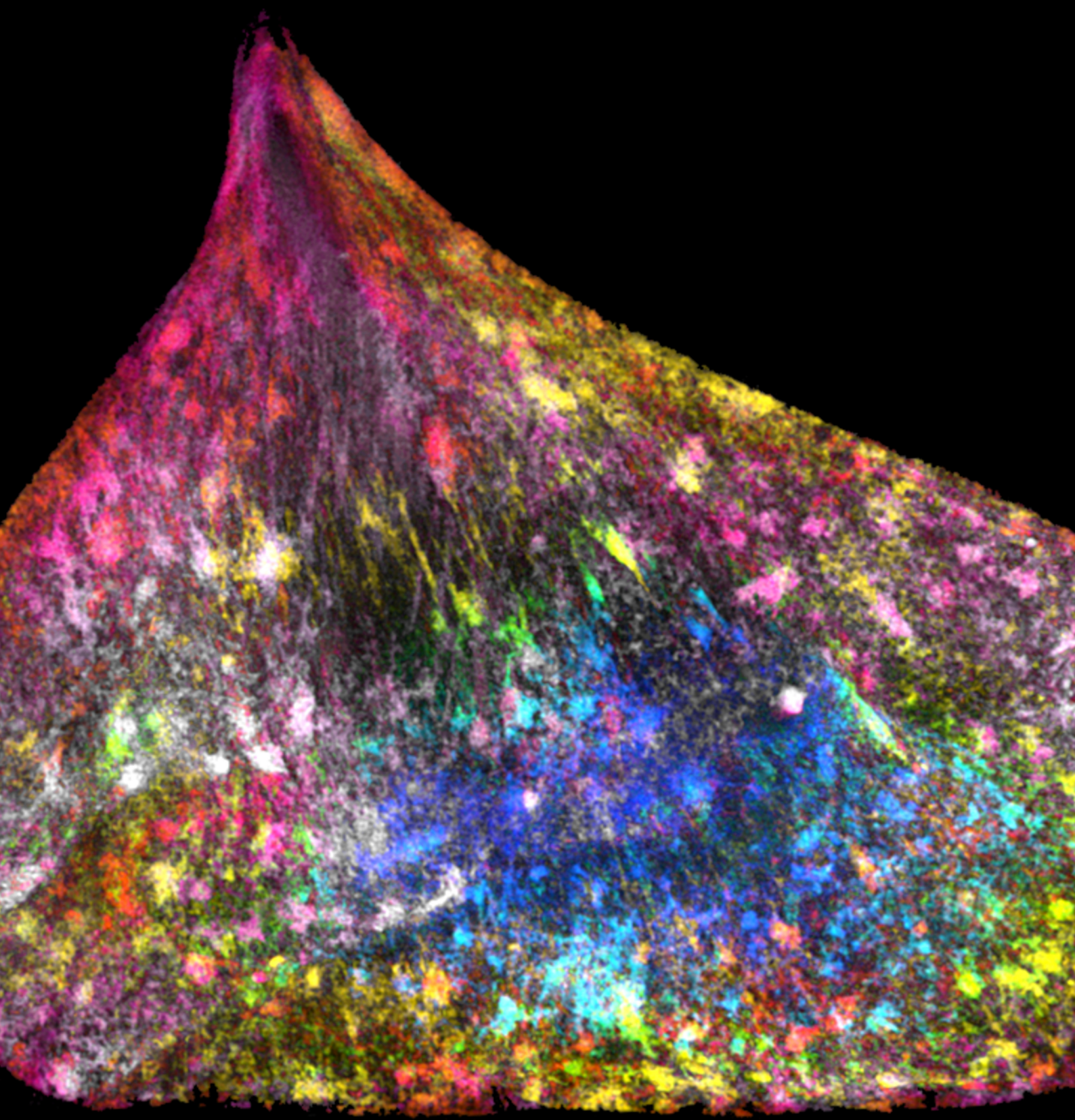
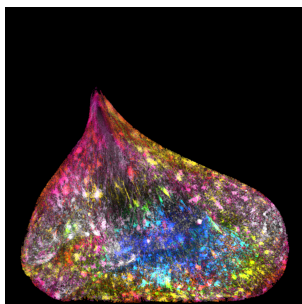






DNA drop
Wellcome Image Award 2017 winner.
Ezequiel Miron.

Comprehensive mapping of the 3D epigenome by high-content super-resolution image analysis

Ezequiel Miron

February 7, 2018

Thesis presented for the degree of Doctor of Philosophy

Discipline of Biochemistry, University of Oxford

Supervisors: Dr. Lothar Schermelleh, Prof. Robert Klose, Dr. David Pinto



*“The heart of the discerning acquires knowledge,
for the ears of the wise seek it out”*

Proverbs 18:15

Foreword

This thesis represents many years of hard work and dedication, not just on my part, but on the part of the many people who have helped me along the way. People whose help is the reason I have made it this far.

I would like to thank my supervisors for their constant input, critical assessment and the ideas shared with me throughout these years. I would like to thank Lothar Schermelleh for making me an integral part of his group. For striking the right balance between letting me develop as an independent researcher but always being interested and at-the-ready to discuss my work. Your passion for the science is evident in your expressions; your excitement is contagious, and it has helped me through the low points.

I want to thank David Pinto. I cannot conceive how I would have achieved all that I have without your guidance. You took me under your wing and taught me so many things. You instilled in me a drive to know the detail and the joy of figuring it out on my own. I take this with me as I look to new challenges. For all your help I am so grateful.

I am happy to have shared many years in the lab with Justin Demmerle and CassandraVictoria Innocent. You have been my day-to-day source of fun and I think I'm right in saying we have been a support to each other during our daily grind. Thank you for the many memories made together. I want to also extend my gratitude to the many people involved at the Micron advanced bioimaging facility. Everyone who has ever had me knocking on their door looking for people, reagents, equipment and answers to all my microscope related problems. Thank you for taking the time to help in all these small things.

I want to thank my wife, Rebeca. I thank you for patiently believing in me all these years. Thank you for your honesty, for being my constant drive to push myself to be the best me that I can be. Thank you for lifting my spirit when I needed it the most. I am so glad to have your smile and company, they mean the world to me.

Finally, I need to thank my family: my father, mother and my two sisters. I owe you all unimaginable thanks for decades of immeasurable support.

I am blessed.

Contents

Foreword	iii
Contents	iv
List of Figures	vii
List of Tables	x
Summary	xi
1 Introduction	1
1.1 The hierarchical organisation of the genome	1
1.2 3D architecture - Approaches to study higher order epigenetics	14
1.3 Microscopy and the genome	24
1.4 Aims of the investigation	36
2 Materials and Methods	39
2.1 DNA methods	40
2.2 Cell culture	47
2.3 Cellular staining	53
2.4 Replication-based foci labelling	57
2.5 TAD labelling methods	64
3 ChaiN - a high-throughput pipeline for super resolution image analysis	73

3.1	Acquisition and reconstruction	75
3.2	Feature segmentation	82
3.3	Measurements and statistics	90
4	G₁ 3D topology of the genome and epigenome	101
4.1	Interphase chromatin topology	101
4.2	The 3D epigenome	106
5	Replication-induced local perturbations in S phase	121
5.1	Local chromatin decompaction	123
5.2	Rearrangement of the 3D epigenome	125
6	Mechanisms of global chromatin perturbation	139
6.1	Global chromatin decompaction	142
6.2	Global chromatin compaction	144
7	Single TAD labelling developments	163
7.1	LAD-ID imaging	165
7.2	CRISPR imaging	167
7.3	Raser FISH imaging	177
8	Discussion	183
8.1	Aims and ChaiN developments	183
8.2	Chromatin landscape	184
8.3	The 3D epigenome	190
8.4	Concluding remarks	200
9	Appendix	203
9.1	Controls	203
9.2	Statistical analysis	209
9.3	Optical set up	211
9.4	Supplementary data	211
9.5	Materials	217

9.6 Code	222
Bibliography	227

List of Figures

1.1	The genome in higher dimensions	3
1.2	Histones and nucleosomes	8
1.3	Histone PTMS	9
1.4	2D chromatin looping	12
1.5	Chromatin conformation capture	16
1.6	Hi-C	17
1.7	Insulation at TAD boundaries	20
1.8	Cohesin and CTCF	21
1.9	Loop Extrusion model	23
1.10	Chromosome territories	25
1.11	Super-resolution	30
1.12	Resolving chromatin <i>in situ</i>	33
1.13	Functional zonation	38
3.1	ChaiN workflow	74
3.2	Biological alignment sample	80
3.3	Modulation contrast clean-up	83
3.4	Make workflow	85
3.5	Chromatin masking	86
3.6	Centroid segmentation	88
3.7	Chromatin density classes segmentation	91
3.8	Chromatin network dimension	94

3.9	Count and normalisation of marker in chromatin classes	97
3.10	Distance to IC	98
4.1	Chromatin volume	102
4.2	Chromatin density	103
4.3	Chromatin Surface area	104
4.4	Chromatin Network	105
4.5	Active marker distribution	107
4.6	Chromatin remodellers	108
4.7	Active marker distribution	109
4.8	Inactive marker distribution	112
4.9	Mid marker distribution	116
4.10	Structural marker distribution	118
4.11	Distance to IC	120
5.1	Sphase stages	122
5.2	S phase density per class	124
5.3	Sphase changes to all density distributions	127
5.4	Sphase changes to RNAP-S2P distribution	129
5.5	Sphase changes to H3K36me3 distribution	129
5.6	Sphase changes to H3K27me3 distribution	131
5.7	Sphase changes to HP1A distribution	132
5.8	Sphase changes to H4K20me1 distribution	134
5.9	Sphase changes to CTCF distribution	134
5.10	Sphase changes to SAF-A distribution	135
5.11	Sphase changes to cohesin distribution	136
6.1	G ₀ chromatin	141
6.2	Chromatin decompaction	143
6.3	Modulation contrast of different chromatin topologies	144
6.4	Chromatin compaction	146

6.5	Chromatin to nuclear volume	147
6.6	chromatin to nuclear volume	149
6.7	Network filament changes	149
6.8	Chromatin surface area to volume	150
6.9	Chromatin compaction changes to all density distributions	154
6.10	Compaction effect on distance to IC	156
6.11	Senescent changes to all density distributions	160
7.1	Imaging TAD	164
7.2	LAD-ID	166
7.3	LAD-ID vs RD	168
7.4	CRISPR methodology	171
7.5	<i>in vitro</i> cleavage methodology	172
7.6	Results of cleavage assay	173
7.7	CRISPR-delivery	175
7.8	Raser FISH	179
7.9	Raser FISH vs IF	180
8.1	TAD filament network model	187
8.2	Perturbations of the TAD filament network model	189
8.3	Epigenetic functions on the TAD filament network model	192
8.4	TAD functional model	197
9.1	Resolution controls	205
9.2	Antibody penetration	207
9.3	Reversible S phase disruption to chromatin	208
9.4	Lightpath	212
9.5	Nuclear volumes	212
9.6	S phase distributions	214
9.7	Exogenous compaction distributions	215
9.8	Senescence distributions	216

List of Tables

9.1	Spot numbers	209
9.2	Reagents	217
9.3	Solutions	218
9.4	Equipment	219
9.5	Antibodies	220
9.6	Cell lines	221

Summary

A full understanding of the relationship between the density and packing arrangements of chromatin in interphase mammalian nuclei and epigenetic function at the nanometer size scale remains an elusive goal in the field of chromatin biology. Over the last years great technical leaps have been made through sequencing-based methods to address this question. Advancements have also allowed research to bypass long-held optical limits to resolve chromatin at a scale where its 3D topology can start to be analysed.

The work in this thesis represents attempts at high-throughput data mining of 3D SIM datasets, encoding rich nanometer-scale spatial information from immunofluorescence detection of histone modifications and key epigenetic markers relative to a chromatin landscape. To do so, an automated and high-throughput image analysis workflow (ChaiN , for ChaiN analysis of the *in situ* Nucleome) was developed.

Novel metrics for the quantitation and correlation of chromatin and the 3D epigenome reveal a chromatin network of filaments at the size scale proposed from sequencing approaches, and with segregated regions of genomic activities as a function of chromatin accessibility (its density and depth). Furthermore, ChaiN has allowed characterisation of the local and global rearrangements of chromatin when subjected to replication pressures or exogenous perturbation showing for the first time how individual genomic markers are affected at different locations throughout the chromatin network.

The model hypothesised from these results tries to reconcile previous data obtained from Hi-C population ensemble studies and *in silico* modelling, to single cell observations at super-resolution.



