure 10g**). Fourthly, H3K4m1 which marks active and poised enhancers, shows depletion of this mark towards the center for every subcompartment (**Figure 10h**). H3K4m1 is mostly present in A1 which is in turn mostly located at the center of the nucleus and consequently, in bulk analysis this histone modification is observed more in the center. Lastly, H3K27me3, whose deposition is mediated by the Polycomb repressive complex and characteristic of the B sub-compartment, seems enriched in the middle layers of the nucleus along its radius (**Figure 10i**).

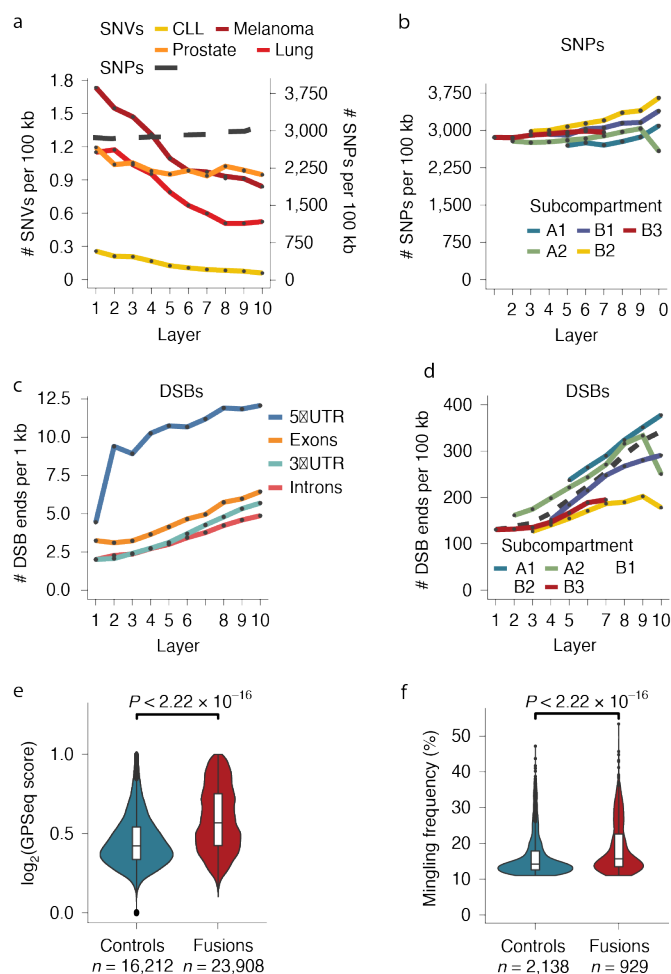
Additionally, the GPSeq maps were intersected with DNA replication timing data acquired using the Repli-seq method<sup>122</sup>, which identifies genomic windows that undergo replication in early S phase (S1), mid S2 and S3, or late S phase (S4). As expected, replication

timing seems to be radially arranged, with the central portions of the genome replicating first and those most peripheral replicating as last (**Figure 11**).



**Figure 11** – Repli-seq signal at 1 Mb resolution radially presented for each chromosome along the different cell cycle phases.

To understand how mutational processes relate to chromatin radiality, the GPSeq score of HAP1 was integrated with single-nucleotide polymorphisms (SNPs) classified by the 1000 Genome Project and cancer-associated single-nucleotide variants (SNVs) obtained from publicly available data of four different cancer tissues. Most of the SNVs were observed at the periphery of the nucleus, in line with the so-called bodyguard hypothesis that states that the outer layer of heterochromatin protects the more inner euchromatin from mutagenic agents<sup>123</sup>, while SNPs show mildly opposite trend (**Figure 12a**) especially when the analysis was stratified by the sub-compartments (**Figure 12b**). Interestingly, double strand breaks (DSBs) sites that were detected using another NGS-based method called breaks labeling in situ and sequencing (BLISS)<sup>124</sup>, were shown to be more frequently located at the center irrespectively of which part of the gene they reside in (**Figure 12c**) or their sub-compartment assignment (**Figure 12d**). Moreover, DNA regions that carry genes often found engaged in cancer gene fusions tend to be more centrally located (**Figure 12e**) and they tend to engage more in intermingling events than control regions (**Figure 12f**) indicating a certain mechanism of their formation likely takes place in the very center of the nucleus in between highly expressed regions that engage in frequent inter-chromosomal mingling.



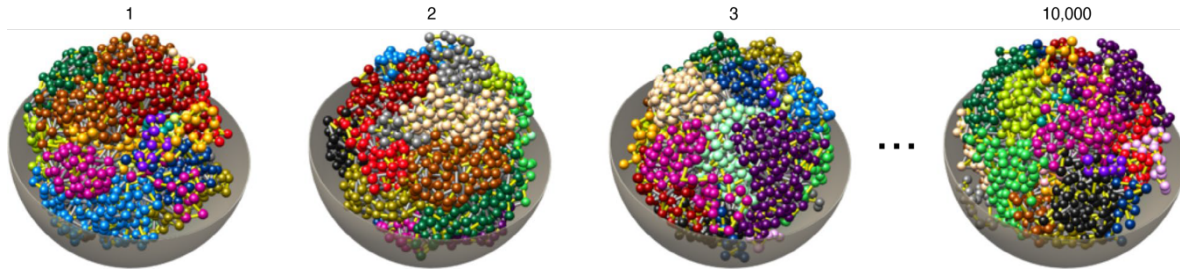
**Figure 12** – Mutations and DNA breaks along the radial distribution. (a) Radial representation of SNVs for 4 different cancer tissues and SNPs. (b) SNPs divided per sub-compartment and represented radially. (c) Endogenous DSB distribution in HAP1 cells towards the center, stratified according to the genic region. (d) The DSBs divided by sub-compartment and distributed by radial layers. (e) Violin plot for GPSeq score in all controls or fusion regions annotated in the cancer atlas of TCGA. (f) The chromosome arrangements were generated with the modeling software, *chromoflock*, where the 10% of genomic windows that are mingling the most with other regions are analyzed for fusions occurrence in comparison to non-overlapping windows (controls) based on annotated in TCGA.

## 5.2.4 3D genome modeling

Important previous work has attempted to recreate 3D genome models using Hi-C data. However, when our lab generated 3D models of the genome structure using the state-of-the-art tool for generating ensembles of single-cell structures using Hi-C data only, called PGS<sup>125</sup>, such structures did not correlate well with iFISH measurements. Then, a new tool named *chromflock* was developed for 3D genome modeling using both Hi-C and GPSeq data. The *chromflock* structures (at 1 Mb resolution) generated based on Hi-C data only did not correlate



well with our DNA FISH data (PCC: 0.421). The combination of Hi-C and GPSeq datasets from HAP1 could potentially improve the structures reliability, since they probe complementary aspects of genome architecture; specifically, Hi-C informs on DNA contacts while GPSeq complements this by conveying the radial distribution of the sequences. Those new 1 Mb structures were generated (**Figure 13**) and a considerably higher correlation with DNA FISH was observed (PCC: 0.947).



**Figure 13** – Examples of 4 out of 10,000 structures generated by *chromflock* by combining Hi-C with GPSeq data at 1 Mb resolution which is represented as a bead.



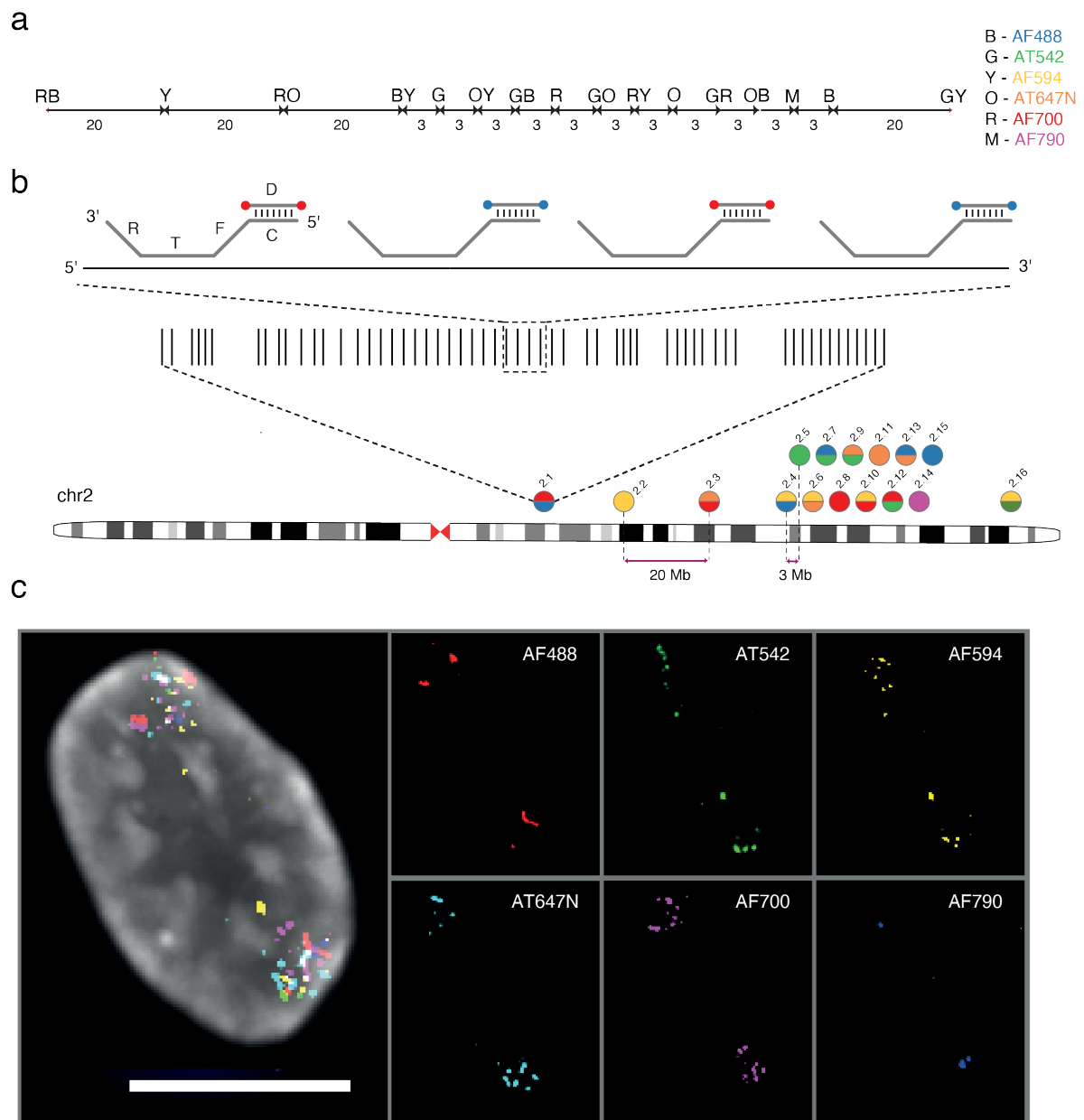
### **5.3 Paper III: Simultaneous visualization of DNA loci in single cells by combinatorial multi-color iFISH**

In this project, the focus was on the higher-order folding of entire interphase chromosomes. The online set of oligos used for iFISH probes was also employed to design probes with multiple colors in the same experiment, to simultaneously detect multiple reference points along a given chromosome. Here, we aimed to visualize as many single DNA loci along a relatively large fragment of a chromosome as possible and connect the FISH signals following the linear genome, in order to study the higher-order arrangement of chromatin within individual chromosomes. This project was motivated by the recent discovery of an unexpected folding pattern of mitotic chromosomes, which seem to adapt a solenoid-like structure with 20 Mb turns all along the chromosome length.<sup>34</sup> The motivating hypothesis for this project was that such a helical twisting of entire chromosomes is preserved to some extent in interphase cells.