Introduction

1.1 The hierarchical organisation of the genome

The three-dimensional (3D) organisation of chromatin in interphase nuclei is increasingly understood to be an important epigenetic regulator of genome function. Such understanding comes from decades of meticulous studies, from early evidence that DNA in the nucleus encodes genetic inheritable information (Hershey and Chase, 1952). The interest in the study of the nuclear compartment has grown exponentially as further evidence has mounted in the decades since about the importance of DNA, not just for hereditary of traits, but also as a physical molecule that can respond to internal and environmental stimuli. The plasticity and capacity to seemingly “override” the final expression of hard-coded genes by mechanisms in the nucleus has since lead to the blossoming field of “epigenetics”, currently proposing a variety of DNA biology models to explain observable genome regulation paradigms.

It has become clear that control of the genome does not occur through a single centralised mechanism but by layers of mechanisms. Indeed reliance on a singular pathway would be a risky strategy when considering the value of correct interpretation of the genome. The genome is thus organised in a

way that allows multiple checkpoints and controls to act in unison. In this way, mutations which may affect any single control mechanism have their effect reduced by the compensating action from any other checkpoint.

The regulation of genome function is also now known to be hierarchical (Figure 1.1)(Gibcus and Dekker, 2013). Starting with simple motif-specificity based on Watson-Crick pair recognition, to the simultaneous firing of multiple neighbouring origins of replication in small domains (Pope et al., 2014) up to whole chromosomal interactions with the nuclear lamina (Guelen et al., 2008). The existence of some previously described architecture is now questioned, and debate is ongoing on the nature of higher order organisation at the megabase scale beyond simple looping (Figure 1.1).

Genome organisation

At the simplest level the whole genome is encoded in four nitrogenous bases: adenine (A), cytosine (C), guanine (G), thymine (T). A 3 base codon system which redundantly encodes the 20 naturally occurring amino acids is transcribed to mRNA by the action of RNA polymerase enzyme complexes in a process known as transcription. The mRNA is exported from the nucleus where it is read and translated to polypeptide chains by ribosomes in conjunction with tRNA. This central dogma of DNA biology is completed by the action of DNA polymerase, duplicating the genome during the synthesis (S) phase of the cell cycle, preparing a cell for division. This dogma is part of textbooks on the subject (Voet and Voet, 1995; Alberts et al., 1983). The minutia of these mechanisms are a point of continued investigation but fall beyond the scope of this introduction.

The open reading frames (ORFs) of genes are directly under the control of promoters. Promoters are sequence-specific motifs regulating specificity and strength of binding-interactions between the RNA polymerase (RNAP) and the upstream region of the gene (involved in transcriptional initiation). There are many different promoter motifs, for weak vs strong expression or

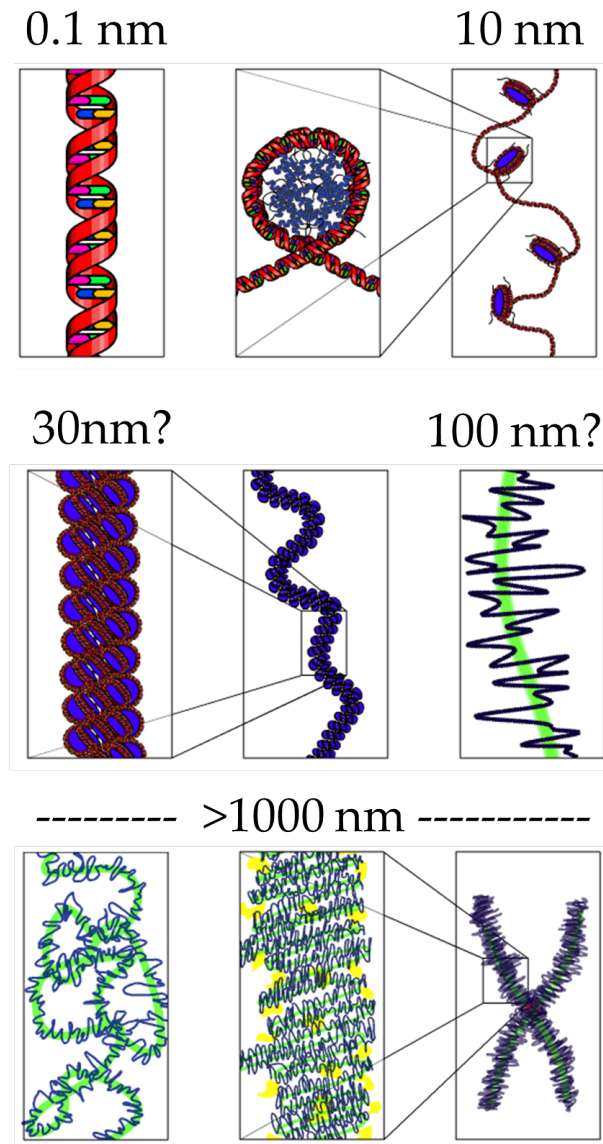


Figure 1.1: **The genome in higher dimensions.** Schematic of dimensions of genome organisation beginning from double stranded DNA at 0.1 nm, to 10 nm chromatin fibre up to full mitotic chromosome architecture. Mid-scale organisation at 30 nm to 100 nm during interphase is still under debate. Figure adapted from Richard Wheeler (Creative Commons Attribution).

constitutive vs temporally restricted expression. However there is still only three canonical RNAPs (I/II/III).

The bridge between these DNA elements and the transcription machinery is composed of transcription factors (TFs) which have domains that can recognise specific promoters and other domains to recruit RNAP. The most common are “general TFs” (GTFs) which are required for basal transcription. These are the TFIID-A/B/E/F/H TFs (Thomas and Chiang, 2008). TFs add an immediate layer of complexity to this one-dimensional regulatory paradigm. Although many TFs do act to upregulate transcription through increased promoter-RNAP contact, many TFs act as antagonists of such interaction. Furthermore as previously suggested, many TFs are expressed as required after endogenous or exogenous cues for a short period of time. For example, the Hox gene cluster TF in embryonic development (Moens and Selleri, 2006; Lobe, 1992; Pawson, 1993; Shamovsky and Nudler, 2008; Benizri et al., 2008).

In eukaryotes a core promoter of the RNAP type II (responsible for transcribing most mRNA) is composed of the transcription start site (TSS, +1), RNAP binding site, and TF binding sites (Juven-Gershon et al., 2008). One of the most well known of these TF binding sites is the TATA box. Approximately 24% of human gene promoters contain this motif (Yang et al., 2006) and it is found not just in eukaryotes but also in archaea (Kadonaga, 2011). Bacteria also have short consensus sequences upstream of genes, most notably the -10 Pribnow box (TATAAT) and the -35 element, separated by 17bp without consensus sequence (Pribnow, 1975). Other promoter not immediately upstream of the TSS are also found. These are generally termed “proximal” for promoters approximately 250bp upstream of the TSS or “distal” for any genomic distance further upstream.

Genes are not the only genomic element. In fact, of the approximately 19,000-20,000 protein coding genes found in the human genome (Ezkurdia et al., 2014) all their exons together make up only <2% of the genome (Con-

sortium, 2001). At first the remaining ~98% of the genome was considered junk, however we now know of the existence of many essential non-protein-coding sequences. Between genic exons lie the introns, making up a large portion of the genome, ~26% (Gregory, 2005). At least 8% (Consortium, 2012) maybe as high as 20%-40% (Stamatoyannopoulos, 2012) of the genome is composed of regulatory sequences of gene expression, eg: enhancers. And studies performed by the ENCODE consortium in the last 5 years suggests that the percentage of transcribed genome (genome which is bound by regulatory proteins or has other known function) may be as large as 80% (Stamatoyannopoulos, 2012). It is worth pointing that not all transcription may be biologically relevant and some may be due to unspecific polymerase activity (Palazzo and Akef, 2012).

Studies of comparative genomics have also found 5% of non-coding DNA to be highly conserved (Nix and Eisen, 2005), with much of these being composed of macro-scale structural elements of chromosome anatomy: anchoring points for spindle kinetochores in cellular division (centromeres, Pluta et al. (1995)) or hexanucleotide repeats which protect erosion of the ends of linear genomes after replication (telomeres, (Mondello and Scovassi, 2004)), respectively. There are of course many legitimate transcribed but not translated RNAs whose role is also vital to cellular maintenance. Many transcribed RNAs have roles as ribosomal and transport RNA (r/tRNA) (Iyer et al., 2015) used in translation, micro and short nuclear RNA (mi/snRNA) for RNA silencing and post-transcriptional splicing and regulation (Ambros, 2004; Eddy, 2001) as well as long non coding RNA (lncRNA) (Managadze et al., 2013). Pseudogenes, non-functional copies of actual genes, are also present in the genome totalling ~13,000, constituting ~1% of the genome (Pei et al., 2012).

By far the largest proportion of the genome is made up of repetitive and or mobile genetic elements comprising up to ~50% of the genome. These highly variable, low complexity regions often occur in tandem and have been historically exploited as markers for forensic and paternity DNA analyses (Tre-

angen and Salzberg, 2011). Transposons (transposable DNA elements from viral genomes) can replicate copies of themselves and insert at distant sites throughout the host genome, for example the Alu transposon, with ~50,000 copies found throughout intra/inter-genic regions (Bennett et al., 2008).

The precise figures of all these elements is still highly contested, but the orders of magnitude give a decent estimate as to their true representation on the linear genome.

First dimension of epigenetics

The DNA molecule is well suited to encode every gene, gene variant and regulatory sequence with room to spare and it is also robust enough for the faithful transmission of information from parent to progeny. However, it does pose challenges for its recurrent use by organisms. Firstly, DNA is an extremely large molecule; all the DNA from every chromosome in a single nucleus would be approximately 2 m long, yet it must be packaged in nuclei which are in the order of 10 μm to 20 μm in diameter and 4 μm to 15 μm tall for adherent epithelial diploid cells. Even with its narrow width of just 2 nm, the density of DNA packing is very high. Many textbooks still display an artistic representation of a random and disorganised interphase genome akin to a tangled string or a bowl of spaghetti. Nevertheless, if random packing would be employed it would be difficult to organise the genome such that functionally similar genomic regions could be read and processed efficiently. Furthermore, as individual chromosomes must be repeatedly compacted for division, random packing would not result in efficient sub-grouping of the nuclear compartment to facilitate this process.

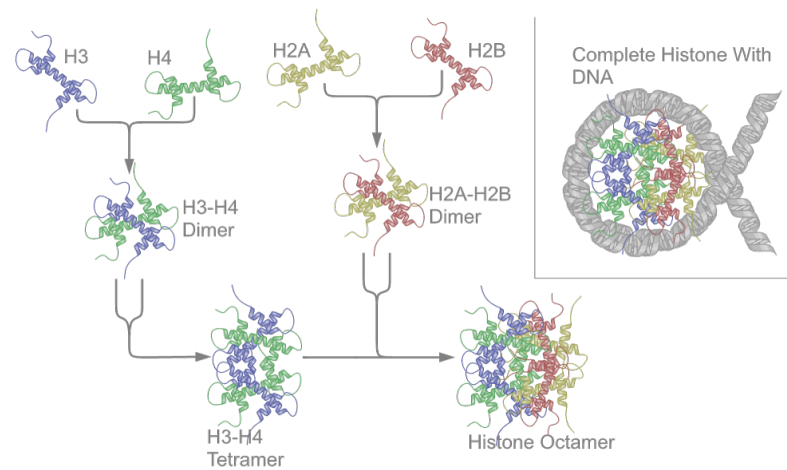
It is now known that the basic unit of DNA packing is the nucleosome. Nucleosomes are composed of an octamer of two sets of core histones (histone-domain-containing proteins: H2A, H2B, H3, H4). DNA is spooled around the nucleosome such that 146bp are in contact with its circumference as well as at least an additional 10bp of non-nucleosomally wrapped DNA, acting as a

linker between neighbouring nucleosomes (Figure 1.2(a))(Luger et al., 1997). The size of the linker region is variable but has a average length of 20bp and remains under 80bp. The linker histone H1 protein (although not a true histone given its lack of histone fold domain) binds to the inter nucleosomal DNA to stabilise this region (Ramakrishnan et al., 1993). DNA sequentially wrapped in nucleosomes leads to the “chromatin” fibre. A 10 nm thick fibre whose appearance in electron micrographs when first studied resembled “beads on a string” (Figure 1.2(b))(Olins and Olins, 1974).