Initially, images generated in the context of the iFISH project using the probes from the iFISH repository were analyzed, in which 6 consecutive loci on chromosome 18 were detected (**Figure 6c, d**). Unfortunately, we soon realized that 6 loci were not enough to study this property of chromosomes, given the very high variability of structures present in single cells. This limitation prevented us from detecting any structural patterns, such as a helix or a rosette, since the 6 loci were heterogeneously arranged in the 3D nuclear space. Therefore, we needed to increase the number of loci detected individually in one single experiment. This was achieved by encoding information about a given DNA locus by making use of unique color combinations. Nevertheless, the chromosome topologies obtained from using datasets with these color combinations still demonstrated that interphase chromosomes have extremely heterogenous structures. We then re-purposed this project to become a public data descriptor, since there is an acute need for high quality, open access, curated and reproducible DNA FISH datasets. As a result, the 9 datasets used for this project were uploaded in Figshare and IDR platforms to be publicly available, allowing one to access both the raw images and the 3D coordinates of all dots identified in this project. These datasets might be important for other researchers to apply to their own scientific questions that can involve many different subjects like 3D genome architecture studies, relating sequencing data such as Hi-C or ChIP-seq with direct probing of the nucleus as is done by FISH, and may possibly lead to better automatic DNA FISH analysis algorithms.

### 5.3.1 miFISH probe design

In the iFISH paper, the number of loci imaged in one round of hybridization was increased to 6 loci because 6 different colors were implemented without detectable crosstalk or bleedthrough in the other channels. Paper III aims at expanding this number to 16 distinct loci so that in one single hybridization more information can be acquired. To accomplish this, 5 out of the 6 dyes available are used in different combinations, while the far-red dye Alexa790 remains as a single-color probe to be an experiment control. The resulting technique consisted of a multi-color DNA iFISH (miFISH) and follows a workflow similar to iFISH to generate the datasets. For the 16 miFISH loci we used oligos already present in the iFISH database and we took advantage of the iFISH4U webtool for their design ([www.ifish4u.org](http://www.ifish4u.org)). The probes were placed along the q-arm of chr2 because the arm is particularly long, providing enough space to design probes without centromere interference. The distance between 11 probes is 3 Mb, to sample in relative smaller steps the high-order structure, and for comparison 3 probes were placed 20 Mb away from those (**Figure 14a**). The oligo number in each probe is 700, which is higher than for iFISH probes to accommodate having 2 alternating colors and to increase the detection rate. In turn, a single probe spans a larger region in the genome than in iFISH in which the distances were around 10-20 kb, while miFISH probes span 62-73 kb. The oligo sequence design is the same as in iFISH, having a reverse primer (20 nt), target region (40 nt), forward primer (20 nt) and color barcode (20 nt). However, the multi-color probes have the color barcode alternating between the consecutive oligos to yield color mixing homogeneity, and the forward primer in these probes also alternates to allow separate amplification (**Figure 14b**). To maximize the number of probes used, 10 probes have two colors alternating, out of combinations of 5 colors, while counting as well in the use of 6 probes labeled with a single color out of 6 colors (**Figure 14b**). Besides, to complete the observations made with the miFISH probes, classical iFISH probes were retrieved to target chr1, 2 and 10 with evenly spaced probes, with each probe having only a single color. In total, we used 60 additional iFISH probes in this project, which are also publicly available on the same platforms, to validate the pairwise distance patterns observed in the miFISH dataset, as will be explained in the coming sections.



**Figure 14** – miFISH experimental design. (a) The 16 miFISH probes are located along chr2 with varying inter-probe genomic distances. Each single or double-color probe is assigned one or two letters that indicate the fluorescent dye(s) for probe labelling. (b) The same probes represented in (a) are positioned along chr2 and the individual oligos (vertical black lines) composing probe 2.1 are shown as an example. Each oligo is composed of 4 parts as already introduced in iFISH. T, target sequence. F and R, adapter sequences for selective amplification. C, color sequence to which the labelled oligo (D) binds. (c) An example of a retinal pigment epithelium (RPE) nucleus imaged with an epifluorescence microscope, where all the 16 miFISH probes are detected. Scale bar: 10  $\mu$ m. The smaller panel detects all the miFISH probes for all 6 fluorescent dyes as expected in (a). The larger panel contains the DNA stained with Hoechst 33342.

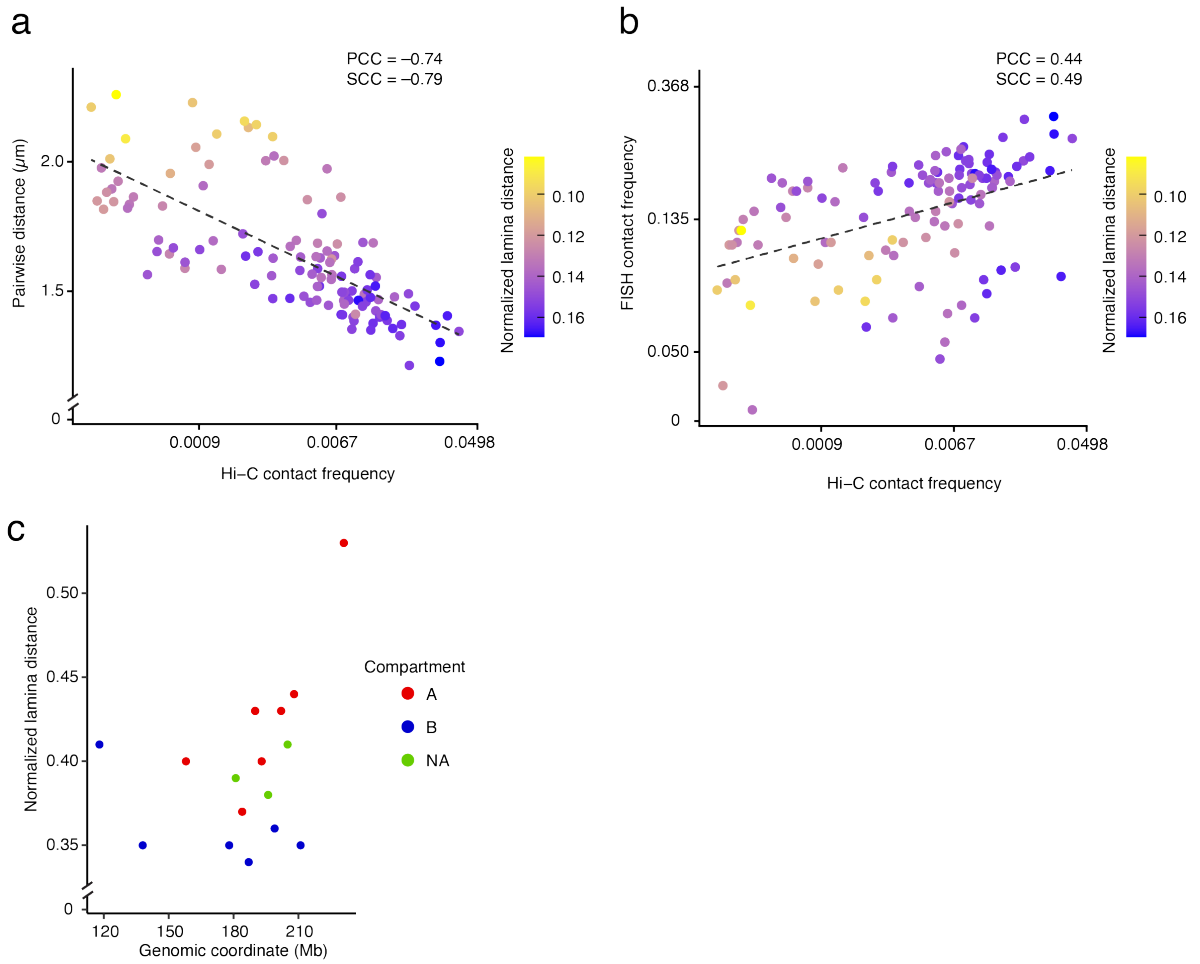
For the signal identification, the 3D coordinates of dots coming from different color channels were overlapped to co-localize the dots (**Figure 14c**). Subsequently, the 16 probes

were identified based on the expected color or color combination. The matching distance between the dots coming from different channels should be close to zero, due to the chromatic correction. Nevertheless, the technical errors were also measured with some extra experiments in which only one multi-color labeled probe was tested and the matching distance for each allele was quantified. These sets of experiments defined the cut-off for co-localization, so that a multi-color probe can only be found if the co-localization distances are lower than this defined cut-off. Therefore, the defined thresholds limited the detection of all 16 probes per allele which led to 37 out of 466 (8%) alleles identified based on the dots clustering to contain all the 16 probes. Most of the alleles contained 13-14 probes.

With this dual-color approach and the thresholding explained above, more loci can be detected simultaneously in the same cell, which allows calculating inter-probe distances with less bias and accessing different genomic architectural structures in a more reliable and easier manner.

### 5.3.2 Validation of miFISH pairwise distances with Hi-C

As has been shown in previous sections, a common way to validate new DNA FISH techniques is to reproduce the results of Hi-C. The Hi-C contact frequencies at 100 kb resolution from retinal pigment epithelium (RPE) cells were compared with 3D pairwise distances measured by miFISH. At first, we extracted contact frequencies from Hi-C data, which generated 120 pairwise contacts, and then compared this to the median of 120 pairwise distances from all miFISH alleles (Pearson's correlation coefficient, PCC: -0.75; Spearman's correlation coefficient, SCC: -0.77, **Figure 15a**). However, the FISH pairwise distances include long distances (often >1  $\mu\text{m}$ ) that are not captured by Hi-C.<sup>66</sup> Therefore, another approach was tested by setting a threshold of 1  $\mu\text{m}$  for miFISH pairwise distances to reflect the Hi-C contacts. This approach gave a weaker correlation of miFISH and Hi-C data (PCC: 0.40; SCC: 0.47, **Figure 15b**).



**Figure 15** – miFISH validation using Hi-C data. (a) Hi-C contact frequencies were correlated with the median pairwise distances obtained by miFISH. The linear regression is the dashed line and the dots are colored according to their lamina distance difference. PCC, Pearson's correlation coefficient. SCC, Spearman's correlation coefficient. (b) Hi-C contact frequencies were correlated with miFISH contact frequency, which is defined as dots located less than 1 μm apart. (c) For each miFISH probe, the normalized lamina distance was computed at the respective genomic coordinate. The color code is based on the A/B compartment assignment.