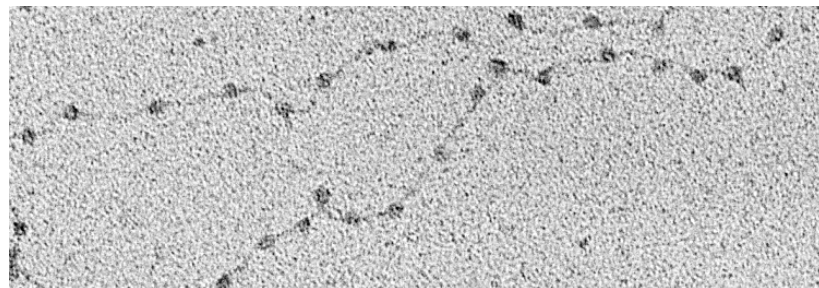
Histones contain unstructured N-terminal domains that are susceptible to post translational modifications (PTMs) by chromatin remodeller enzymes. The deposition of a diverse range of chemical moieties such as acetyl (Ac), methyl (Me), phospho (Ph) and ubiquitin (Ub) groups in combination with a diverse number of positions along these N-terminal tails have been shown to directly regulate genomic activities at these sites, leading to much investigation over the years on this “histone code”(Figure 1.3).

Acetylation is the result of histone acetyl transferases (HATs) and is removed by histone deacetylases (HDACs). Acetylation is known to conduce a more open chromatin environment, euchromatin. This is thought to be the product of neutralising the positive charges on the arginine (R) and lysine (K) amino acids of the N-terminal tails which could otherwise serve to stabilise DNA-nucleosome interaction from multiple Rs and Ks to the negative DNA backbone. Acetylated nucleosome tails extend further away radially from the nucleosome, also decreasing the efficiency of nucleosome packing.

Methylation is also known to occur in both R and K amino acids, but its functions are far wider. The methylation of certain lysines are to date the most studied, but even so a comprehensive understanding of their roles continue to elude consensus. A factor compounding the complexity of the histone code is its non-linearity. The position of the PTM relative to the tail’s end has no correlation to function or strength, and neither do the number of the same PTM on the same amino acid. For example, the mono-methylation of lysine 9 on



(a) Nucleosome assembly



(b) Beads on string

Figure 1.2: Histones and nucleosomes. (a) Nucleosome octamer made from a tetramer of two H3-H4 (blue and green, respectively) histone dimers and a tetramer of two histone H2A-H2B (yellow and red, respectively) dimers. Inset shows DNA (grey) wound around the nucleosome. Image Source: Richard Wheeler (Creative Commons Attribution). (b) Electron micrograph of chromatin fibre showing DNA wound around nucleosome octamers in what can be described as “beads on a string”. Modified from Chris Woodcock - Cell Image Library 709 (Creative Commons Attribution).

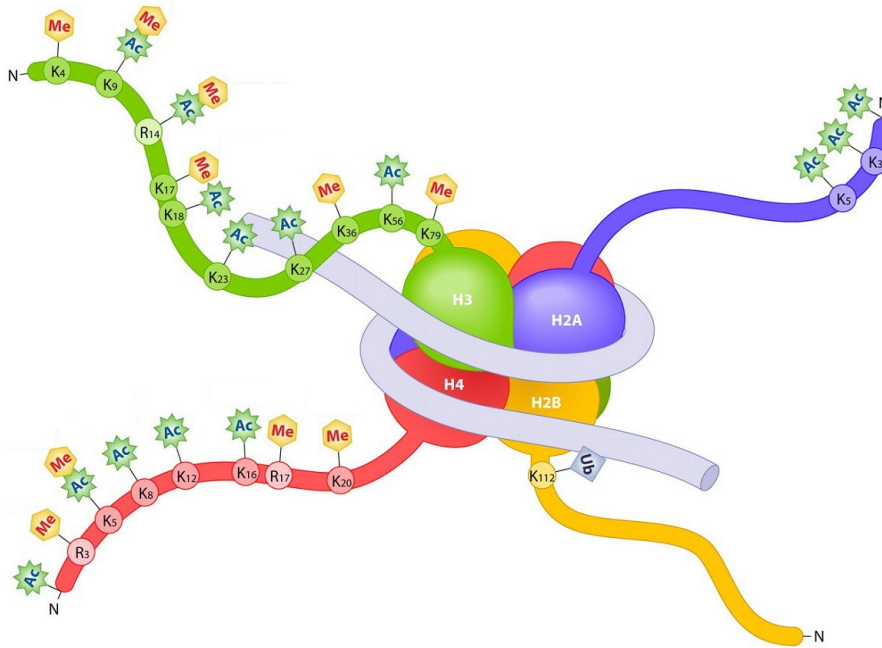


Figure 1.3: **Histone PTMS.** Schematic of covalent histone post-translational modifications (PTMs): methylation (Me), acetylation (Ac), and ubiquitination (Ub). Histones are modified on the N-terminal tails changing the interaction between DNA (grey) and the nucleosome and between DNA and DNA-binding proteins. Modified from Cui and Miao (2010)

histone 3 (H3K9me1) is conducive to genomic activation, while subsequent methylations cause repression (Figure 1.3, green)(Barski et al., 2007). Over 4000 known individual histone PTM are found on gene promoters in total throughout the genome. But in fact, the combinatorial mathematics suggests an order of magnitude of over 10^{11} possible combinations (taking all known levels and types of modifications and the amino acids which can be modified). Within all these possibilities a set of 17 PTM are present together in over 3000 genes (Wang et al., 2008). What biochemical elements contribute to the molecular recognition of histone PTMs by chromatin readers and writers is still an open question (Wang et al., 2010) and it is not known whether further strong correlations of PTM patterns will ever be found.

The remodelling of the chromatin fibre by histone PTMs can lead to eu-chromatin and heterochromatin (Dillon and Festenstein, 2002). These names

are vestigial remnant of their first description as bands of different intensity (heteropycnotic, weaker or stronger, respectively) of Giemsa band staining in cytogenetic karyotyping experiments (Passarge, 1979). Euchromatin tends to replicate in early S phase and it is more sensitive to nuclease digestion than heterochromatin, possibly due to the lack of protection offered by the steric effect of tightly packed nucleosomes and their N-terminal tails. Euchromatin is also generally gene-rich and the proportion of genes transcribed in this environment is relatively high. Heterochromatin on the other hand is characterized by a denser nucleosome array. It is replicated in late S phase and hypo-sensitive to nuclease activity. Facultative heterochromatin is comprised primarily of tissue-specific genes and can be re-organised to euchromatin through the action of developmentally regulated and tissue specific chromatin remodellers. Constitutive heterochromatin is often gene poor, and correlates to repetitive satellites of the linear genome, often located at sites of structural non-coding DNA regions (eg: at pericentromeric regions) .

2D epigenetics

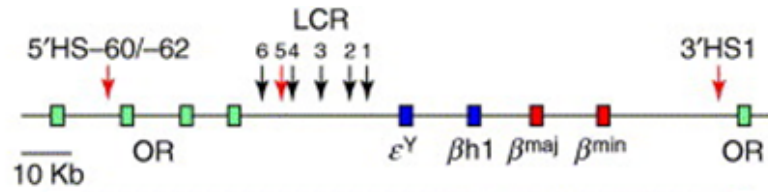
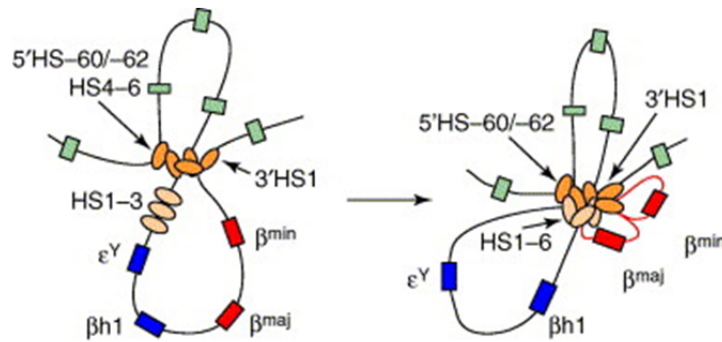
The organisation of DNA to a 10 nm fibre compacts the length occupied by the genome by over 7 times, yet even this compaction is by far not enough to fit lengthwise in nuclei. Additional levels of higher order chromatin structure are required. Replacing histones by protamines is an effective way to achieve the required compaction. This is the mechanism employed by mammalian sperm cells. However, when considering this organisation it is worth noting that the high degree of compaction which results from exchange to protamines is almost crystalline in nature. The rigidity and tight packing makes for a high hydrodynamic structure, excellent for transporting the genetic material, but also prevents practically all protein-DNA interaction. During interphase in diploid somatic mammalian cells, this organisation would not be compatible with the many genome processes that must take place.

The 10 nm chromatin fibre is proposed to be arrayed into a 30 nm solenoid

fibre, stabilised by the local addition of excess H1 linker (Figure 1.1)(Robinson et al., 2006). This fibre has been found *in vitro* from electron micrographs and is thought to be the compact state chromatin takes when packaged as heterochromatin. Chromatin remodelling may remove sufficient nucleosomes to destabilise the 30 nm solenoid, making DNA accessible to the action of DNA interacting enzymes (Smith and Peterson, 2005). Debate about the occurrence of the 30 nm fibre in normal cell physiology is still ongoing as it has not been found *in vivo* (Hansen, 2012). Further solenoid-based supercoiling has been proposed from the 30 nm fibre as 100 nm to 250 nm wide fibres, coiled again into a 700 nm fibre (the chromonema model, Figure 1.1) (Woodcock et al., 1976; Belmont and Bruce, 1994). However, with the existence of their smallest sub-unit, the 30 nm fibre, in question it is possible these supercoiled structures may only be achieved *in silico* or under extreme non-physiological conditions *in vitro*.

At the very least, chromatin must fold loops in two dimensions. This is true because of the nature of genomic elements called enhancers. Enhancers, as their name suggests, act on promoters to enhance their transcriptional output. However, it has been shown that enhancers may act on specific promoters from over 1Mb upstream or downstream from the TSS in both eukaryotes and prokaryotes alike (Kulaeva et al., 2012). These expanses of genome simply cannot be bridged by any protein complex in 1D. Organization into flexible loops could bring enhancer and promoter elements together in 2D space (Figure 1.4). Yet flexibility does give freedom for enhancers to potentially interact with many multiple promoters, which may not be desirable. The discovery of the presence of genetic insulator elements as boundaries between enhancers and promoters in 1D, which also block their interaction in 2D gives cells additional epigenetic control in the latter dimension (Burgess-Beusse et al., 2002).

One of the best examples of loop control of gene expression comes from the haemoglobin gene cluster (Figure 1.4(a)). The globin protein variant changes during development from predominantly ϵ in primitive erythroid cells, to

(a) β -globin locus

(b) ACH

Figure 1.4: **2D chromatin looping.** (a) Enhancer(green)-promoter(blue/red) interactions in 2D space at the β -globin locus in mouse foetal liver cells . (b) Active chromatin hub (ACH) rearrangement during distinct developmental stages. Changes in the enhancer-promoter interaction causes change in the transcriptional programme of the β -globin locus. Adapted from Dean (2005).

β and δ in adult liver cells (Grosveld et al., 1993). However, some adults suffer anaemia in the form of β -thalassemia, despite having a non-mutated β locus (Van der Ploeg et al., 1980). It was found that these patients had perturbed genomic sequences upstream of the globin gene cluster, in a region which became known as the locus control region (LCR) (Kioussis et al., 1983), which was also shown to be hypersensitive to nuclease degradation (i.e.: in an open euchromatin conformation) in erythrocytes (Tuan et al., 1985). The LCR is now known to contain enhancer elements, but may only act upon a single target gene through long range looping (Wijgerde et al., 1995). The distance between the LCR and the target gene is negatively correlated to the effect of the LCR on this gene, and offers a clear explanation as to why the genes in the globin locus are in the order they occur during development as they ap-

pear downstream of the LCR (Hanscombe et al., 1991). The globin locus possesses sufficient distance between the LCR and the gene cluster to allow the required looping, but the act of forming interactions with any single one of the loci greatly reduces the chance of a looping interaction with any of the other competing loci. In mice, the looping structures formed are termed the “active chromatin hub” (ACH) (Palstra et al., 2003) and genetic experiments confirmed the strong correlation between β -globin expression levels and the appearance of favourable LCR to β -globin locus interactions at the ACH (Figure 1.4(b))(Simonis et al., 2007).