### 5.3.3 Lamina distances

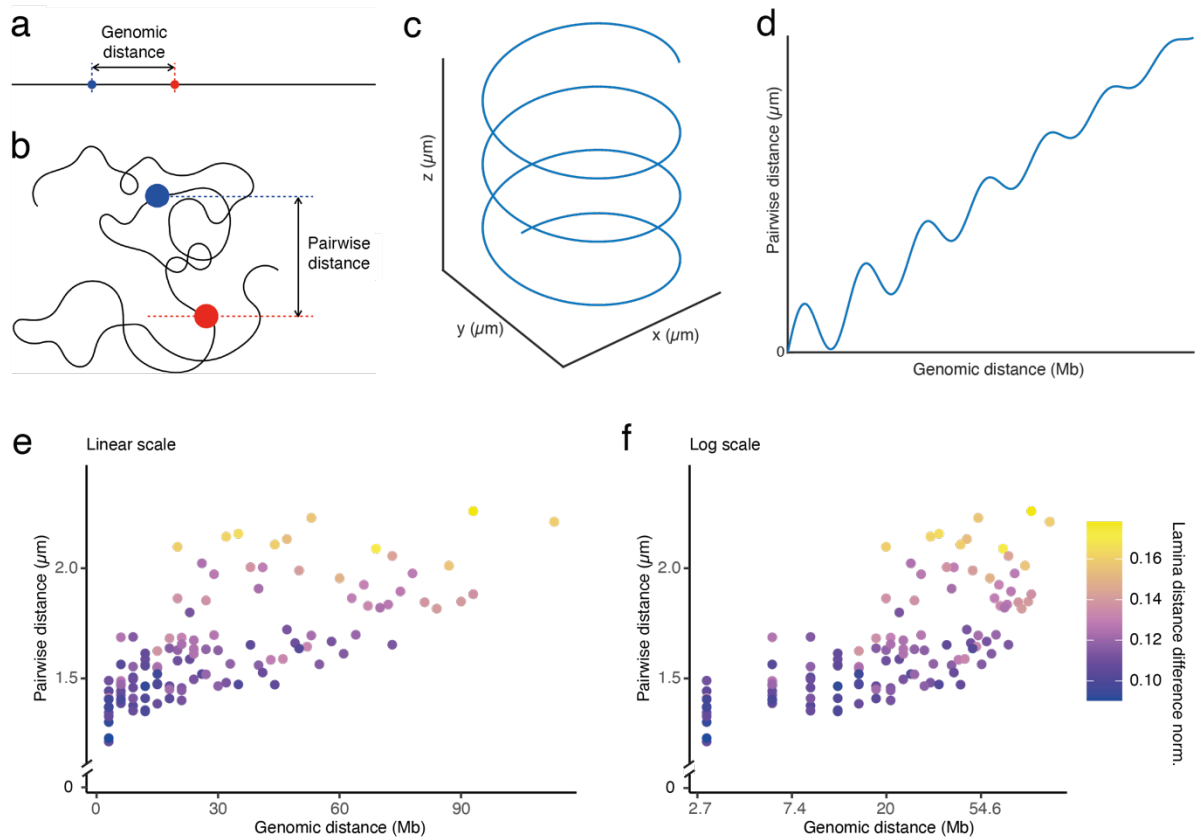
The nuclear lamina has an extensive influence on the chromatin architecture, as paper II demonstrated, and several techniques including DamID proved that nuclear proteins play an important role in creating lamina-associated domains (LADs). LADs consist of mostly genomic regions belonging to the B compartment and the regions close to the periphery are more densely packed.<sup>1</sup> Consequently, the miFISH probes placed in regions classified as B compartment are expected to be located closer to the lamina than the probed loci classified as A compartment. Therefore, we computed the lamina distance of all clusters detected for the 16 miFISH probes as described in the Materials and Methods section. The A and B compartment classification was extracted from Hi-C eigenvectors that were calculated through PCA

clustering.<sup>7</sup> The sign of the eigenvector is arbitrarily defined so in order to associate a region to an A/B compartment, the sign is correlated to ChIP-seq data. ChIP-seq looks at epigenetic markers that are known to locate mostly in inactive and repressive chromatin like the H3K9me3 and in active chromatin like H3K4me3 which respectively informs for compartments B and A. The correlation for this dataset was already performed in Darrow *et al.*<sup>126</sup> To discard ambiguous regions, a threshold was defined at an absolute eigenvector value of 0.01 for a region to be called A or B compartment. As expected, the lamina distances for probed loci belonging to A compartments were higher than for loci belonging to B compartments, showing that the probes assignment using the miFISH pipeline associates the FISH signals to the genomic region correctly (**Figure 15c**).

#### 5.3.4 Pairwise distances and higher-order chromosomal structures

To uncover chromosomal structures, the pairwise distances were used as a proxy to infer possible repeated structural patterns (**Figure 16a,b**). Our initial hypothesis posed that chromosomes form a helical structure in interphase cells (**Figure 16c**) in the same way as was previously described for mitotic chromosomes.<sup>34</sup> We assumed that the helical turn could occur between 10 and 25 Mb, consequently the 12 miFISH probes were spaced every 3 Mb along a genomic region of 33 Mb. The remaining 4 miFISH probes were used as controls and placed evenly at 20 Mb away from the 3 Mb probes. All 16 miFISH probes combined generated 120 different pairwise distances and the pairwise distances were computed against the genomic coordinates. In case the helical structure is the typical genomic structure, the pairwise distance would display a sinusoidal pattern and would periodically increase along larger genomic distances (**Figure 16d**). However, the miFISH results showed a gradual increase without sinusoidal behavior (**Figure 16e**). Thus, we discarded the option of a helical structure and replaced it with a less defined structure that travels gradually away from the starting position.

By plotting the distances using the logarithmic scale, the linear curve can be decomposed into multiple different curves with a characteristic slope. Therefore, different chromatin aggregates are visible that would not have been detected using the linear scale. miFISH probes displayed two clear distinct curve slopes with the inflection point at 20 Mb (**Figure 16f**). The reason for chromatin to change behavior at 20 Mb is not clear and not discussed in this paper since this result should ideally be confirmed in other chromosomes first.

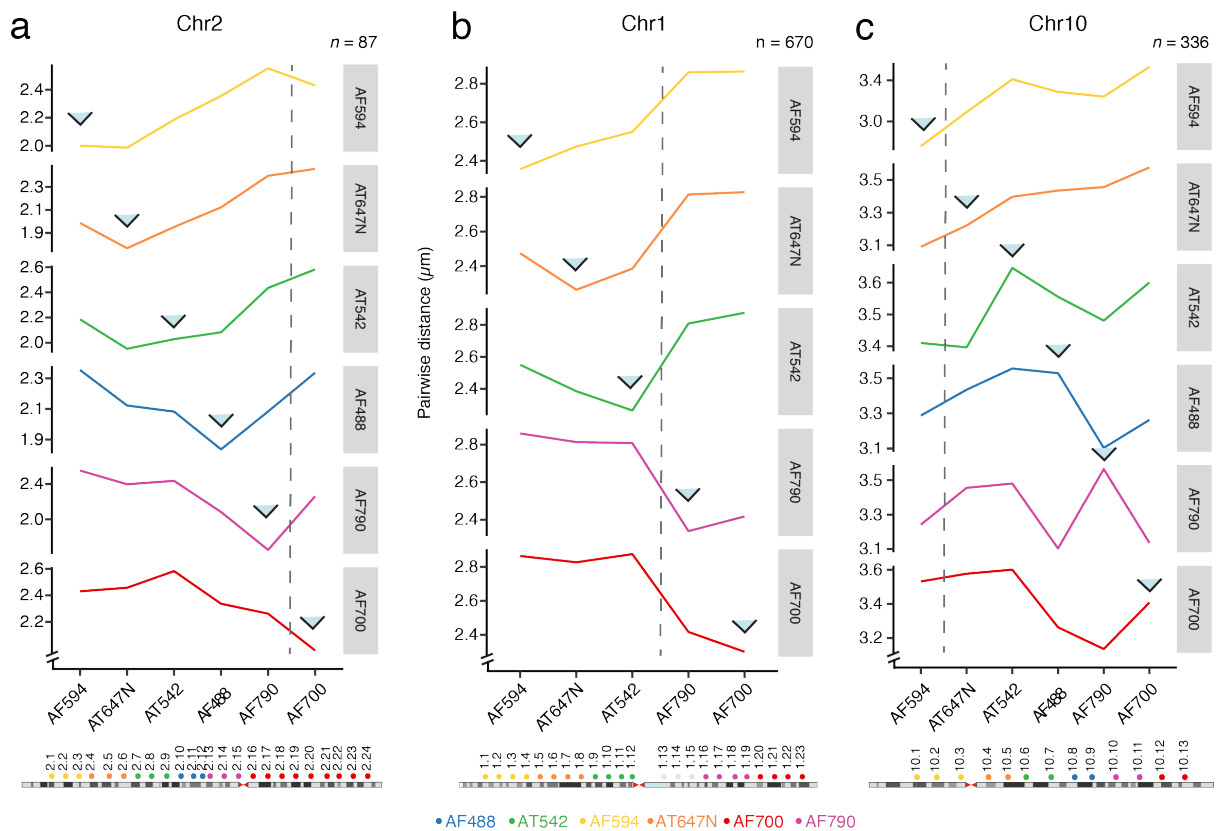


**Figure 16** – Higher-order chromosome structure model. (a) Representation of DNA as a black line and the two points at different DNA locations simulate the genomic distance in 1D. (b) Representation of a genomic structure in 2D space, where the pairwise distance between the two points does not resemble the linear DNA distance. (a) and (b) were adapted from Sazer & Schiessel, *Traffic*, 2018. (c) A scheme of a helical structure in 3D space. (d) Expected correlation between the pairwise distance and genomic distance for a helical structure like shown in (c). (e) Correlation between genomic distance and median pairwise 3D distances for all possible combinations of 16 miFISH probes, resulting in 120 pairwise comparisons. The color code ranges from the minimum to maximum normalized distance from the lamina for each dot. (f) Same as (e) but with a logarithmic scale, the pairwise distances after 20 Mb increase exponentially.

### 5.3.5 Global chromosome shapes

Individual chromosome topologies were previously described to be highly heterogeneous<sup>67,68</sup> and indeed, miFISH displays individual genomic structures that strongly point to high levels of variability. To validate the elongated configuration observed for chromosome 2, additional FISH experiments were performed on chr1, 2 and 10 with the classical iFISH probes.<sup>44</sup> For every chromosome, the probes were divided in 6 groups with a variable number of singly-labelled probes depending on the chromosome length and arm length. To access the pairwise distances of different genomic regions, the probes of the same color group were compared against the probes of another color group as well as against the probes of the same color group.

When looking at iFISH results from chr2, there is an increase of pairwise distances as was observed with miFISH probes (**Figure 17a**). The same was observed for chr1 (**Figure 17b**), but not for chr10 (**Figure 17c**), suggesting that chromosome elongation is a rather specific pattern of individual chromosomes.



**Figure 17** – iFISH pairwise distance test on chr1, 2 and 10. (a) The ideogram for chr2 is shown below the plots, where the probes and respective color are also mentioned. (b) Same as in (a) for chr1. (c) Same as in (a) for chr10. The grey dashed line shows centromere location.