Beyond this specific example, evidence for the formation of long-range loops of ~100kb on average is increasingly abundant. Its correlation with transcriptional enhancement has led to models of “transcription factories”; regions in which a complex formed from aggregation of actively transcribing RNA polymerases spool multiple chromatin loops from a central foci (Iborra et al., 1996; Cook, 2002). Multiple genes being transcribed from different loci at the same factory is hypothesised to cause looping between these genomic regions. However, there are various unknowns to this model. Primarily, although DNA is a thin molecule compared to the proposed sizes of transcription factories, the chromatin molecule is still very large. It is difficult to reconcile how transcription would pull whole chromatin regions to such a small local event in the spooling manner described (Gavrilov et al., 2010), rather than translocate itself along the DNA strand towards the rest of chromatin. So far no known factors are involved in the immobilisation of RNAP to the nuclear matrix to provide an anchoring support (Pederson, 2000). Secondly, there is evidence that RNAP can transcribe genes on both + and - strands at the locus it interacts with. The mode of action that allows this to occur relies on free rotation of the transcription complex. A stationary RNAP would have to employ yet-undiscovered transcriptional mechanisms to achieve this.

1.2 3D architecture - Approaches to study higher order epigenetics

Despite decades of work, very little is certain about the way chromatin folds in single interphase cells at the “meso-scale”, beyond the 30 nm fibre, from tens of nanometers up to the micrometer size scale. Sub-chromosomal features above the micrometer scale (chromosomes, chromosome arms, centromeres) are “large-scale” and have been characterised by decades of confocal imaging. Over the last few years, great strides have been made towards addressing these unknowns by the advent of new genomic and imaging technologies. Nevertheless, even with state of the art instrumentation, open questions still remain. The issue in part, stems from the complex nature of chromatin, the limitations of all techniques currently employed to its investigation, and the lack of cross-correlative studies between newly developed techniques. The main tools in recent attempts to characterise meso scale interphase chromatin organisation are discussed subsequently.

Chromatin capture methods

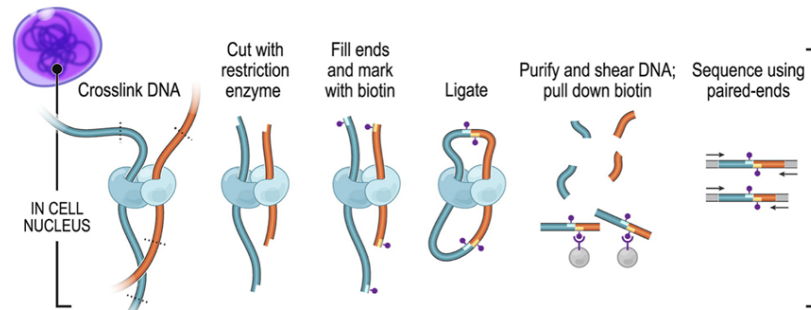
Current research in the field of chromatin biology is dominated by genomic based techniques such as chromatin immuno-precipitation sequencing (ChIP-Seq), and a variety of techniques based on chromatin conformation capturing (3C)(Figure 1.5)(Dekker et al., 2002). These variants (4C, 5C and Hi-C) benefit from the reduced cost of genome sequencing by the availability of massively paralleled next generation sequencing methods (de Wit and de Laat, 2012). Their high linear genomic resolution allows researchers to correlate chromatin interactions with results on the cross-linked position of proteins, histone modifications, etc (ChIP-Seq). 3C studies have also proven useful in elucidating the relative positioning of target sequences to other genomic features, such as the promoter-enhancer interactions via looping discussed previously for the globin locus (Tolhuis et al., 2002). This example is particularly true for 3C, a

so called “one vs one” technique which works between a single pair of known genomic loci (Figure 1.5(b), 3C). Furthermore 3C allows capturing of transient chromatin interaction from an *in vivo* state (Carter et al., 2002).

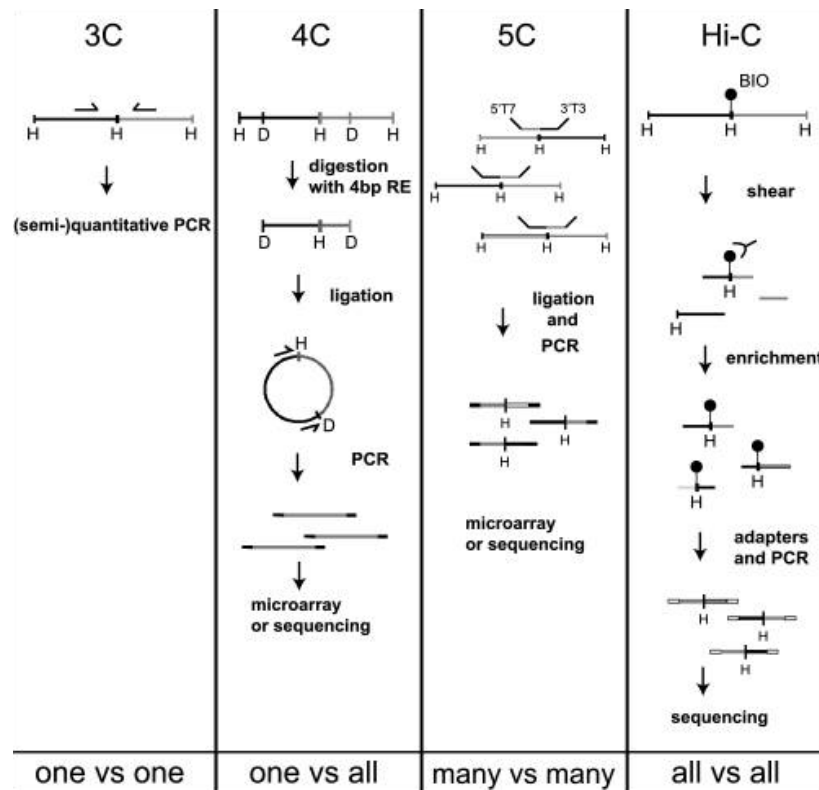
3C-on-chip (4C) is an approach which only requires *a priori* knowledge of one genome loci to generate contact frequencies from it to all other interacting chromatin regions (Simonis et al., 2006). This makes it useful to “fish” for unknown interactions, e.g.: of new enhancers and promoters, as a “one vs all” method (Figure 1.5(b), 4C). 3C-carbon-copy (5C, an “all vs all” method) requires the sequence of both chromatin regions studied to be known but yields a 2D map of all interaction frequencies between these two regions (Figure 1.5(b), 5C)(Dostie et al., 2006). However, 5C is not practical for assaying the whole genome as it would require primers in the order of millions to tens of millions. 5C is therefore always used for local regions as a “many vs many” approach. Hi-C is a true “all vs all” method employing high-throughput sequencing to map genome-wide chromatin-chromatin contacts frequencies (Figure 1.5(b), Hi-C)(Lieberman-Aiden et al., 2009). The pair of cross-linked regions which were in contact *in situ* can then each be aligned against a reference genome (a prerequisite of the technique), and counted to yield an individual pairwise contact. A pool of crosslinked genomes must be assayed simultaneously to build up a tally of contact frequencies. This is important in order to normalise true readout from noise.

Genomic studies using Hi-C, have shown that the linear genome is clustered in 3D regions of ~1Mb in size on average. These have been termed topological associating domains (TADs). TADs are regions of elevated chromatin contact frequencies over background levels (Lieberman-Aiden et al., 2009). The TAD organisation has been shown to remain conserved for many genomic loci across different tissues within an organism and, more strikingly, across related species (Jost et al., 2016).

Another genomic based approach, Dam-Identification (Dam-ID), relies on the introduction of an exogenous enzyme, the prokaryotic Dam DNA methyl-

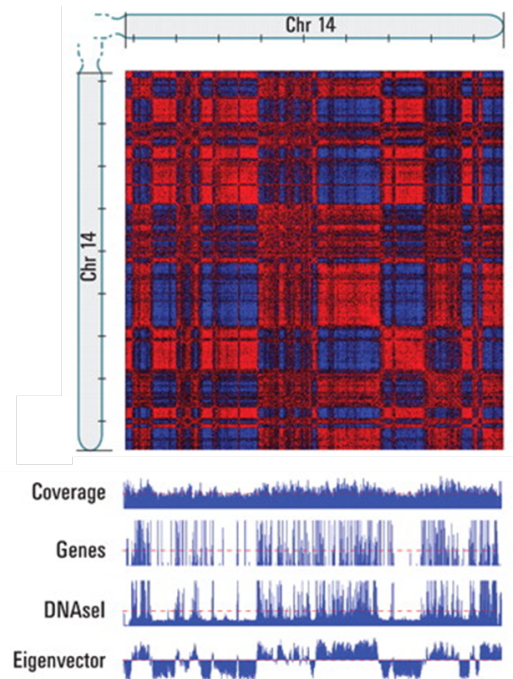


(a) Chromatin capture

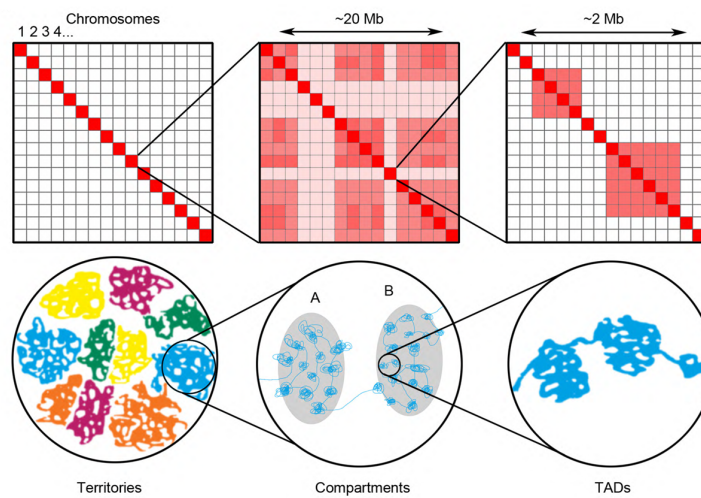


(b) Conformation capture techniques

Figure 1.5: **Chromatin conformation capture.** (a) Protocol for fixation of 3D conformation *in situ*. (b) Differences in protocols for 3C, 4C, 5C and Hi-C techniques which make each suitable for different experimental requirements, from “one vs one” to “all vs all”. Adapted from Lieberman-Aiden et al. (2009) and de Wit and de Laat (2012)



(a) Hi-C correlation



(b) Hi-C interpretation

Figure 1.6: **Hi-C.** (a) Eigenvector heatmap of Hi-C data for chromosome 14 correlating gene density and DNase sensitivity. Modified from Lieberman-Aiden et al. (2009) and Selvaraj et al. (2013). (b) Schematic of general Hi-C contact frequency heatmaps (top) showing a fractal pattern at various genome scales. Corresponding schematic of 3D representations of Hi-C heatmaps (bottom) showing chromosome territories down to the final unit of high chromatin interaction frequency: Topologically Associating Domains (TADs). The 3D arrangement of the meso-scale “A and B” sub-chromosomal compartments in single cells is not as depicted from pure interpretation of Hi-C data. Adapted from Phillips-Cremins (2014).

transferase (Greil et al., 2006). Dam-ID was developed for the identification of chromatin sequences which have undergone transient interactions with a particular nuclear protein. By fusing Dam to a target protein e.g.: a lamin protein of the peripheral nuclear lamina, any DNA within molecular distance to the lamina becomes methylated (m6A) (Kind et al., 2013). Due to the fact that the methyl mark is deposited by an enzyme of exogenous origin and it does not occur naturally in mammals, it is not recognised by any endogenous readers or remodellers of DNA methylation. Using the exogenous m6A marker the Dam-marked DNA can be tracked (§2.5).

Parallel studies to Hi-C using a variant Dam-LaminB fusion have shown that interactions between chromatin and the nuclear lamina also occur preferentially in domains of ~1Mb, termed lamina associating domains (LADs). Correlative bioinformatics on these datasets revealed a clear correlation between the domain boundaries of LADs and TADs, further supporting this modular architecture. They have also revealed that LADs harbour mostly heterochromatic environments, being gene poor with low expression levels. LADs are not merely artefacts of chromatin at the lamina. Due to the physical limitation of the surface area of the lamina not all LADs detected from a pooled cell population can be at the lamina in all cells. It has been shown however, that despite becoming internalised, and away from the lamina, that LADs retain their modularity. Many internal LADs reposition to the edge of nucleoli, these are called nucleolar associating domains (NADs). NADs share the same heterochromatic characteristics as LADs, being in fact composed of a selection of the same domains and also have strong boundary correlations with TADs.

On average TADs are flanked by sites of elevated insulator interaction in 1D. Given the role of insulators in dividing the genome into smaller domains of synchronous regulation (discussed previously), it is easy to suppose that their strong correlation with TAD boundaries (Figure 1.7(a)), confers individual TADs some degree of regulatory independence from their genomic neigh-

bours (Perkel, 2015; Ong and Corces, 2014). Theories of DNA looping between enhancer and promoters precede recent TAD characterisations, however the same tools that enabled TAD detection have confirmed loops occur within TAD loci. Loops interacting within TADs are hypothesised to be of a more transient nature and therefore less stable than the whole TAD architecture. In fact TADs remain stable during development and differentiation despite looping rearrangements between enhancer and promoters within them. As such, many now believe that TADs may represent the fundamental unit of genome architecture, harbouring (and possibly facilitating) a defined number of smaller transient looping interactions within them. Important evidence to support this comes from work carried out on the deletion and translocation of TAD boundaries (Lupiáñez et al., 2015). Due to the insulating environment at TAD boundaries, most enhancers are not able to influence promoters across them on the 1D linear epigenome. Removing the TAD boundary can therefore allow these interactions to occur, resulting in a non-physiological gene transcription programme. One clear example is evident by the appearance of polydactyl phenotypes in mammals (Figure 1.7(b))(Lupiáñez et al., 2015).

Of all insulators found at TAD boundaries, the CCCTC-binding factor (CTCF) has been strongly implicated with the stabilisation of TAD architecture (Phillips and Corces, 2009). Although the exact mechanisms of TAD formation are still not known, there is strong evidence suggesting that looping caused in part by the co-occupancy of CTCF with the cohesin protein may serve to define the ends of TADs on the linear scale (Figure 1.8(a))(Ji et al., 2016). Cohesin derives its name from its role in the cohesion of sister chromatids during mitosis or meiosis by the entrapment of chromatin in its ring-shaped complex (Gruber et al., 2003).

The co-occupancy of cohesin and CTCF is important for the formation of TADs and it is taken into consideration in the most current polymer models of TAD formation. Of these, loop extrusion (LE), dynamic (diffusion-driven) looping (DL) or Strings and Binders Switch (SBS) have been the strongest

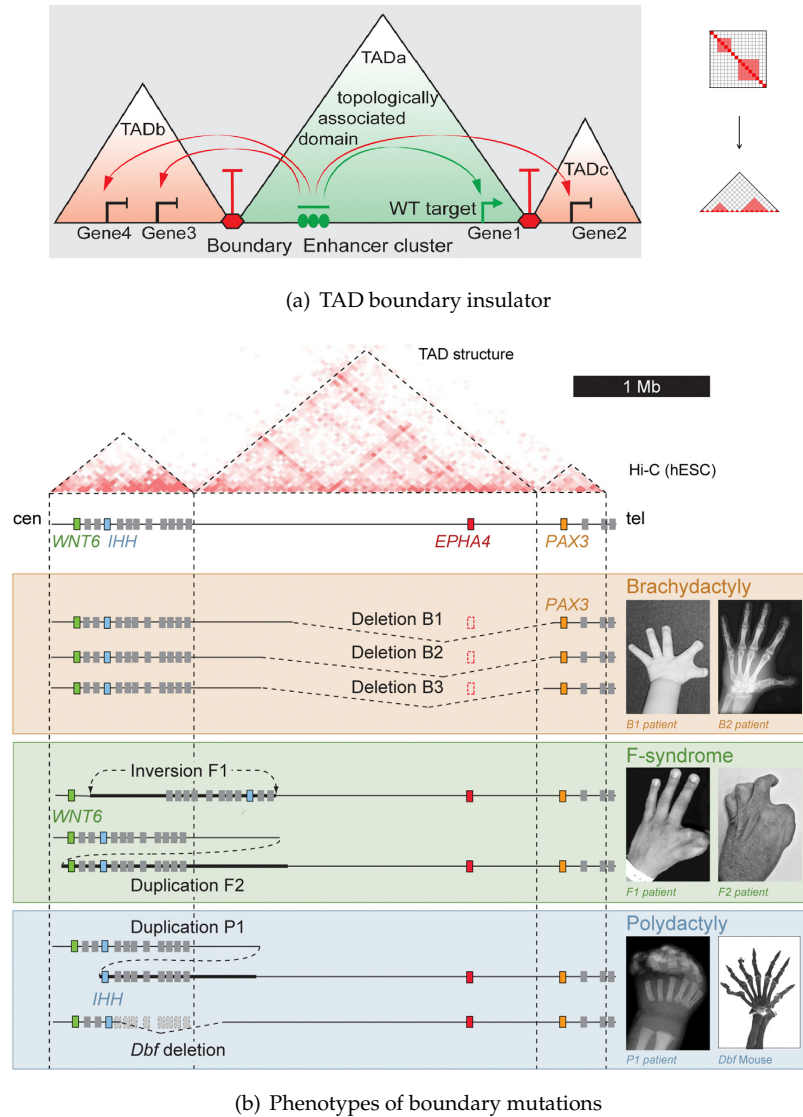
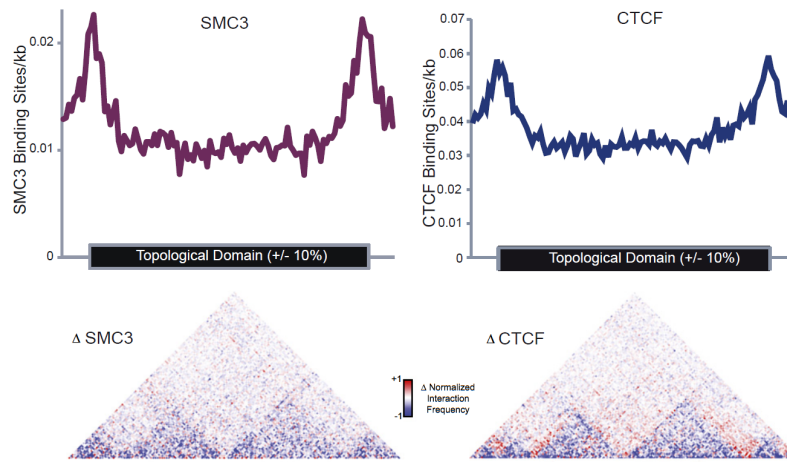
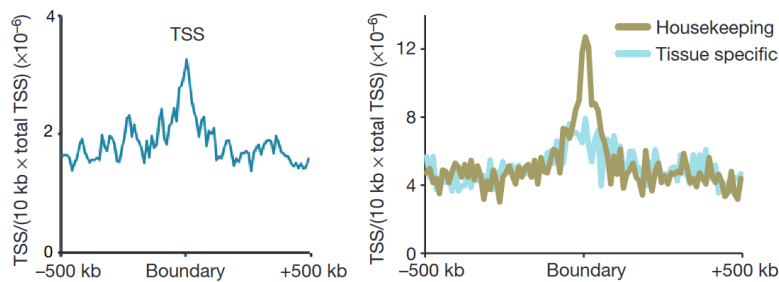


Figure 1.7: Insulation at TAD boundaries. (a) Grey box: Insulation at TAD boundaries (red) prevents distal enhancer inter-TAD interaction, where TADs are represented as triangles from one half of the mirrored Hi-C heatmap (right, from Figure 1.6(b)). (b) Disruptive rearrangements of TAD boundaries such as deletions, inversions and duplications are enough to cause polydactyl phenotypes in the mammalian *WNT6*/*EPHA4*/*PAX3* TAD architecture. Adapted from Lupiáñez et al. (2015).



(a) TADs and structural proteins



(b) TADs and transcription

Figure 1.8: Cohesin and CTCF. (a) Cohesin (Smc3 subunit) and CTCF are enriched at TAD boundaries when ChIP-Seq reads are aligned relative to TADs called *in silico* (top). Ablation of each of them causes a break down of TAD structure, shown by a decrease (blue) intra-TAD the chromatin interaction frequency (bottom left and right) and increased (red) inter-TAD interaction frequencies (bottom right). Adapted from Zuin et al. (2014). (b) Cohesin and CTCF are also co-enriched with TSSs (left) at TAD boundaries and active transcription of housekeeping genes (right). Adapted from Dixon et al. (2012).

models. As DL and SBS have failed to gain the same degree of support from both simulated and experimental data as LE in the field over the last few years this introduction will focus on LE. For a comprehensive look at DL and SBS see Bohn and Heermann (2010) and Nicodemi et al. (2008), respectively.

LE is able to simulate TADs in Hi-C contact frequency maps, as well as sub-TAD features and scale up to replicate whole chromosome territory segregation (Sanborn et al., 2015). The basis of LE is the unimpeded extrusion of the chromatin fibre through a cohesin-based extrusion complex until a collision with a TAD boundary element (BE), predominantly CTCF (Figure 1.9(a)). Collision and stalling of LE only happens if the CTCF motif is both bound by CTCF and in the correct strand and therefore orientation relative to the LE complex. Tension generated by multiple extrusions along the fibre condense chromatin from a loose random conformation to an organised extruded globule (Figure 1.9(b)).

Cohesin and CTCF have high correlation with active transcription start sites as these are also enriched towards TAD boundaries (Figure 1.8(b), Dixon et al. (2012)). It is therefore speculated that TSS and TAD boundaries should be found preferentially at the surface of TADs, accessible to TFs and nuclear components (Figure 1.9(c)). However, many questions still remain about the nature of TADs *in vivo* in single cells. What is the internal organisation of a TAD, and how is it maintained? What is the difference in organisation between TADs in active vs inactive compartments? How do TADs accommodate conflicting observations of nucleosome “clutches”? (Ricci et al., 2015), a model that claims orders of magnitude fewer nucleosomes in a TAD-size volume than would be predicted from a megabase of DNA even at extremely low nucleosomal packing densities.

All the genomics-based methods described are characterised by the need to pool sequencing reads from a large population of cells, their average maps do not output absolute 3D positions nor are they informative on spatio-temporal associations in specific cellular states (as single cell variations are