#### **5.4 Paper IV: FRET-FISH probes chromatin density at individual genomic loci in single cells**

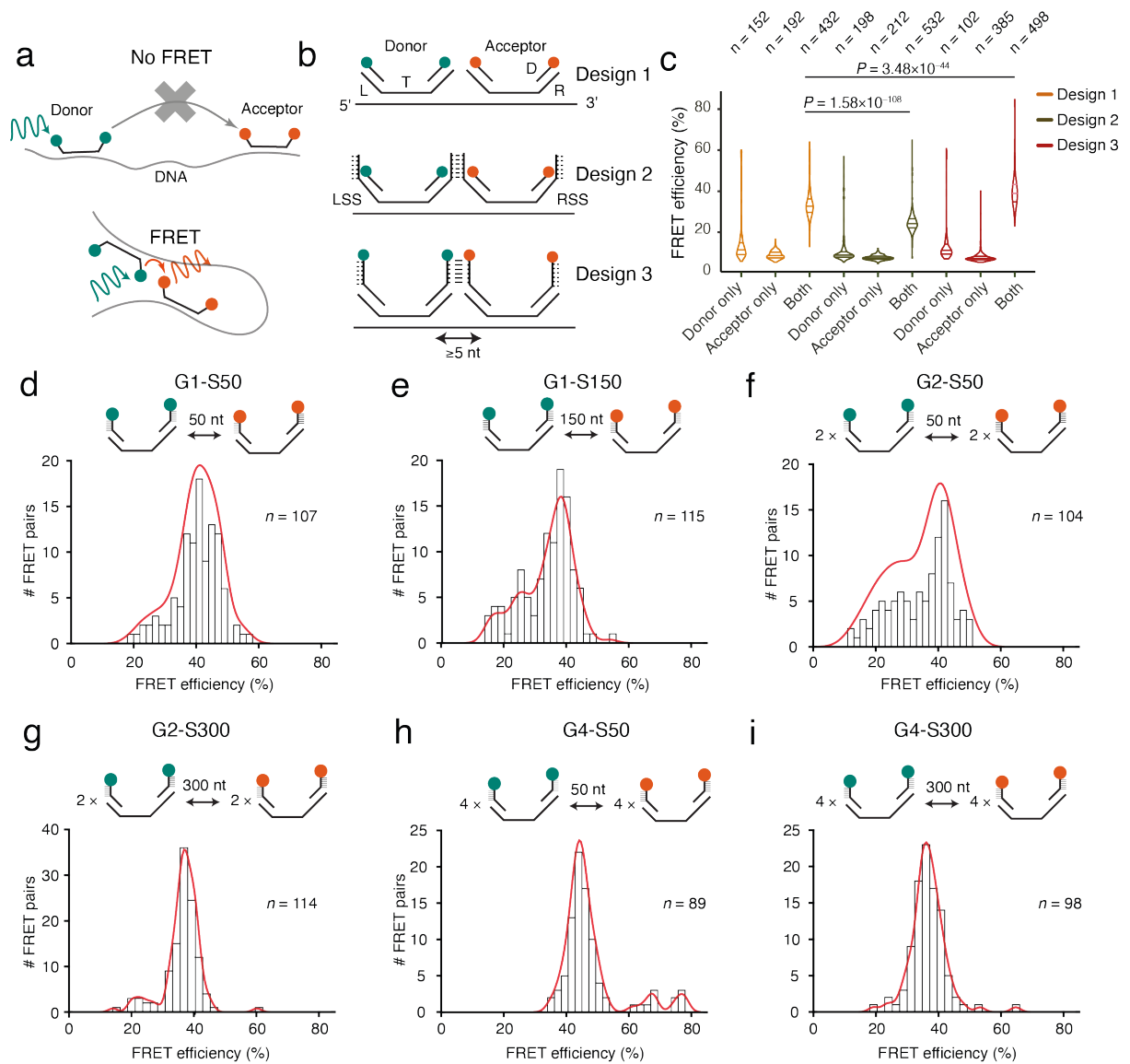
Throughout this thesis, the localization of a genomic region in the nucleus has been the major focus of the methods developed. In this paper, the local arrangement of chromatin is the main emphasis of the project since the local environment can affect the gene activity. For instance, a certain genomic region is organized in a relatively small volume where one can find high density of DNA. In turn, DNA have its accessibility obstructed to the transcription machinery and consequently, hinders certain genes to be expressed. Hence, we describe a novel method for assessing chromatin condensation named FRET-FISH, which combines two well-known microscopy techniques: fluorescence resonance energy transference (FRET) and DNA FISH. Together, sensitive measurements of spatial distances are made possible within a targeted genomic region. FRET-FISH can reveal properties of genome organization that thus far could only be inferred indirectly through sequencing methods involving DNA nucleases (as in DNase-seq)<sup>127</sup>, restriction enzymes (as in MNase-seq)<sup>128</sup> or transposases (as in ATAC-seq)<sup>72</sup>. Those methods have been widely used to infer chromatin accessibility by reasoning that digested DNA regions were accessible to the enzymes used due to a more open chromatin structure. Other sequencing methods, such as Hi-C<sup>129</sup>, measure the frequency of pairwise contacts genome-wide and have suggested that highly interactive regions may portray condensed chromatin regions. All the mentioned methods have been used as a proxy for chromatin compaction due to the lack of a well-established and widely accepted method for direct assessment of chromatin compaction in the scientific community.

To setup FRET-FISH, the FISH probe design was first refined to enable distinguishing different compaction states without compromising FRET signal detection. The FRET-FISH probe design pipeline enforces rigid spacing between the oligos that target specific genomic sequences. Moreover, the labelled oligos that carry the FRET dyes were designed to have stabilizing sequences. The FRET-FISH probes used here target different X chromosome genes and the results are in agreement with ATAC-seq (PCC: -0.80), demonstrating a relation between accessibility and DNA density. Superior correlation was observed with Hi-C data (PCC: -0.94), which proves that Hi-C is a better proxy for compaction than ATAC-seq.

At last, the method showed DNA density differences when (1) applying drug treatments that affect chromatin density, (2) prolonging the cell culture, (3) detecting different cell cycle phases and (4) dividing the nucleus into radial layers. This proves the method's high sensitivity to measure chromatin condensation.

#### 5.4.1 FRET-FISH probe design

FRET-FISH takes the iFISH methodology one step further by acquiring additional layers of information, such as the compaction of the chromatin. For such enterprise, FRET technology is combined with FISH oligos that target a specific locus. If the two oligos carrying complementary fluorescent molecules are positioned at a close distance (**Figure 18a**), fluorescence is transferred from the molecule with the highest excitation state to the molecule with the lowest excitation as described in the Jablonski diagram. The distance over which this energy transference occurs is defined as the Förster distance,  $R_0$ , and this value is specific for each combination of fluorescent molecules, due to the variation in spectral overlap between the donor emission spectrum and the acceptor excitation spectrum. One of the most extensively used FRET pairs is Cy3 and Cy5, thanks to its reasonable brightness and low price in comparison to other dyes. Taking this fluorescent pair, the Förster distance  $R_0$  is at 5.2 nm, where it displays 50% FRET efficiency. Besides the influence of molecular distance, there are many other factors that are known to affect FRET efficiency and the dipole orientation of the fluorescence molecule is one of those parameters since the electromagnetic field of the fluorophore mediates its interaction with other fluorophores. If the dipole orientation matches the FRET transfer the FRET efficiency should not be affected, however, if the dipoles are in opposite directions, the FRET transfer might not occur. Following this principle, we first tested three different FRET-FISH probe designs to identify which design yields the highest FRET efficiency. The test probe targeted the *MYC* gene locus in human haploid cells (HAP1) and encompassed 20 kb (**Figure 18b**). The oligos used to generate the probes have a similar structure as the iFISH oligos, although the target sequence is 60 nt instead of 40 nt to bind to the complementary genomic DNA with higher affinity. The oligo also includes the two flaps on the left (L) and right (R) of the target sequence for PCR amplification and for the binding of the fluorescently labelled (D) oligo (**Figure 18b**, Design 1). The second and third designs introduce an extra sequence to stabilize the oligos and restrict the free movement. This extra sequence is only 6 nt long, in order to not compete with the binding of the fluorescently labelled oligo of 20 nt to the complementary flap of the oligo targeting the genome. Longer sequences are more stable since more nucleotides contribute for the connection. In the case of the second design, the L and R flaps are extended with a linker sequence of 6 nt (**Figure 18b**, Design 2). Design 3 introduces the linker sequence in the fluorescently labelled oligo next to the fluorescence molecule (**Figure 18b**, Design 3). For all designs, the minimum distance between the consecutive oligos is 5 nt and the probe comprises 134 donor-acceptor oligo pairs. The experiments were performed in a 9-well chamber so that every design would go through exactly the same experimental procedure to reduce experimental bias.



**Figure 18** – Design optimization of FRET-FISH probes. (a) Schematic representation of the oligos (black line) that compose a probe of two different dyes. Orange color represents the barcode sequence specific for the labelled oligo carrying the acceptor dye while green represents the donor. Cartoon representation of FRET-FISH probes that hybridize to DNA strands with different compaction level, low condensation (top) and high condensation (bottom). At high density, the excited donor will transfer the fluorescence energy to the acceptor that will emit it in turn. (b) Schematic representation of 3 different designs that were tested in this study. The target sequence contains 60 nt and has 20 nt at both ends for hybridization to the labelled oligomers. For design 2 and 3, the linker sequence is 6 nt. The green dye represents the donor and orange is the acceptor and the target oligos are spaced at least 5 nucleotides. (c) Violin plots of FRET efficiencies, the color code is associated with the design.  $n$ , number of loci.  $P$ , p-value (Wilcoxon test, two-tailed). (d-i) The 6 different designs of targeting oligos are shown on top of each panel where GX-SX represents the number of oligos of the same color alternating and the inter-distance between the oligos of different color. On the bottom of each panel, the histogram distribution of the 6 different designs is shown using MEF cells. The orange line represents

the fitting curve looking for extra peaks arising from different compaction states of chromatin.  $n$ , number of loci.

The FRET efficiency was calculated based on **Eq. 2** on Methods and Materials section which divides the signal intensity detected in the FRET channel by the corresponding signal in the donor channel (Cy3) plus the FRET channel. As expected, Design 3 has the highest FRET efficiency, 39%, followed by Design 1, 32%, and Design 2, 24% (**Figure 18c**). In Design 3, the stabilizing sequence of the oligo binds to the stabilizing sequence of the neighboring oligo, which keeps the dipoles in a more parallel orientation than in Design 1. Design 2 was surprisingly worse than the other two designs, possibly due to the interference of the primary oligo that might introduce a larger distance between the fluorescent molecules and therefore act as a quencher. Additionally, to calculate the cross-excitation and crosstalk of these probes, controls were accounted for in the 9-well chamber, where some wells had only donor primary oligos or acceptor primary oligos for each design tested. The influence of cross-excitation and crosstalk is considered low and more importantly, it is similar between the different designs which means that it does not affect the comparisons.

#### 5.4.2 Optimization of FRET-FISH probe design to study local DNA density

After the promising results, we further tailored the probe design to allow measuring local compaction. The previous probe design placed the probes 5 nt apart, which is too close in the linear genome to display differences in 3D spatial organization or compaction. By increasing the spacing between the oligos, the FRET sensor can identify over wider ranges of distances, whether the DNA strand is more or less stretched in space. To do so, we based the distance between the oligos on a previous publication where purified nucleosomes were shown to have DNA wrapped around histones with repetitive modules of 300 nt.<sup>130</sup> In this study, shorter and longer spacing (S) distances between consecutive oligos are considered (50, 150, and 300 nt). Moreover, as DNA FISH hybridization efficiency is unknown and possibly never achieves 100%, adding more consecutive oligos that are either donor or acceptor could maximize the hybridization chances. To circumvent a possible loss of oligos, the probe design includes groups (G) that can have donor and acceptor oligos alternating in groups of 1, 2 or 4. The different designs were named G1-S50, G1-S150, G2-S50, G2-S300, G4-S50, and G4-S300 (**Figure 18d-i**). We used a limit for the probe span of 100 kb, to avoid probes spanning too large genomic regions and the likelihood of reducing FRET detectability to nuclear background. The proof-of-concept system chosen for the chromatin density test is X chromosome inactivation (XCI) in female cells. Previous studies have concluded that the genes

in the inactive chromosome X (Xi) are mostly transcriptionally silenced and have generally lower accessibility than those in the active allele (Xa). The designed FRET-FISH probes target the *Ogt* gene, which has a low (~6%) frequency of transcription from the Xi in *mus musculus* based on single-cell RNA-seq.<sup>63</sup> The inactivated locus is expected to be organized more compactly than its active counterpart on the Xa.