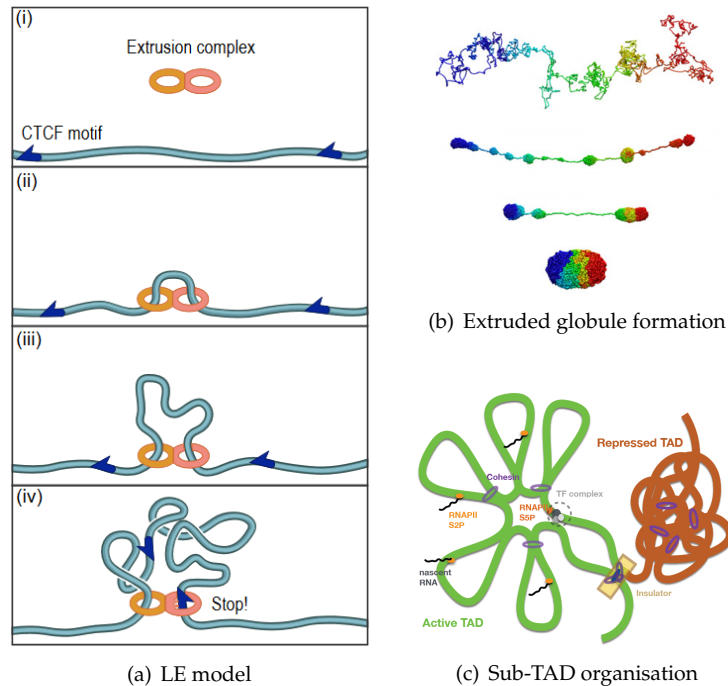


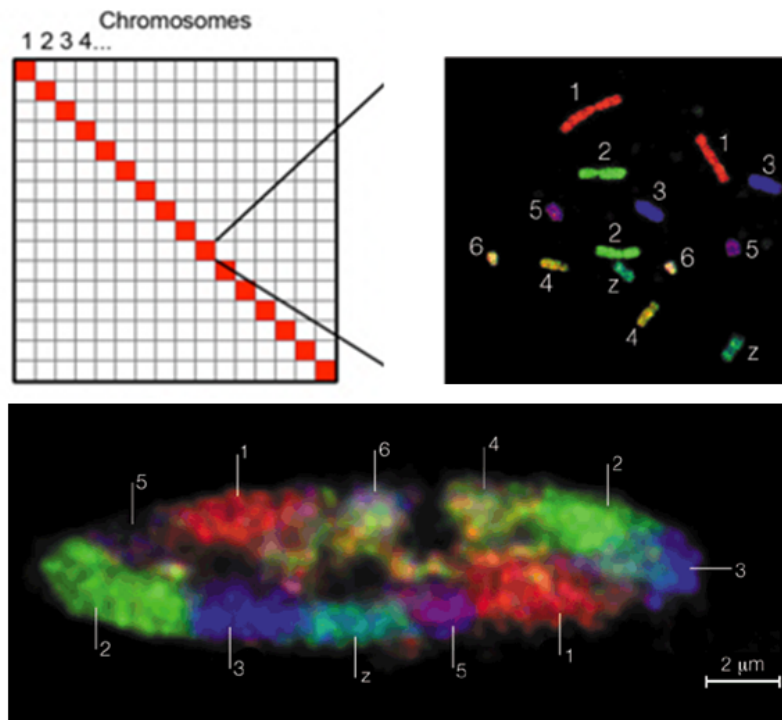
Figure 1.9: Loop Extrusion model. (a) The extrusion complex associates on the chromatin fibre and extrusion begins, pulling nearby CTCF motifs towards it (i-iii). Encountering a CTCF-bound CTCF motif in the correct orientation stalls extrusion of one flank but not of the other (iv). Note, the stoichiometry of the extrusion complex is currently unknown and may range from a single cohesin ring, a doublet (pictured) or bundles. (b) From top to bottom: Extrusion at various locations across a disorganised fibre forms localised organised loops which generate tension shortening the whole fibre. Eventually the fibre becomes organised into a TAD as an extruded globule. LE predicts a somewhat fractal sub-TAD organisation. (c) Sub-TAD organisation is currently unknown. Loops within TADs must be arranged in ways to facilitate interaction between distal regulatory elements within a TAD and whole TADs must be able to be re-organised as required by the cellular transcriptional programme into active or inactive compartments. The relationship between single TAD organisation and functional compartmentalisation (represented by nascent RNA) is still an open question in the field. The density and nucleosome packing arrangement of TADs is postulated to contribute to the overall accessibility of sub-TAD loci to nuclear proteins. Adapted from Fudenberg et al. (2016) and Sanborn et al. (2015) NonCommercial Creative Commons Attribution.

not accounted for). They must therefore be interpreted *in silico* in order to obtain models for relative chromatin architectures or orientations. Genomic based approaches therefore can provide high functional information of relative contacts but lower absolute spatial information for single cells. The best way to obtain concrete spatial relationships, is to look at them directly.

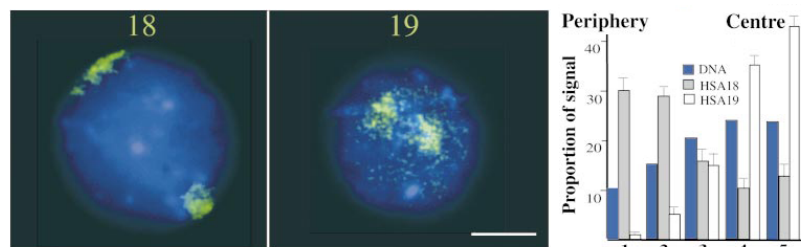
1.3 Microscopy and the genome

The supposed organisation of chromatin into distinct domains proposed by the genomic techniques is backed up by early microscopic observations. Fluorescence microscopy, indeed microscopy in general has been one of the only available methods to peer into chromatin organisation. Starting with the discovery of cells, organelles (such as the nucleus) and the description of a densely stained “chromatin” entity. Only recently have the decreased costs of genome sequencing precipitated the explosion of genomics sequencing techniques, mentioned previously.

When the entire genome is modelled with the globular nature of TADs as base subunits, chromosomes remain in segregated environments and free from knotting within and between them (Sanborn et al., 2015). This is directly confirmed by the observed presence of chromosome territories (CTs, reviewed in Cremer and Cremer (2001)) imaged with (FISH) DNA paint (Figure 1.10(a)). One of the first demonstrated relationships between gene activity to chromosomal structure was a direct consequence of the discovery of the CT organisation. When two chromosomes with distinct densities of active genes are labelled using whole-chromosome paint (Croft et al., 1999; Cremer et al., 2003; Bolzer et al., 2005) their radial locations show a clear separation. Gene-rich chromatin is preferentially found at the nuclear interior, while gene-poor chromatin at the periphery (Figure 1.10(b)). This has since been confirmed through further imaging experiments including the translocation of chromatin to ectopic chromosome regions (Matarazzo et al., 2007).



(a) Chromosome paint



(b) Radial position of CTs

Figure 1.10: Chromosome territories. (a) Chromosome paint reveals separated chromosome territories (CT) which are now also visible in Hi-C experiments as higher intra-chromosomal than inter-chromosomal contacts. Adapted from Cremer and Cremer (2001). (b) The position of these CTs are non-random. Higher gene density (chromosome 19) resulting in CTs preferentially found at the centre of the nucleus and gene poor CTs (chromosome 18) preferentially at the nuclear periphery. Adapted from (Croft et al., 1999).

Imaging of *in situ* DNA replication (observed via the incorporation of labelled or traceable nucleotide homologues) reveals a punctate pattern throughout the nucleus (Nakayasu and Berezney, 1989; O’Keefe et al., 1992; Jackson and Pombo, 1998; Schermelleh et al., 2001). Studies on the linear genome have shown that multiple replication origins fire almost simultaneously within clusters (average 4-5 origins per cluster). These clusters occupy ~0.8Mb, termed replication foci (RF) or constant timing regions (CTR). For an in depth review see Rhind and Gilbert (2013). The 3D separation of where the majority of these RFs occur throughout S phase bares striking similarities to the nature and position of CTs, LADs and NADs. During early replication low density, gene rich RF fire. These can be observed throughout the interior of the nucleus but excluding the periphery and nucleolar surfaces. During mid replication stages the opposite is true, with RF firing at the nuclear lamina and around nucleoli. Finally the RF pattern focuses on the densest chromatin, such as chromocentres in the mouse nucleus. This suggests a link between RFs, gene density and LADs/NADs.

Further evidence is supplied by bioinformatics studies which have demonstrated a strong correlation between the alignment of the genomic boundaries between early and late RFs to TADs (Pope et al., 2014). In light of the congruence of evidence from many assays the emerging picture is suggestive of a single unit of genome organisation which may be termed TAD, LAD, NAD or RF depending on the technique used. Although not all TADs, LADs, NADs and RFs have perfectly aligning boundaries it is worth pointing that all domain-calling software used in their characterisation is trained from data of millions of pooled genomic topographies, yielding an average. Cross correlative studies in single cells using multiple techniques may not be feasible currently, but if achieved may prove useful in testing such hypothesis.

Further investigation of the 3D arrangement of chromatin at the TAD scale and its relationship with DNA-interacting factors and histone PTMs is essential to understand the fundamental biochemistry behind genomic regu-

lation within the nucleus. Yet, microscopic chromatin TAD topologies remain largely unexplored by conventional fluorescence microscopy techniques as these lack the capacity to resolve such structures. The implementation of super-resolution microscopy techniques able to bypass previous resolution limits, will become instrumental for the characterisation of chromatin at these scales.

Super-resolution microscopy

The last few years have led to many useful advances in microscopy for the study of chromatin. New and improved dyes and labelling methods let us examine sub-chromosomal regions of the genome for longer, in multiplex and even *in vivo*. Even more significantly, has been the seminal development of super-resolution microscopy techniques. Fluorescence microscopy, as all microscopy based on the electromagnetic spectrum, is limited by Abbe's diffraction limit. Ernst Abbe estimated in 1873 that the limit of resolution of any microscope d could be give as a relationship between the wavelength of the propagating wave λ , and the numerical aperture NA of the objective lense. Where the NA is proportional to the refractive index of of the medium n and the half angle θ from the lens to the focal point.

$$d \approx \frac{\lambda}{2n \sin \theta} = \frac{\lambda}{2NA}$$

The NA can be pushed to 0.9/1.2/1.3/1.4 in air, water, silicon and oil, respectively, with state of the art oil-immersion optics; a maximum resolution is approximated by $\lambda/2.8$. What this means for microscopes using the visible spectrum is that the highest resolution possible is ~ 200 nm using Rayleigh's formula (or ~ 142 nm with Abbe's formula) for the smallest wavelength used (blue, ~ 400 nm) and up to 300 nm for the longer wavelengths (red, ~ 600 nm). It is important to note that this represents the maximum theoretically achievable resolution under optimal conditions. In practice, this is almost never obtained due to contrast and aberration limitations. These compounding factors be-

sides wavelength-dependent resolution affect the quality of all microscopy techniques discussed subsequently, particularly contrast for 3D SIM stripe generation.

Structures in chromatin organisation at the proposed TADs scale fall just below the diffraction limit of fluorescence microscopy. The only way to increase resolution therefore is to move away from optical microscopy based on the visible spectrum to parts of the electromagnetic spectrum of intrinsically smaller wavelengths (such as X rays), or to use electrons. While the resolution for these techniques can extend to the sub-nanometer range they suffer from limitations which restricts their use to fully explore chromatin architecture. Firstly, they are expensive and secondly sample preparation is difficult, introducing artefacts if not done correctly. Electron microscopy (EM) has the limitation that there are no dyes to generate a multiplex channel image. Instead EM relies on electron-dense materials such as heavy metal ions (e.g.: immuno-gold particles for target specific labelling) or on the differences in electron-density of the targets (such as viral capsules on cell membranes). This limits the use of experiments that can be done at this resolution assessing the spatial relationship between two proteins of interest. Furthermore EM has limitations of contrast generation as well as the preservation of structures (unless employing cryo-EM). Finally, the cost for EM instruments place them above the range of common laboratory tools, which has in turn lead to little development in the production of automation of high-throughput analysis.

Optical fluorescence microscopy requires a less expensive set up and set of reagents making it more convenient, thus in the last decade great pioneering work was undertaken to bypass the diffraction limit, allowing research of the nano architecture of chromatin to move back to fluorescence microscopy. The implications of these achievements have been profound, and rightly recognised in “The Nobel Prize in Chemistry 2014”.

Super-resolution microscopy can be achieved through several methods. An in-depth explanation of all falls beyond the scope of this introduction

but can be gained from several reviews (Schermelleh et al., 2010; Fornasiero and Rizzoli, 2014). Super-resolution microscopy can be divided in two main camps, Near-field and Far-field techniques. Near-field include atomic force microscopy and total internal reflection microscopy. Although these truly go beyond the diffraction limit (by exploiting excitations from evanescent waves) they are limited to a few tens of nanometers from the coverslip, making them incompatible with studies of the large nuclear environment. Of the Far-field techniques, three of the most known are PALM/STORM (pointillism based single molecule localisation microscopy, SMLM), STED (stimulated emission depletion) and SIM (structured illumination microscopy).

Photoactivated localization microscopy (PALM)(Betzig et al., 2006), stochastic optical reconstruction microscopy (STORM)(Rust et al., 2006), and fluorescence photoactivation localization microscopy (FPALM)(Hess et al., 2006) are the names given to a few technical set ups based on the principle of photoactivation for single-molecule detection. A single fluorophore, if bright enough, can be detected with nm precision, however, when many fluorophores are illuminated and excited simultaneously their point spread function (PSF) of emission merge. By synthesising chemical dyes able to blink in a non-continuous manner many images of single fluorophores can be acquired consecutively (or sequentially) and all fluorophores localised to very high precision. This is essentially temporal unmixing, using time as an added dimension to separate nearby fluorophores in 3D, as the probability that the same two neighbouring fluorophores being active in all images is very low. By doing so, the resolution (or full width half maxima of localisation - $FWHM_{loc}$) becomes inversely proportional to the root of the number of photons acquired (N) (Bobroff, 1986; Thompson et al., 2002).

$$FWHM_{loc} \approx \frac{FWHM_{PSF}}{\sqrt{N}}$$

Thus to achieve a high resolution (20 nm to 50 nm in $X - Y$ and 100 nm in Z) thousands of images need to be acquired, to increase the number of photons

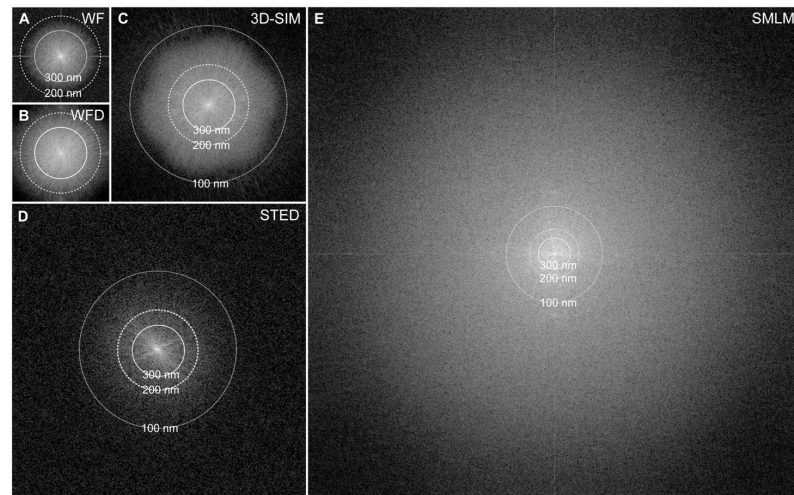
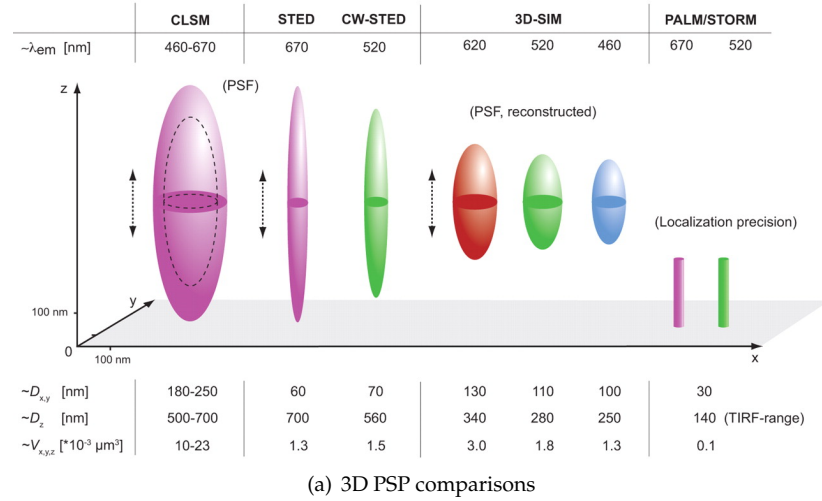


Figure 1.11: Super-resolution. (a) Schematic point spread functions (PSFs) at various emission wavelengths resolved from different super-resolution methods compared to each other and against deconvolved confocal laser scanning microscopy (CLSM). Adapted from Schermelleh et al. (2010). (b) Fast fourier transforms (FFTs) of PSFs empirically acquired for widefield (WF, A), widefield deconvolved (WFD, B), 3D-SIM (C), STED (D) and single molecule localisation microscopy (SMLM, E). Wider diameter denotes higher resolution. Adapted from Demmerle et al. (2017).

N (Figure 1.11). This increases bleaching and the length of exposure (to a few minutes) also increasing phototoxic effects. A much higher resolution can be achieved by combining this method with Near-field optics, down to a range of 10 nm to 20 nm in $X - Y$ and 10 nm to 25 nm in Z (Figure 1.11(a)), but this of course is also not applicable to the study of micrometers of chromatin beyond the coverslip.

Localisation microscopy has the highest resolution of all the super-resolution techniques and has been useful in the study of synapsis for neurodegenerative diseases in live cells Hoze et al. (2012). By far the greatest drawback of this technique is not the number of images or the dedicated software to reconstruct the super-resolved image but the availability and expense of the dyes required. Furthermore, high density of labelling can start to encroach on the limit of temporal resolution leading to artefacts whereby localisations are placed at an average distance from two or more fluorophores which could not be segmented from each other.

Stimulated depletion (STED) was first proposed in 1994 (Hell and Wichmann, 1994) and works on the principle of using a second beam to stimulate the outer ring of the excitation beam from the ground state to the first state, preventing spontaneous fluorescent emission. The effective volume of excited molecules which can emit is artificially reduced by STED, leading to a smaller PSF and higher resolution. The resolution d can therefore be controlled by the width of the STED ring, being directly proportional to its intensity, I (Westphal and Hell, 2005).

$$d \approx \frac{\lambda}{2 \sin \theta \sqrt{1 + \frac{I}{I_{sat}}}}$$

2D STED boast a typical resolution of 40 nm to 80 nm in $X - Y$ (going below 40 nm in a best case scenario). 2D STED has no improvement in resolution in Z over diffraction limited microscopy, with a resolution of 500 nm to 700 nm (Figure 1.11(a)). A 3D STED set up is available (Harke et al., 2008). This has increased complexity in the optical path but greatly improves the Z resolution

to 20 nm to 100 nm. However, in practice its often the case that power provided to the 3D STED beam parallel to the sample (which provides the 3D increase in resolution) is diverted from the perpendicular STED beam path. What this means in practice is loss in the $X - Y$ resolution to ~ 100 nm range to increase the Z resolution (Figure 1.12(a)). STED has been successfully used in resolving dendritic spines in live cell imaging (Nägerl et al., 2008).

STED, just like the single molecule localisation techniques, suffers from a restricted dye palette which must comply with both the STED and excitation beam line photochemistry, increasing the challenge of achieving highly multiplexed channel experiments. Furthermore the high complexity of optics and alignment required for 3D STED makes it less flexible to the introduction of new wavelengths. As described by the adapted Abbe limit equation, high STED laser power (I) is required to decrease the emission PSF, increasing resolution. This however is problematic for the many samples more prone to bleaching or phototoxicity. Another drawback of the system is that STED has to occur around a focal point, this means that STED systems must be inherently point scanning, making them at least an order of magnitude slower than widefield set ups such as the single molecule localisation or structured illumination systems. For whole nuclei imaging tens of minutes might be required for a STED dataset. Moreover, at high labelling-density samples (such as chromatin) the contribution of out-of-focus signal is much too high to use 2D STED alone. However the use of 3D STED as previously mentioned places the $X - Y$ resolution in the range of that achieved by SIM, with only a marginal improvement in Z (Figure 1.12(a)).

Structured illumination microscopy (SIM) relies on projecting an illumination pattern that has first been structured by passing through a periodic diffraction grating, such as many small stripes. When this illumination is combined with the sample structure of unknown periodicity a convolved Moire interference pattern is formed. The pattern is acquired, and given that the pattern for the diffraction grating is known, the unknown sample structure

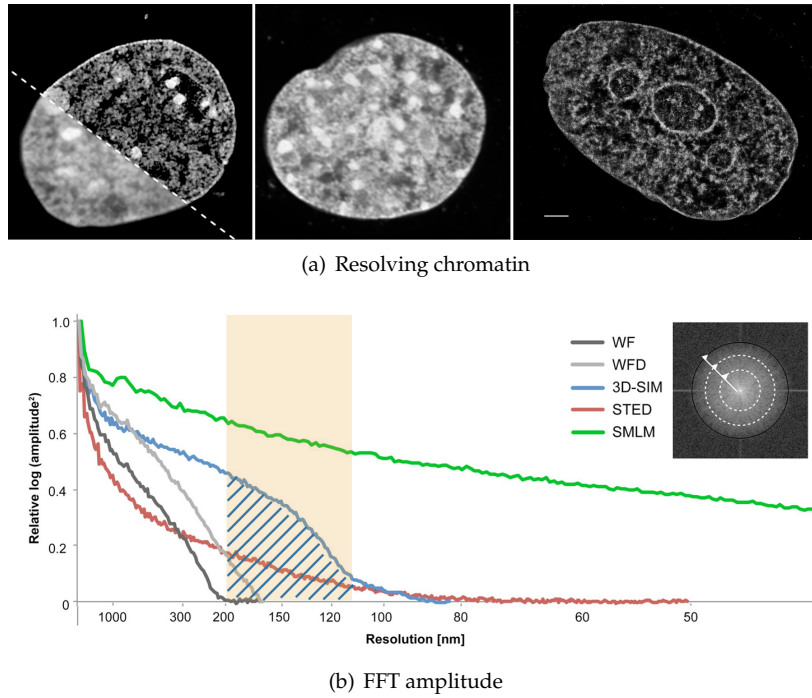


Figure 1.12: **Resolving chromatin *in situ*.** (a) Exemplary acquisitions of chromatin under different systems: 3D SIM, STED, STORM from left to right respectively. SIM is compared to widefield deconvolved in the bottom left inset. Note the loss of $X - Y$ resolution in 3D STED compared to SIM. STORM: Human fibroblast (HfB) courtesy of Melike Lakadamyali (ICFO Barcelona) (b) Range of empirical resolutions calculated as radial averages of FFTs of PSFs. SIM has a sharper decrease in resolution, but its amplitude is greater than STED in the range of size scale of chromatin macro structures (shaded area). Adapted from Demmerle et al. (2017).

can be reconstructed mathematically. SIM can also be achieved in 3D (3D SIM) with a resolution of 100 nm to 130 nm in $X - Y$ and 250 nm to 300 nm in Z (Figure 1.11) (Gustafsson et al., 2008). In terms of nominal resolution, 3D SIM is the least powerful of the available super-resolution methods and also requires dedicated reconstruction software (as does single molecule localisation). Yet, despite these limits, 3D SIM has key advantages.

The optical pathway for 3D SIM is not very complex and can be achieved from a widefield instrument. This also confers 3D SIM very high temporal resolution; with acquisition in the millisecond range 3D SIM is cable of acquiring

a whole nuclear 3D stack in under 30 seconds. This has come very useful in imaging very fast processes in cells, which don't necessarily benefit from any higher spatial resolution. This is exemplified in the recent work by the group of Nobel laureate Dr. Eric Betzig with their development of Bessel beam and subsequent lattice light sheet microscopes (Planchon et al., 2011). This development, in his words *"increases the speed of live imaging by about the same order of magnitude in the time dimension as PALM did in the space dimension"* (Beyond the Nobel Prize - New approaches to microscopy, International Society for Optics and Photonics Lecture). 3D SIM also benefits from having no restriction in the dyes needed, as its mechanism is independent of the photochemistry of the dye. Thus 3D SIM is routinely capable of multiplexing over 4 wavelengths, and more if given higher numbers of dedicated light paths.

3D SIM has brought us within reach of extracting accurate 3D and 4D topographies of genomic architecture at those Mb scales described by genomic techniques outlined previously. It can resolve the chromatin landscape not as a blurred amorphous cloud, but as an intricate network of de-condensed chromatin channels and *lacunae* pervading through and in between regions of denser chromatin packing (Figure 1.12(a))(Schermelleh et al., 2008). And, as suggested, at a reasonable cost of set up, and time for each acquisition given its resolution is on par with the other techniques at the scales of chromatin meso/macro structures (Figure 1.12(b)). Previous studies have successfully implemented 3D SIM to quantify the relative position of nascent RNA (Markaki et al., 2010) and H3K4me3/H3K27me3 in Barr bodies (Smeets et al., 2014). Moreover this super-resolved spatial information can be gained from single cells in both fixed and live-cell experiments, which can be pooled for statistical testing if needed, but not as a prerequisite.

Image analysis - an unmet need

Super-resolution techniques coupled with robust image analysis workflows and statistical algorithms enable an in-depth characterisation of the nuclear

environment. It can allow researchers to investigate the spatial relationship between key effectors of genome regulation to each other and to their target, chromatin. This in turn allows the refinement of models on the mechanisms of the central dogmas of nucleic biochemistry: Genome maintenance and repair, transcriptional regulation, replicative regulation, post-transcriptional processing, and transport across nuclear pores. However, despite its increasing availability, the widespread implementation of super-resolution microscopy to investigate genome structure-function relationships has been hampered by the shortage of robust and replicable image analysis workflows; the necessary tools for statistical analysis and hypothesis testing. Without such tools, super-resolution microscopes are useful only for generating beautiful pictures. Although this certainly has its uses for outreach of lay audiences (see cover picture for example) or for pleasing readers and editors on journal covers, it is a waste of the deep information content of most super-resolution micrographs.

The waste of valuable information is compounded by a lack of variability in the few available metrics to extract patterns from such complex spatial biological relationships. A commonly reported metric is the 2D concentric segmentation of nuclei and the radial positioning of features of interest (from full chromosomes to single gene loci) therein (Figure 1.10(b))(Mahy et al., 2002). Unfortunately, such simple measurements take little advantage of super-resolution information. In fact, 2D projections do not even make use of the full information available in 3D conventional deconvolved confocal imaging, due to the collapse of information in the third dimension. More advanced metrics capable of data extraction in n dimensions do exist in other scientific fields, and could allow a more efficient analysis of super-resolution biological datasets. Nevertheless, a general deficiency of core programming skills in the life sciences means that even if such algorithms could be sourced from disciplines like astronomy and physics, few researchers are able to integrate them as part of a robust automated pipeline (an essential requirement for the processing of terabytes of imaging data). The use of super-resolution

imaging combined with an array of automated image analysis metrics in single cells is imperative for understanding chromatin organisation, more so as many genomic studies are based on pooled populations of cells and may not be capable of distinguishing subtle structural variations between cells *in situ*.