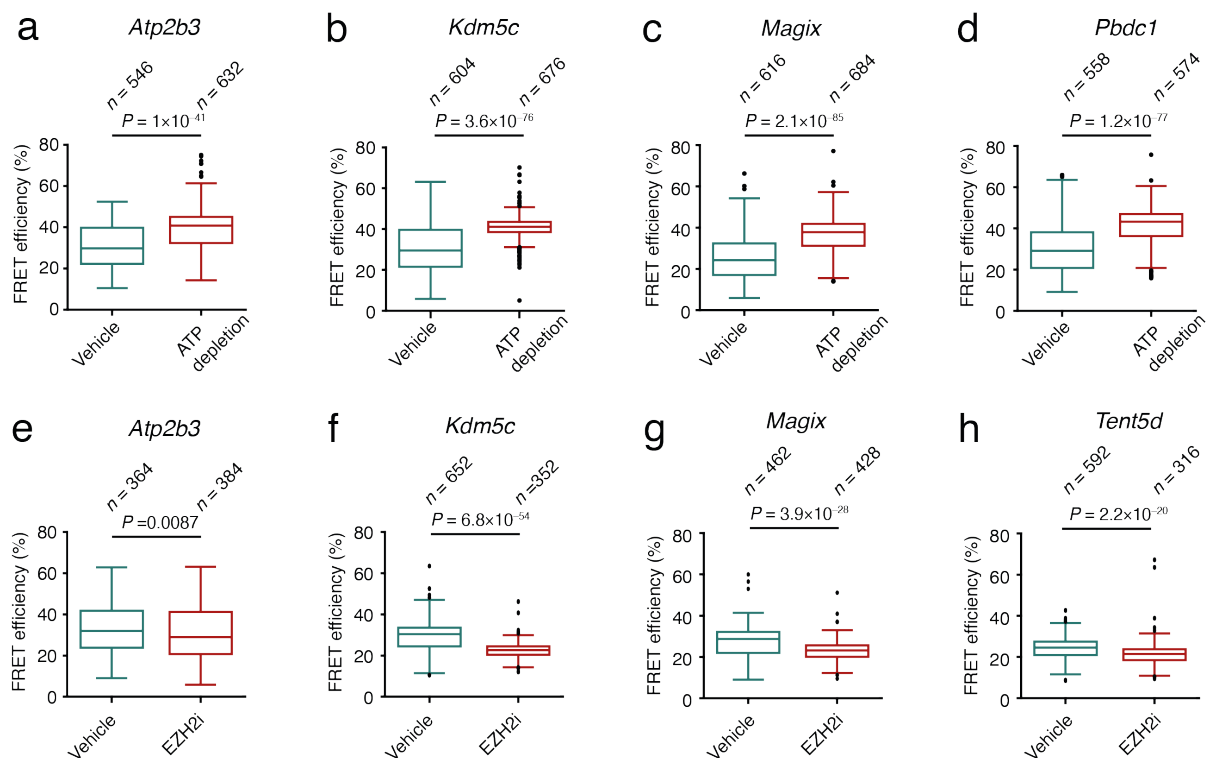
Two replicates were performed on two different cell lines, mouse embryonic fibroblast (MEF) and NIH Swiss murine fibroblast (NIH3T3). The results revealed that all designs had the major peak of FRET efficiency distribution at around 40% but the designs G1-S150 and G2-S50 have a second peak at lower FRET efficiencies (**Figure 18e,f**). The appearance of the second peak can be reasoned as a detection improvement of more varied chromatin states that are associated with different DNA densities. The intensity signal distributions of the acceptor in all designs are very similar, which further cements that the FRET efficiency differences come from DNA density differences and not from other technical or experimental variations. Nevertheless, if the signal average is compared, the G1-S150 probe shows a lower signal intensity due to the large region that the probes spans, which spreads the oligos over larger physical distances.

Once the optimization of the FRET-FISH probe design was successfully tested, different FRET dyes were contemplated to reduce crosstalk and bleed-through, improve fluorescence signal stability over several imaging steps and increase quantum yield. Therefore, in all subsequent experiments, AlexaFluor 488 (AF488) and AlexaFluor 594 (AF594) dyes were adopted with the G1-S150 probe design to capture different compactness states with reliable signal intensity.

#### 5.4.3 Validation of FRET-FISH to chromatin changes upon drug treatment

FRET-FISH sensitivity was tested with two different drug treatments that are responsible for chromatin condensation (10mM sodium azide and 50mM 2-deoxy-d-glucose) or decondensation (EZH2 inhibitor, Ezh2i). The first treatment depletes cells from ATP by blocking two energetic pathways with sodium azide and 2-deoxy-d-glucose. Sodium azide inhibits oxidative phosphorylation in mitochondria via inhibition of cytochrome oxidase, the final enzyme in the mitochondrial electron transport chain, thereby resulting in a rapid depletion of intracellular ATP. 2-Deoxy-d-glucose is a glucose molecule whose 2-hydroxyl group is replaced with a hydrogen. This molecule competes with glucose products to yield glucose-6-phosphate and consequently inhibits the glycolysis pathway. Subsequently, ATP depletion causes an increase of the number of free polyamines and divalent cations, which are no longer bound to the most abundant negatively charged molecule ATP. Instead, the

polyamines and divalent cations bind to DNA by a charge neutralization mechanism and contribute to the chromatin compaction.<sup>131</sup> Without ATP, a cell halts many other ATP-dependent processes such as transcription.<sup>132</sup> We confirm the treatment efficacy by staining nascent RNA with the Click-iT™ RNA kit, as transcription is an ATP-dependent process. Indeed, the treated cells have a significantly weaker staining of nascent RNA than control cells ( $P < 2.22\text{e-}16$ ). The FRET efficiency for four different genes (*Atp2b3*, *Kdm5c*, *Magix* and *Pbdc1*) was measured and the ATP depleted cells show a clear increase of FRET values, representing an increase in chromatin compaction in the treated cells in comparison to the controls (**Figure 19a-d**).



**Figure 19** – ATP depletion and EZH2i drug treatments on 4 different loci targeted with FRET-FISH probes. (a-d) Boxplot plots for control (vehicle) samples in comparison to the ATP-depleted cells for each gene. (e-h) Boxplot plots for control (vehicle) samples in comparison to the EZH2 inhibition cells for each gene. *n*, number of FRET pairs for control cells in first and treated cells in second. *P*, p-value (Wilcoxon test, two-tailed).

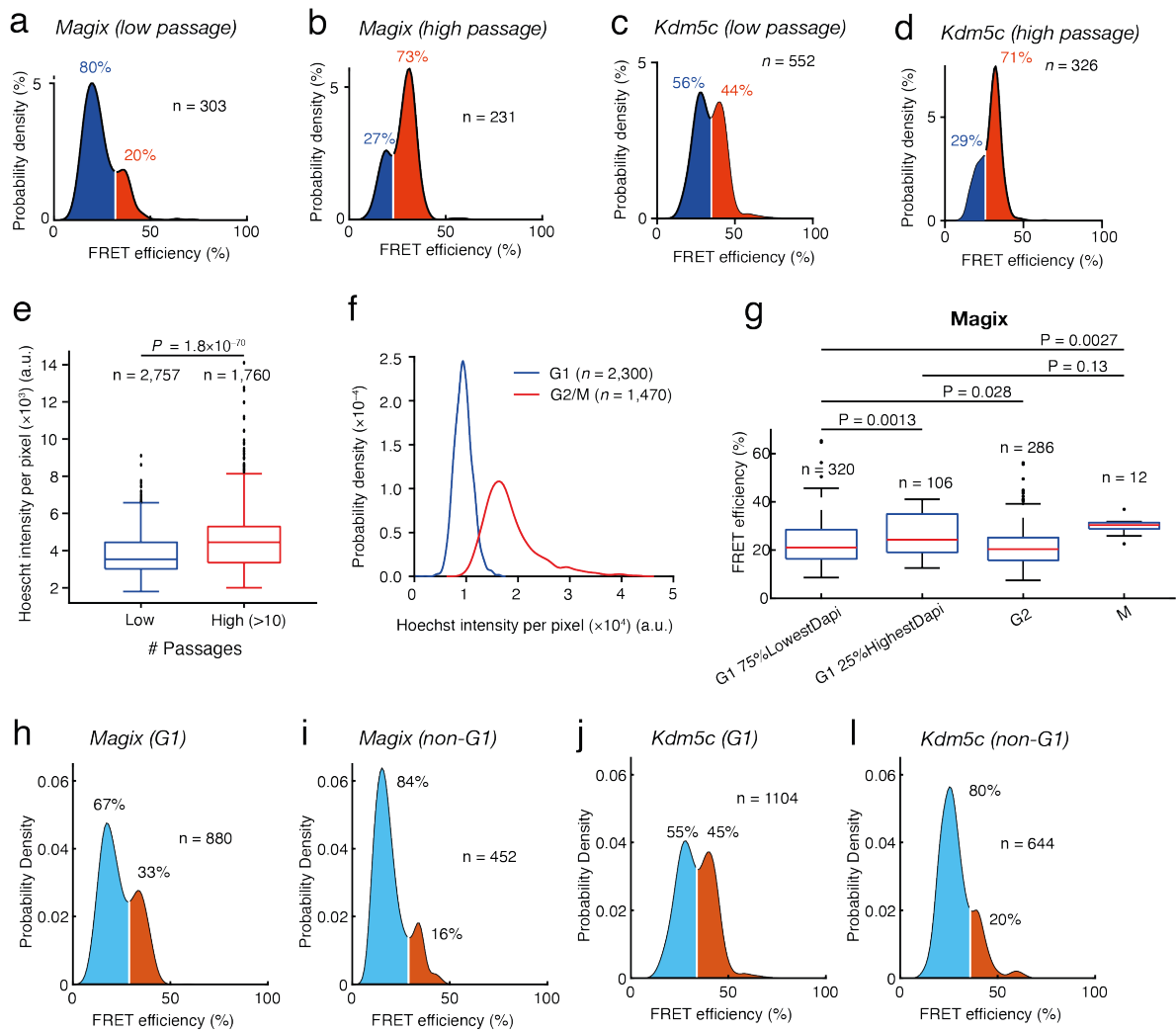
In contrast, in the decondensation treatment the enhancer of zeste homolog 2 inhibition (Ezh2i) obstructs the enhancer of zeste homolog 2 (EZH2) which is the catalytic subunit of the Polycomb repressive complex 2 (PRC2). PRC2 is the epigenetic regulator that induces histone H3 lysine 27 methylation (H3K27me3) and silences specific gene transcription. Ezh2 inhibition decreases Xi compaction since Xi is rich in Polycomb and decorated with H3K27me3 and, consequently, this treatment leads to the unfolding of the Xi.<sup>133</sup> We confirmed

Ezh2 inhibition with immunofluorescence staining of H3K27me3 antibodies, showing a profound depletion of this histone mark. As expected, the FRET efficiency decreases in the Ezh2i treated cells (**Figure 19e-h**).

#### 5.4.4 Prolonged cell culturing leads to chromatin condensation

Long maintenance of cells in culture is known to affect their epigenetic traits and subsequently threatens experiment reproducibility. Prolonged cell culturing also increases the levels of DNA methylation<sup>134</sup> and DNA damage<sup>135</sup>. Moreover, linker histone H1 is found at sites affected by DNA damage and in consequence, this histone reduces accessibility to other enzymes such as repair proteins.<sup>136,137</sup> Another study has shown that histone H1 influences genome organization through compaction.<sup>17</sup> Nevertheless, a direct correlation between prolonged cell culture and compaction has not been shown in the literature.

FRET-FISH displays typically a clear bimodal distribution of FRET efficiency in MEFs, as can be observed for the *Kdm5c* and *Magix* loci. FRET efficiency in higher passage cells is shifted to the mode with higher FRET efficiencies, which supports the hypothesis that prolonged culturing leads to chromatin condensation (**Figure 20a-d**). To further validate our method, we employed Hoechst fluorescence intensity as a proxy for DNA density, since it stains minor grooves of double-stranded DNA. Chromatin was found to be significantly more compact in cells with more than 10 culturing passages than cells that underwent less than 5 passages (**Figure 20e**). Hence, prolonged cell culture affects chromatin density, possibly by the mentioned DNA damage pathway.



**Figure 20** – Cell culture passages and cell cycle influence on chromatin compaction. (a) Representation of the probability density of FRET efficiency values for the *Magix* locus at low passage in MEF cells. The two peaks of the bimodal distribution are divided by the inflection curve between the peaks so that the mode with lower FRET efficiencies is blue and the mode with higher FRET efficiencies is red.  $n$ , number of loci. (b-d) Same as in (a) but for *Magix* at high passage, *Kdm5c* at low passage and *Kdm5c* at high passage. (e) The nuclear Hoechst intensity is averaged per pixel and its distribution is represented here in cells that underwent only few passages in comparison to cells that underwent more than 10 passages.  $n$ , number of nuclei. (f) Representation of the probability density of Hoechst averaged intensity per pixel for the *Magix* and *Kdm5c* loci for cell cycle phases G1 (blue line) and G2/M (red line).  $n$ , number of nuclei. (g) The G1 cells were divided into two groups according to the Hoechst intensity with the first including the cells that have the 75% lowest intensities and the second group with the remaining cells that have the 25% highest intensity. The G2 and M cell cycles are represented with their corresponding FRET efficiency distribution.  $n$ , number of loci (h-l) Same function of probability density as in (a) applied to G1 cells and for cell cycle phases after G1. (h) *Magix* at low passage, (i) *Magix* at high passage, (j) *Kdm5c* at low passage and (l) *Kdm5c* at high passage.  $n$ , number of loci.  $P$ , p-value (Wilcoxon test, two-tailed).



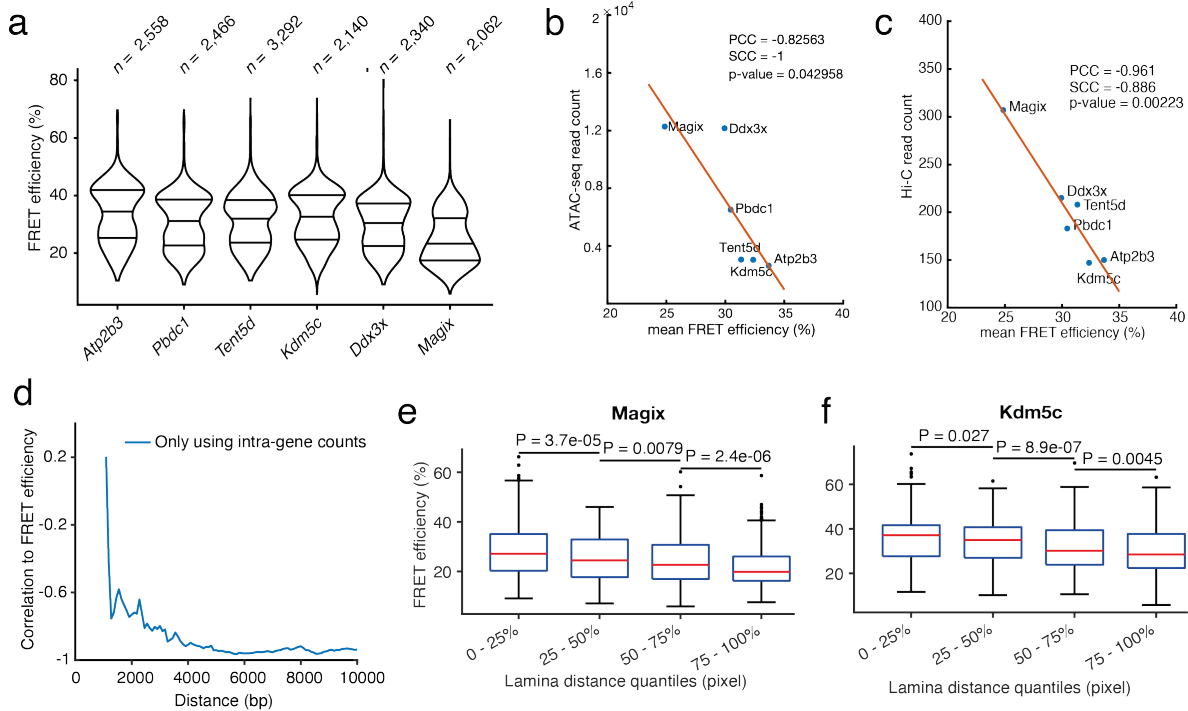
#### 5.4.5 Chromatin density changes during the cell cycle

Throughout the cell cycle, chromatin changes its density tremendously<sup>138</sup>, in particular when cells enter mitosis there is a stark compaction of the chromatin.<sup>139,140</sup> Here, cells in the different cell cycle phases were separated by quantifying the sum Hoechst signal in each segmented nucleus. G2 cells are particularly large cells due to the generation of duplicated genetic material and to the synthesis of proteins before entering mitosis, thus contributing extensively to the high Hoechst intensity. Therefore, the resulting histogram displays two peaks corresponding to G1 cells and the other to G2/M cells. As performed for the prolonged culturing, the local density was assessed with Hoechst on a global scale and by FRET-FISH as gene-specific. In contrary to summing whole-nucleus Hoechst intensity, the local density of chromatin is calculated by averaging the global intensity per pixel. Curiously, the Hoechst staining suggests that G2 cells have a higher density than G1 cells (**Figure 20f**) while FRET efficiencies display the opposite behavior (**Figure 20g**). This occurrence was also observed in ATAC-seq<sup>91</sup> where G2 cells had higher accessibility than G1 cells. One possible explanation is that the nuclear enlargement might help to decompact multiple regions of the nucleus to improve the accessibility for the replication machinery. On the other hand, the G1 cells were separated into two groups that include the high average intensity of Hoechst that corresponds to ~25% of cells and the low average intensity of Hoechst that corresponds to ~75%, and the high Hoechst intensity have higher FRET efficiency (**Figure 20g**). This is also in line with ATAC-seq that suggests that this high Hoechst intensity comes from G1 cells decondensing from mitosis, in which our data also shows high FRET efficiency (**Figure 20g**).

#### 5.4.6 Validation of the FRET efficiencies with ATAC-seq and Hi-C published data

As shown before with ATAC-seq, FRET-FISH is sensitive to chromatin changes and confirms observations made with orthogonal techniques. To further validate and infer FRET-FISH accuracy, several different genes were tested against the orthogonal sequencing methods ATAC-seq and Hi-C. The probe selection included inactivated genes (*Magix*, *Atp2b3*, *Tent5d*) since they undergo XCI and additional constitutive escapee genes (*Kdm5c*, *Pbdc1*, *Ddx3x*) were included, which are the result of incomplete inactivation of transcription in the Xi (**Figure 21a**). The corresponding ATAC-seq read count for the 6 genes was retrieved from GEO: GSE127926 and showed a strong correlation (Pearson's correlation coefficient, PCC: -0.80, **Figure 21b**). Likewise, Hi-C contact frequencies were extracted from GEO: GSE76479 and showed an even stronger correlation (PCC: -0.94, **Figure 21c**). Even though chromatin density is in high degree relatable with accessibility, FRET-FISH endorses that contact frequencies in

Hi-C are the closest resemblance of chromatin compaction. Furthermore, when focusing the analysis of Hi-C data on near-diagonal contacts, the correlation between Hi-C and FRET-FISH decreases, suggesting that FRET-FISH detects longer range contacts (**Figure 21d**).



**Figure 21** – Correlation of FRET efficiency with ATAC-seq and Hi-C for 6 genes. (a) Violin plots showing distribution of FRET efficiencies for each gene. The horizontal lines indicate the first, second (median) and third quartile (from bottom to top), and the whiskers extend from the minimum to the maximum value. *n*, number of loci (b-c) Correlation between (b) ATAC-seq and (c) Hi-C read counts for the same genomic region as the FRET-FISH probes, with the mean FRET efficiency for each gene that was shown in (a). p-value (Wilcoxon test, two-tailed). PCC, Pearson's correlation coefficient. SCC, Spearman's correlation coefficient. (d) The p-value correlation value is acquired at different Hi-C contact windows that are typically shown in Hi-C maps close to the diagonal. The curve shows the fitness of intra-gene contacts at defined ranges when compared to the FRET data. (e-f) Representation of the progression of FRET efficiency at radial layers of the nucleus located progressively at higher distances from lamina for (e) *Magix* and (f) *Kdm5c* from 3 independent replicates. *P*, p-value (Wilcoxon test, two-tailed).

#### 5.4.7 Radial localization in the nucleus influences chromatin condensation

Finally, the final result demonstrates that the radial organization of the genome in the nucleus influences the chromatin compaction. GPSeq method presented in paper II predicted that more accessible and actively transcribing genes are located in the center of the nucleus while the silenced or inactivated regions are positioned closer to the nuclear lamina. FRET-FISH was performed in 3 replicates for *Magix* and *Kdm5c* gene and the acquired signals were divided in

quantiles according to their radial position in relation to the nuclear lamina (0-25%, 25-50%, 50-75% and 75-100%). Hence, the FRET efficiencies coming from FRET-FISH signals captured within those lamina distance intervals prove that more compacted chromatin is located closer to the nuclear lamina and chromatin progressively decompacts until more central layers of the nucleus (**Figure 21e-f**) in line with GPSeq predictions.



## 6 DISCUSSION AND CONCLUSIONS

DNA FISH was first established in the '80s<sup>141</sup> and was responsible for the pioneering steps in the genome organization field. FISH, together with other fluorescence techniques, revealed links between genome organization and functionality, by tracking transcription and proteins as well. However, in the years after the DNA FISH technique was developed, its applicability rarely extended beyond clinical laboratories. The first DNA probes were based on bacterial artificial chromosomes (BAC) and the production of all oligos for a single probe was laborious and expensive. This changed with the MERFISH project, which already introduced a more efficient workflow to produce many probes simultaneously, an effort that the iFISH paper aimed to improve with more cost-effective probe production and higher gene coverage. Besides the cost reduction, the probe design implementation has become more within reach for every interested research group. The iFISH probes datasets are publicly available and the iFISH4U tool was implemented to assist with probe design for a specific region. These advances can bring DNA FISH at a closer level to high-throughput sequencing techniques in terms of genomic sequences targeted and provide valuable unprecedented insights into single-cell genome structures.

Since the iFISH workflow allows for high-scale probe production, a single chromosome can have several probes hybridizing in very precise locations by spanning small regions. At the moment, 6 chromosomes can be spotted individually in the same experiment by targeting 20 probes per chromosome and imaging each chromosome with one of the 6 colors available for DNA FISH. We have also used the iFISH data to identify chromosome territories and assess the intermingling level between the chromosomes, showing more loose territories in embryonic cells, accompanied by higher levels of intermingling.

To expand the number of loci imaged beyond the 6 colors available, miFISH creates combinations of 2 out of 6 colors to allow capturing 16 different loci in one single hybridization step. miFISH is based on the same simple workflow of iFISH but without the microfluidic system. In the future, implementation of miFISH in the MERFISH workflow would reduce the mechanical shifts and the consequent errors from the decoding step by imaging more loci in one single imaging round.

The structures imaged on chr2 highlight the heterogeneity observed in single-cell methods, nonetheless, the *q* arm of chr2 showed that the pairwise distances between the probes were growing along their genomic distances. This trend was observed in chr1 but was not repeated in chr10, which displays a more random organization. Chr10 is more centrally located and

smaller in comparison to chr1 and chr2, which could result in chr10 being more easily affected by entropic variables. The miFISH dataset and the iFISH spotting probes are freely available for other groups to analyze and the genomic structures can be employed for chromatin modelling purposes.

The vast production of iFISH probes proved useful for the validation of GPSeq, a novel sequencing-based technique to study chromatin organization. In itself, GPSeq brings new knowledge of the radial positioning of genomic sequences. When comparing GPSeq with the epigenetic markers that have been used as predictors of chromatin organization and transcriptional activity, they seem to support the pattern that active chromatin is positioned in the center of the nucleus while inactive chromatin is closer to the lamina. correlate well with the predictions of the radial position for several markers agree with GPSeq Therefore, GPSeq can also be used as a predictor for chromatin state when certain region is preferentially located at the surface or at the center of the nucleus. This hypothesis is also in accordance with DamID which proved that DNA close to the nuclear lamina interacts with proteins like lamin B that are responsible for silencing genes.<sup>142,143</sup> Even when comparing with the golden-standard method of DNA FISH, GPSeq was shown to correlate highly with DNA FISH data.

Furthermore, GPSeq results were complemented with Hi-C data so that both techniques provide information on genome organization, respectively of the radial positioning and the contacting chromatin regions in its neighborhood. Then, several 3D genome structures were generated leveraging the *chromflock* modeling tool. In the advent of single-cell sequencing tools, more realistic structures are expected to be achieved by modeling with both chromatin contacts and radial information.

An unaddressed point that can bias the radial prediction is the enzymatic accessibility to certain locations, in particular where large numbers of proteins interacts to create solid-like structures. This could be the major drawback of the technique, together with the inability to distinguish homologous chromosomes without SNPs to distinguish the paternal allele from the maternal. The accessibility question can be tested by performing DNA FISH in known areas with higher and lower accessibility scores from ATAC-seq. For example, as used in paper IV, the two X chromosome in female cells develops two distinct fates, with one being inactivated so accessibility is believed to be reduced. Therefore, GPSeq scores can be compared with DNA FISH probes average positions, while the inactive allele is simultaneously distinguished in all cells using RNA FISH targeting *Xist* RNA.

The GPSeq paper revealed, like previous work, that heterochromatin is localized at the outmost layers, which was then hypothesized to relate to protection of the more central euchromatin from external agents, borrowing the idea from the old “bodyguard” model.<sup>123</sup> The

model assumes that external factors such as UV light are deemed to cause DNA damage more in the peripheral chromatin, while the most transcribed genes in the center are shielded. Taking the example of irradiating light, the DSBs were shown to occur more in euchromatin than in heterochromatin, because H1 and H5 are found more on heterochromatin and shield DNA from damage, while euchromatin is less wrapped around histones and consequently more exposed to the oxygen radicals.<sup>144,145</sup> Then, the “bodyguard” model does not apply to DSB since it was also shown with GPSeq that occurs mostly in the center. Other types of mutation such as point mutations change DNA in single nucleotide by deletion, insertion or switch to another nucleotide which can lead to SNV. Those type of mutations occur mostly during DNA replication but point mutations are also a consequence of external agents as irradiation or chemicals coming from smoking for instance. It is not clear that those agents create directly more mutations in the outmost layers, what is shown in GPSeq is that they tend to accumulate in heterochromatin. The SNVs accumulation in the outmost layers can be associated to low correction efficacy of DNA damage when positioned in heterochromatic regions due to obstructed accessibility for proteins involved in repair machinery.<sup>146–152</sup> The internal regions are less compacted and more transcribed so also more easily noticed by the repairing proteins. Euchromatin is relatively more essential to the cell functioning so the repair mechanism should start faster and more efficiently at those regions than in gene inactivated regions present in the heterochromatin. Hence, another way to justify the presence of more SNVs at the peripheral layers could be by the inefficacy of the repair pathway in protein-shielded chromatin regions. In the “bodyguard” model, the heterochromatin is implied as a protective coat of the transcriptionally active center from external agents but the protective capacity of heterochromatin does not seem to sustain in any study. Instead, the accumulation of mutations in the external layers can be derived by the lower repair activity in heterochromatin.

The heterochromatin position in more peripheral layers was partially explained in the introduction and attributes the process of liquid phase separation the responsibility to separate different chromatin domains. For instance, heterochromatin is present preferentially at the lamina because of the affinity of heterochromatin proteins for Lamin proteins.<sup>142,153</sup> Moreover, the allocation of most active events into the nuclear center can be seen as a way of achieving a high concentration of complex enzymatic machinery in a smaller space. Otherwise, for transcription conducted more dispersedly throughout the nucleus, it likely takes longer time for the proteins to reach the genes, which would also require higher concentration of enzymes to span all the nuclear volume, which altogether would be an inefficient way of managing the limited energetic resources of a cell.

The final method that was presented in this thesis is FRET-FISH, a technique that emerged from iFISH. Both the introduction of FRET sensor ability and the DNA FISH probe design were improved to acquire optimal FRET detection. For instance, the pipeline that designs FRET-FISH probes is more rigid than previous iFISH pipeline because it enforces a constant distance between the oligos and removes genomic regions with low uniqueness. In consequence of larger inter-oligo distance, the probe span is increased (~100 kb), which reduces the precision of the probes. Nevertheless, there is room for probe span reduction because we hybridized 2.5 times more oligos than in iFISH. On the other side, the probe span can be expanded to wider regions so large-scale structures can also be studied such as topologically associating domains (TADs). The article also aimed to prove the sensitivity of FRET-FISH to discern different compaction states. Two drugs with well-known documentation were used for the sensitivity test but many other drugs can be tried in the future to understand how these drugs affect chromatin organization.

This new method provides a novel tool to study genome organization in single cells and it may in the future be combined with immunofluorescence and RNA FISH to complement the genomic information. FRET-FISH used a regular epifluorescence microscope, so many other groups can easily implement it in their laboratories. This article has begun to study some aspects of cell cycle variation and DNA damage, but more in-depth future studies can continue with targeting relevant genes for processes such as (1) condensation and decondensation in mitosis, (2) embryonic cell differentiation and the consequent (3) formation of euchromatin and heterochromatin regions in the nuclei. FRET-FISH can also be used as a tool to optimize certain parameters of the FISH protocol such as hybridization efficiency of the oligos in the targeted regions.

Besides, this technique may have more versatile implementations such as hybridizing donor and acceptor targeted oligos to distinct loci to measure the relative physical distances between those regions. This new conceptualization can be applied to enhancer-promoter loops, for example at the MYC locus, to gain further insights into the mechanisms of loop formation. FRET-FISH can identify when the loop occurs in single cells and the imaging can be complemented with cohesin/CTCF staining and RNA FISH to assess native transcripts. The same approach can be also applied to study DNA-RNA and RNA-RNA interactions, in particular when those interactions are responsible for liquid-liquid phase separation. The application of FRET-FISH to DNA-protein and RNA-protein interactions could potentially partially replace proximity-ligation assays, so that no ligation step would be required.

Finally, FRET-FISH showed that chromatin compactness is inversely correlated with accessibility. Nevertheless, accessibility might be dependent on other factors that are



independent of chromatin compaction such as DNA-binding proteins. The *Ddx3x* and *Magix* loci are two interesting examples out of 6 genes that do not correlate entirely with ATAC-seq published data. Staining with DNA-binding proteins can be performed simultaneously with FRET-FISH to understand if those proteins can be a factor that affects chromatin accessibility. Some of these suggestions can be tested in simplified settings like in vitro experiments, however, performing these tests in fixed cells attributes higher biological relevance to the results.



## 7 ACKNOWLEDGEMENTS

Just like the epics, the monumental achievements were thanks to many other people that provided assistance or guidance during a turbulent journey. Here I have the opportunity to show my gratitude to those that were like the guiding stars in my explorations.

In first place, I owe to Dr. Magda Bienko the opportunity to pursue a PhD in Stockholm at Karolinska Institutet. The whole experience was life changing and opened doors for career opportunities that otherwise were unthinkable. Nonetheless, on top of everything, I have a deep sense of gratitude for allowing me to dedicate so much of my PhD time to a self-proposed project, a relatively risky project that could have been an immense waste of resources. FRET-FISH was a very challenging project but, if I have grown as a researcher during the PhD studies, it was thanks to this project. I got to know how much it takes to successfully develop a research project and I appreciate that Magda granted this opportunity and supervision.

I would like to appreciate my dissertation committee, Dr. Anna Klemm, Prof. Marcus Wilhelmsson and Dr. Antoine Coulon and my opponent Prof. Beat Fierz for being present in my thesis defense and for contributing to an interesting discussion of my research outcomes. Also, I would like to thank Dr. Claudia Kutter for helping in hosting the thesis defense as a chairperson.

Many thanks to my co-supervisors, Prof. Camilla Björkegren and Dr. Sebastian Deindl. At different PhD stages, I have got the chance to get feedback and discuss project issues with someone of vast scientific experience and that kindly shared a more external perspective.

I am also thankful to Dr. Nicola Crosetto for all the discussions, scientific and non-scientific, for reading my manuscripts and giving valuable feedback. It was a pleasure to have met you and for sharing your views on the science setting.

A very big thanks to all members of Bienko-Crosetto lab for offering so many different perspectives to look at scientific questions, help in troubleshooting, develop new analytic methods and gain resilience. I would like to thank some members in particular for collaborating so closely in my research projects.

Firstly, Maud Schweitzer is the most transversal member of the lab. Maud participated in all the 4 papers of this thesis and she knows all the colors that my PhD journey got. Maud was like a power plant in the lab, she provided constant energy supply, when Maud wasn't there, it was like the lights were turned off.

Xinge and Masahiro had long left our group but they were the responsible to introduce me FISH and both had an extraordinary caring personality, they constantly gave their best and were always available to help, no matter how busy they were.

Katta and Xiaoze are very supportive lab managers that have assisted in different tasks and by sharing their views on lab work and on project-related topics, gained during their long experience in academia.

Erik for supporting in the analysis of the constitutive papers, in particular when deadlines were tight and messages in slack were just flooding.

Federico, Gabriele, Eleni, Kata, Tomek, Quim and Sólrún were all part of a huge team, that I also belonged, to achieve great scientific endeavors where we trailed a long path and learn to put aside conflicts or self-affirmation issues so that we can smoothly overcome obstacles and altogether enjoy the discoveries.

Su, Reza, Michi, Emma, Britta, Roberto, Luuk, Lei, Ning, Silvano, Carla, Hin, Eleonora and Wenjing are all amazing members or former members of Bienko-Crosetto group that enabled a scientific active environment with very interesting conversations and by sparkling curiosity in the most diverse topics which I deeply appreciate for expanding my horizon.

I could not end the acknowledgements without recognizing the invisible supporters that even though their names do not appear in any thesis section or papers, they were crucial for me to start the PhD. In first place comes my family, in particular my parents, for all the sacrifices they endured so that I could study at the university. My aunt Guiomar and uncle Jancinto gave me a hand when the 2011 crisis seemed to threaten my dreams. In Lisbon, even though not part of my family, Delfino's family were complete strangers to me until they allowed me to stay at their place without any charge. Also, I would like to thank everyone that supported me in the Netherlands. I hope that you feel this PhD is all due to you, otherwise I wouldn't been able to pursue it.

To my friends for their advices, for their laughs, for the hiking and for always being there no matter what.

And, I would like to finish my acknowledgements by thanking a very special person that gave an immeasurable support during the whole PhD. It was suggested of creating a certificate for him that states "PhD day care" and it should be recognized by some Public Health Agency. I think that you have a promising career for your part-time. After all, there are really no words for you, because you know all about me. *Faleminderit!*

## 8 REFERENCES

1. Steensel, B. van & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).
2. Gonzalez-sandoval, A. & Gasser, S. M. On TADs and LADs: Spatial Control Over Gene Expression. *Trends Genet.* **32**, 485–495 (2016).
3. Bickmore, W. A. The Spatial Organization of the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
4. Leblanc, B. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472 (2012).
5. Solovei, I. *et al.* Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* **137**, (2009).
6. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).
7. Lieberman-aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–294 (2009).
8. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Science Advances* **5**, (2019).
9. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, (2019).
10. Nasmyth, K. Disseminating the Genome: Joining , Resolving, and Separating Sister Chromatids During Mitosis and Meiosis. *Annu. Rev. Genomics* **35**, 673–745 (2001).
11. Davidson, I. F. *et al.* Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO* **35**, 2671–2685 (2016).
12. Parelho, V. *et al.* Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132**, 422–433 (2008).
13. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
14. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci.* **105**, 8309–8314 (2008).
15. Feric, M. & Misteli, T. Phase separation in genome organization across evolution. *Trends in Cell Biology* **31**, (2021).
16. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, (2019).
17. Willcockson, M. A. *et al.* H1 histones control the epigenetic landscape by local chromatin compaction. *Nature* **589**, (2021).
18. Larson, A. G. *et al.* Liquid droplet formation by HP1 $\alpha$  suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
19. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature*

- 547**, 241–245 (2017).
20. Sanulli, S. *et al.* HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, (2019).
  21. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
  22. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
  23. Maharana, S. *et al.* Chromosome intermingling — the physical basis of chromosome organization in differentiated cells. *Nucleic Acids Res.* **44**, 5148–5160 (2016).
  24. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2**, 437–447 (2008).
  25. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb Perspect Biol.* **2**, 1–23 (2010).
  26. Ay, F. *et al.* Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* **24**, 974–988 (2014).
  27. Branco, M. R., Branco, T., Ramirez, F. & Pombo, A. Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH. *Chromosom. Res.* **16**, 413–426 (2008).
  28. Hoencamp, C. *et al.* 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**, (2021).
  29. Liao, Y., Zhang, X., Chakraborty, M. & Emerson, J. J. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Res.* **31**, (2021).
  30. Hirano, T. & Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* **79**, (1994).
  31. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, (2019).
  32. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, (2018).
  33. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
  34. Naumova, N., Nuebler, J., Kanemaki, M. T. & Xie, L. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
  35. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
  36. Nir, G. *et al.* Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, (2018).
  37. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially

- resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, 1360–1363 (2015).
38. Id, C. G. E. *et al.* Walking along chromosomes with super- resolution imaging , contact maps , and integrative modeling. *PLOS Genet.* **14**, e1007872 (2018).
  39. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
  40. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
  41. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, (2021).
  42. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
  43. Gelali, E. *et al.* An Application-Directed, Versatile DNA FISH Platform for Research and Diagnostics. in *CpG Islands: Methods and Protocols* (eds. Vavouri, T. & Peinado, M. A.) 303–333 (Springer New York, 2018). doi:10.1007/978-1-4939-7768-0\_17
  44. Gelali, E. *et al.* iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* **10**, (2019).
  45. Shammas, M. A. Telomeres, lifestyle, cancer, and aging. *Curr. Opin. Clin. Nutr. Metab. Care* **14**, (2011).
  46. Verdaasdonk, J. S. & Bloom, K. Centromeres: Unique chromatin structures that drive chromosome segregation. *Nature Reviews Molecular Cell Biology* **12**, (2011).
  47. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: Advances through genomics. *Nature Reviews Genetics* **10**, (2009).
  48. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* **21**, (2011).
  49. Kumar, A. & Kono, H. Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. *Biophysical Reviews* **12**, (2020).
  50. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
  51. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three- Spread Across the X Chromosome. *Science* **341**, 1–9 (2013).
  52. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
  53. Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249 (2018).
  54. Fang, H., Disteche, C. M., Berletch, J. B. & Morey, C. X Inactivation and Escape: Epigenetic and Structural Features. *Front. Cell Dev. Biol.* **7**, 1–12 (2019).
  55. Shvetsova, E. *et al.* Skewed X-inactivation is common in the general female

- population. *Eur. J. Hum. Genet.* **27**, (2019).
56. Galupa, R. & Heard, E. X-chromosome inactivation: A crossroads between chromosome architecture and gene regulation. *Annual Review of Genetics* **52**, (2018).
  57. Cerase, A. *et al.* Phase separation drives X-chromosome inactivation: a hypothesis. *Nature Structural and Molecular Biology* **26**, (2019).
  58. Strickfaden, H. *et al.* Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, (2020).
  59. Tatavosian, R. *et al.* Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J. Biol. Chem.* **294**, (2019).
  60. Plys, A. J. *et al.* Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, (2019).
  61. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, (2014).
  62. Berletch, J. B. *et al.* Escape from X Inactivation Varies in Mouse Tissues. *PLOS Genet.* 1–26 (2015). doi:10.1371/journal.pgen.1005079
  63. Reinius, B. *et al.* Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* **48**, (2016).
  64. Berletch, J. B., Yang, F., Xu, J., Carrel, L. & Disteche, C. M. Genes that escape from X inactivation. *Hum. Genet.* **130**, (2011).
  65. R. Barth, K. B. and H. A. S. Coupling chromatin structure and dynamics by live super-resolution imaging. *bioRxiv* 1–57 (2019).
  66. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 1–9 (2016).
  67. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, (2019).
  68. Altschuler, S. J. & Wu, L. F. Cellular Heterogeneity: Do Differences Make a Difference? *Cell* **141**, (2010).
  69. Cattoni, D. I. *et al.* Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* **8**, (2017).
  70. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, (2007).
  71. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, (2010).
  72. 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. **10**, (2013).
  73. Morozova, O. & Marra, M. A. Applications of next-generation sequencing



- technologies in functional genomics. *Genomics* **92**, (2008).
74. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nature Reviews Genetics* **12**, (2011).
  75. Zaidi, S. K. *et al.* Bookmarking the Genome: Maintenance of Epigenetic Information. *J. Biol. Chem.* **286**, 18355–18361 (2012).
  76. Xu, J. & Liu, Y. Probing Chromatin Compaction and Its Epigenetic States in situ With Single-Molecule Localization-Based Super-Resolution Microscopy. *Frontiers in Cell and Developmental Biology* **9**, (2021).
  77. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, (2005).
  78. 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**, (1988).
  79. Ali, M. M., Li, F. Z. Z. & Kaixiang Zhang, Dong-Ku Kang, a James A. Ankrum, e X. C. L. and W. Z. Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.* **43**, 3324–3341 (2014).
  80. Clausson, C. *et al.* Compaction of rolling circle amplification products increases signal integrity and signal-to-noise ratio. *Sci. Rep.* **5**, 1–10 (2015).
  81. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, (2008).
  82. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, (2000).
  83. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, (2006).
  84. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, (2006).
  85. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, (2015).
  86. Han, Z. & Porter, A. E. In situ Electron Microscopy of Complex Biological and Nanoscale Systems: Challenges and Opportunities. *Front. Nanotechnol.* **2**, (2020).
  87. Bazett-Jones, D. P., Li, R., Fussner, E., Nisman, R. & Dehghani, H. Elucidating chromatin and nuclear domain architecture with electron spectroscopic imaging. *Chromosome Research* **16**, (2008).
  88. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, (2013).
  89. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
  90. Zuk, P. A. *et al.* Multilineage cells from human adipose tissue: Implications for cell-based therapies. in *Tissue Engineering* **7**, (2001).

91. Chen, X. *et al.* ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, (2016).
92. Xie, L. *et al.* 3D ATAC-PALM: super-resolution imaging of the accessible genome. *Nat. Methods* **17**, (2020).
93. Stephens, A. D., Banigan, E. J., Adam, S. A., Goldman, R. D. & Marko, J. F. Chromatin and lamin A determine two different mechanical response regimes of the cell nucleus. *Mol. Biol. Cell* **28**, (2017).
94. Nava, M. M. *et al.* Heterochromatin-Driven Nuclear Softening Protects the Genome against Mechanical Stress-Induced Damage. *Cell* **181**, (2020).
95. Haase, K. & Pelling, A. E. Investigating cell mechanics with atomic force microscopy. *Journal of the Royal Society Interface* **12**, (2015).
96. Zidovska, A. Chromatin: Liquid or Solid? *Cell* **183**, (2020).
97. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR / Cas System. *Cell* **155**, 1479–1491 (2013).
98. Chen, B., Zou, W., Xu, H., Liang, Y. & Huang, B. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nat. Commun.* **9**, (2018).
99. Qin, P. *et al.* Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.* **8**, (2017).
100. Wang, Y. *et al.* Genome oligopaint via local denaturation fluorescence in situ hybridization. *Mol. Cell* **81**, (2021).
101. Ooga, M. & Wakayama, T. FRAP analysis of chromatin looseness in mouse zygotes that allows full-Term development. *PLoS One* **12**, (2017).
102. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, (2013).
103. Nozaki, T. *et al.* Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol. Cell* **67**, (2017).
104. Estandarte, A. K., Botchway, S., Lynch, C., Yusuf, M. & Robinson, I. The use of DAPI fluorescence lifetime imaging for investigating chromatin condensation in human chromosomes. *Sci. Rep.* **6**, (2016).
105. Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The phasor approach to fluorescence lifetime imaging analysis. *Biophys. J.* **94**, (2008).
106. Jalink, K. & Rheenen, J. Van. *FilterFRET: Quantitative imaging of sensitized emission. FRET and Flim Techniques* **33**, (Elsevier B.V., 2009).
107. Jares-Erijman, E. A. & Jovin, T. M. FRET imaging. *Nature Biotechnology* **21**, (2003).
108. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, (2002).
109. Miyawaki, A. *et al.* Fluorescent indicators for Ca<sup>2+</sup> based on green fluorescent proteins and calmodulin. *Nature* **388**, (1997).
110. James, J., Swift, S., Norman, D. G. & Lamond, A. I. Quantitative analysis of

chromatin compaction in living cells using FLIM-FRET. *J. Cell Biol.* **187**, 481–496 (2009).

111. Llères, D. *et al.* Quantitative FLIM-FRET Microscopy to Monitor Nanoscale Chromatin Compaction In Vivo Reveals Structural Roles of Condensin Complexes. *Cell Rep.* **18**, 1791–1803 (2017).
112. Bienko, M. *et al.* A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat. Methods* **10**, (2013).
113. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
114. Cui, C., Shu, W. & Li, P. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology* **4**, (2016).
115. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* **74**, (2019).
116. Rouillard, J. M., Zuker, M. & Gulari, E. OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* **31**, (2003).
117. Navin, N. *et al.* PROBER: Oligonucleotide FISH probe design software. *Bioinformatics* **22**, (2006).
118. Beliveau, B. J. *et al.* OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).
119. Xu, Q., Schlabach, M. R., Hannon, G. J. & Elledge, S. J. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, (2009).
120. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, (2008).
121. Rao, S. S. P., Huntley, M. H., Durand, N. C. & Stamenova, E. K. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
122. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, (2010).
123. Hsu, T. C. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* **79**, (1975).
124. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, (2017).
125. Hua, N. *et al.* Producing genome structure populations with the dynamic and automated PGS software. *Nat. Protoc.* **13**, (2018).
126. Darrow, E. M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci.* **113**, E4504–E4512 (2016).

127. 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.* **5**, (2010).
128. Yuan, G. C. *et al.* Molecular biology: Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, (2005).
129. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, (2012).
130. Kilic, S. *et al.* Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1 $\alpha$ . *Nat. Commun.* **9**, (2018).
131. Visvanathan, A. *et al.* Modulation of Higher Order Chromatin Conformation in Mammalian Cell Nuclei Can Be Mediated by Polyamines and Divalent Cations. *PLoS One* **8**, (2013).
132. Platani, M., Goldberg, I., Lamond, A. I. & Swedlow, J. R. Cajal body dynamics and association with chromatin are ATP-dependent. *Nat. Cell Biol.* **4**, (2002).
133. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, (2006).
134. Fu, C. *et al.* Assessment of genetic and epigenetic variation during long-term *Taxus* cell culture. *Plant Cell Rep.* **31**, (2012).
135. Simara, P. *et al.* DNA double-strand breaks in human induced pluripotent stem cell reprogramming and long-term in vitro culturing. *Stem Cell Res. Ther.* **8**, (2017).
136. Murga, M. *et al.* Global chromatin compaction limits the strength of the DNA damage response. *J. Cell Biol.* **178**, (2007).
137. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, (2017).
138. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, (1998).
139. Walther, N. *et al.* A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. *J. Cell Biol.* **217**, 2309 (2018).
140. Green, L. C. *et al.* Contrasting roles of condensin I and condensin II in mitotic chromosome formation. *J. Cell Biol.* **125**, 1591–1604 (2011).
141. Bauman, J. G. J., Wiegant, J., Borst, P. & van Duijn, P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Exp. Cell Res.* **128**, (1980).
142. Solovei, I. *et al.* LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, (2013).
143. Robinett, C. C. *et al.* In Vivo Localization of DNA Sequences and Visualization of Large-Scale Chromatin Organization Using Lac Operator/Repressor Recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
144. Cannan, W. J. & Pederson, D. S. Mechanisms and Consequences of Double-Strand

DNA Break Formation in Chromatin. *Journal of Cellular Physiology* **231**, (2016).

145. Pennings, S., Meersseman, G. & Bradbury, E. M. Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.* **91**, (1994).
146. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, (2015).
147. Kallimasioti-Pazi, E. M. *et al.* Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLoS Biol.* **16**, (2018).
148. Zheng, C. L. *et al.* Transcription Restores DNA Repair to Heterochromatin, Determining Regional Mutation Rates in Cancer Genomes. *Cell Rep.* **9**, (2014).
149. Han, C., Srivastava, A. K., Cui, T., Wang, Q. E. & Wani, A. A. Differential DNA lesion formation and repair in heterochromatin and euchromatin. *Carcinogenesis* **37**, (2015).
150. Adar, S., Hu, J., Lieb, J. D. & Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, (2016).
151. Lemaître, C. *et al.* Nuclear position dictates DNA repair pathway choice. *Genes Dev.* **28**, (2014).
152. Fortuny, A. *et al.* Imaging the response to DNA damage in heterochromatin domains reveals core principles of heterochromatin maintenance. *Nat. Commun.* **12**, (2021).
153. Zullo, J. M. *et al.* DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, (2012).