1.4 Aims of the investigation

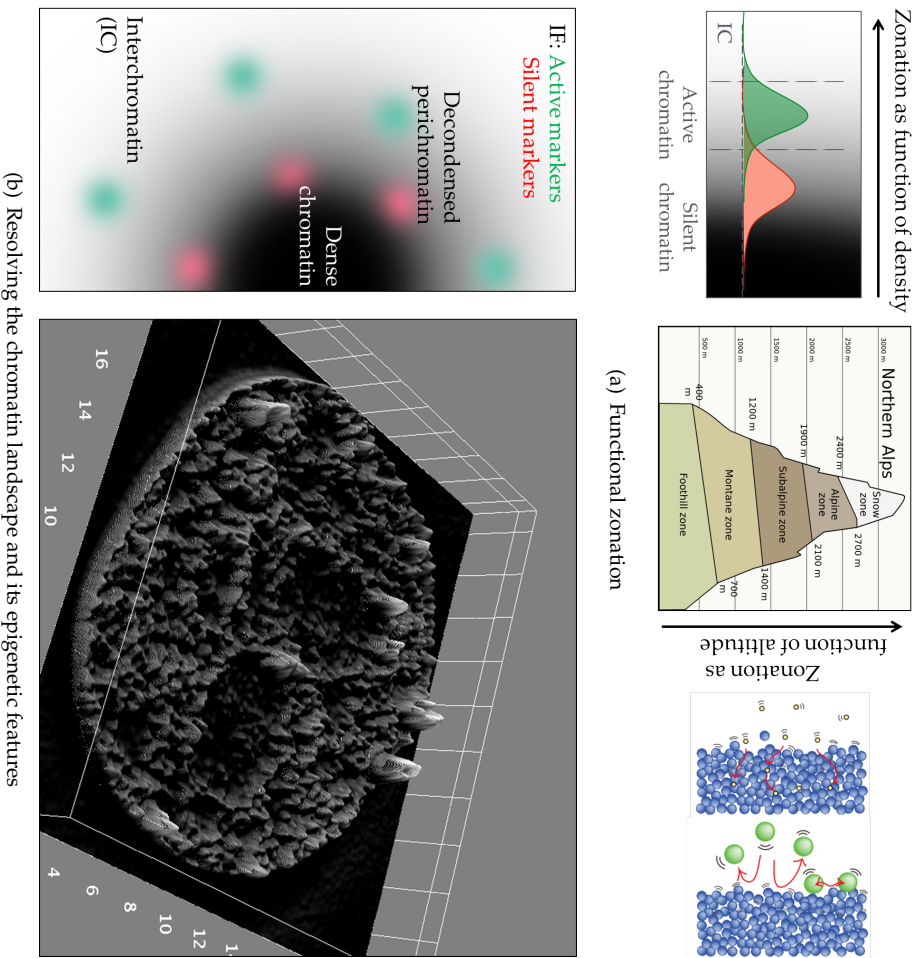
The idea postulated in this investigation is that the action of TFs and other proteins in the nuclear matrix is regulated by the accessibility of chromatin (its own density and steric crowding). Of particular interest is the question of how the density of chromatin may impose a compartmentalisation to epigenetic function, at the size scales of TADs and below (Figure 1.13(b)). This zonation as a function of density can be thought of as analogous to how vegetation zonation occurs as a function of altitude in mountain biomes (e.g.: the northern Alps, Figure 1.13(b), centre)

This view of chromatin is backed by previous *in silico* simulations. These take into account the physical size of nuclear machineries and TFs as well as a range of nucleosomal packing densities to show that chromatin in physiological conditions (modelled as a melted polymer) is a large obstacle for the free interaction of non-chromatin entities to chromatin regions hidden deep inside the polymer melt (Maeshima et al. (2015), Figure 1.13(b), right). This accessibility model likens chromatin to a viscous liquid. The interaction and attachment by large protein complexes acts in the same way as a buoy on the surface of the ocean; it prevents whatever its attached to it from sinking by keeping it afloat. For chromatin this means that the preferential interaction and ongoing transcription keeps active regions at TAD surfaces accessible to more interactions whilst simultaneously occluding inactive chromatin by molecular crowding.

To test the functional zonation predicted from this accessibility model I have developed an automated and high-throughput image analysis workflow (ChaiN , for ChaiN analysis of the *in situ* Nucleome). The ChaiN pipeline har-

bours custom implementations of morphological and geometrical algorithms in order to probe the available range of information. Immunofluorescence (IF) detection of histone modifications and key epigenetic markers in mammalian nuclei are subjected to these metrics and statistics (Figure 1.13(b), left). A spectrum of relative distributions and absolute distance of epigenetic functions to chromatin density (both in wild type and senescent cells), as well as quantitation of surface area to volume of the chromatin networks are here quantified for the first time from 3D SIM datasets. Super-resolution microscopy will be imperative (Appendix 9.1) not just to be able to resolve the individual IF signals for single epigenetic marker epitopes but as a requisite to reveal the “mountainous” landscape of chromatin density on which these epigenetic functions can be mapped (Figure 1.13(b), right).

The model refined from results of the ChaiN analysis in this study tries to reconcile previous data obtained with molecular and biochemical methods, and data from *in silico* Hi-C population ensemble studies, to classical single cell microscopy from conventional and super-resolution observations on the relationship between the complex topology of the interphase genome and many of the core biochemical functions of the nucleome environment.



Materials and Methods

Biochemical methods

Here the biochemical experimental approaches used throughout this study are discussed. Most of the work focuses on the understanding of the 3D epigenome of mammalian cells during interphase. As such the majority of protocols employed have been optimised for mammalian cells, both mouse and human. Identification of diverse epigenetic environments was achieved primarily through immunofluorescence in fixed cells, nevertheless an attempt was made to generate cell lines with stable integrations of fluorescent protein fusions for work in both global chromatin analysis, controls and single genomic loci. For example: histone H2B fusions to new fluorescent proteins or GFP-dCas9 protein fusion utilising the CRISPR system.

Markers for cell cycle stage where necessary in order to select cells at the cell cycle stage of interest. These usually required an additional immunofluorescence incubation (in the case of screening G₀ senescent stage) or replication foci labelling (for S phase stages). In the case of S phase labelling, imaging of this wavelength was conducted. This was required as the 3D epigenome at position of replication foci was a selected target in the experimental design. Markers for G₀ were only used for visual screening of the sample but not ac-

quired.

Finally this chapter explores a whole selection of methods for the labelling of single TADs and genome loci which were trialled, developed or optimised in this study.

Requirements

All chemicals, cell lines, enzymes, antibodies, culture media, buffers and laboratory equipment used in the following protocols are listed in the Appendix 9.5. Where possible, all relevant specifications and details are described, as well as product numbers and manufacturer for ease of replication.

2.1 DNA methods

Fluorescent protein fusions to target proteins of interest have been an invaluable tool for microscopy research over the last few decades (Prendergast and Mann, 1978; Dean and Palmer, 2014). Nevertheless many proteins which were once routinely used in widefield microscopy impose limitations on current work using super resolution techniques. These limits are primarily their quantum yield, excitation coefficient, brightness and photostability. Furthermore, toxicity must be taken into consideration when overexpressing constructs in cells.

GFP and the first generation fluorescent proteins (Red - RFP, Yellow - YFP, etc) suffer from low brightness and low photostability. The latter is a particular problem for use in SIM, a technique that requires repeated exposures to generate reconstructions. Because of this, many of these proteins are unsuitable for use in conjunction with SIM. In this investigation the need for SIM required use of a new generation of fluorescent proteins, e.g.: Long-Stokes-Shift (LSS) Orange, LSSmKate2, mBeRFP, mTurquoise2, mTFP1 (Shcherbakova et al., 2012; Yang et al., 2009; Goedhart et al., 2012). Many of these proteins not only have improved photophysics but have an increase in

the Stoke shift (Lackowicz, 1983). Large Stoke shift fluorophores may be excited by the same wavelength of light as other fluors but emit at a larger wavelength, allowing greater experimental design and lower photobleaching and photo-toxicity from 2 emission channels excited by a single wavelength.

One objective of this study was the construction of an exchangeable plasmid cassette for each of the aforementioned fluorescent proteins, that would allow cloning of their fusion to any protein of interest. For this a pCAG promoter plasmid vector with puromycin or neomycin resistance was cloned in conjunction with an engineered multiple cloning site (MSC) of the following restriction enzymes: XhoI-NheI-BstBI-XmaI-NotI respectively Appendix 9.5. Enzymes in the MCS must have no other restriction sequences within the protein or interest or rest of plasmid vector. The features of the final resulting cassettes followed these relative orders: pCAG-MSC-FP-Ipuro/Neo or pCAG-FP-MSC-Ipuro/Neo. The cassettes were made expressing the mKate2 fluorophore, to allow visual inspection of transfection efficiency after the correct constructs had been amplified by bacterial transformation. Transformations were performed following a modified version of the New England Biolabs (NEB) and Invitrogen transformation protocols.

Bacterial transformation

Solutions and Equipment:

- LB media
- Competent *E. coli* bacteria (DH5 α)
- SOC solution
- LB agar plates
- Antibiotic
- Pipette
- Eppendorf tubes (0.5,1.5,2 ml)
- Plasmid DNA
- Ice bath

- Heatblock
- Heated incubator
- Inoculation loop
- Ethanol
- Ethanol burner
- Heated shaker incubator
- Large centrifuge
- DNA purification kit
- Spectrophotometer (Nanodrop 2000)

Protocol

1. Thaw bacterial cells from -80°C storage on ice.
2. Mix thawed tube gently (avoid vortexing) and pipette $70\text{ }\mu\text{l}$ from stock to a new eppendorf tube. Store stock cells as quickly as possible after the required amount of bacterial solution has been dispensed.
3. Add 1 ng to 100 ng of plasmid DNA and mix gently without vortexing before keeping mixture on ice for 2 minutes.
4. Heat-shock cells at 42°C for no more than 45 seconds in a heatblock.
5. Quickly place mixture back on ice without mixing for another 2 minutes.
6. Pipette $930\text{ }\mu\text{l}$ of SOC solution kept at room temperature to the heat-shocked bacteria and DNA mix.
7. Spread $50\text{ }\mu\text{l}$ of this solution on an agar plate with appropriate antibiotic for selection of transformed cells. If the transformation efficiency is too high a second dilution of this mix can be made with more SOC, at 1:100.
8. Incubate plates over night at 37°C .
9. Once colonies are detected on plates one of these may be picked with an inoculation loop (flamed after ethanol submersion) and transferred to 50 ml LB media with $50\text{ }\mu\text{g ml}^{-1}$ antibiotic selection.
10. Incubate LB over night at 37°C on a shaking incubator. LB should turn from clear to cloudy as bacteria grow.
11. Centrifuge media to collect bacterial pellet and discard solution. For purification of DNA the Qiagen Mini/Midi and Maxi kit were used (QIAGEN Plasmid Mini, Midi, and Maxi Kits, www.qiagen.com/handbooks).

12. Measure the concentration of DNA obtained using a spectrophotometer (NanoDrop 2000/2000c Spectrophotometer). Check the $OD_{260\text{ nm}}/OD_{280\text{ nm}}$ and the $OD_{260\text{ nm}}/OD_{230\text{ nm}}$ ratio. For pure DNA these should be around 1.8 and 2.1 respectively (Interpretation of Nucleic Acid 260/280 Ratios T123 TECHNICAL BULLETIN).
13. DNA may be stored between 4 °C and –20 °C until needed.

PCR and primer design

Primers for PCR (polymerase chain reaction) were designed with PrimerQuest and OligoAnalyzer software tools from Integrated DNA Technologies (IDT). The main criterion for design of these primers is to include the relevant corresponding restriction enzyme sites as well as preserving or adding a Kozak sequence to the final construct. The Kozak consensus sequence (gcc)gccRccAUGG occurs naturally in eukaryotic mRNA and increases the efficiency of protein translation initiation (Kozak, 1987). IDT PrimerQuest and OligoAnalyzer are also useful to check for undesired formation of hairpin loops through hetero/homo-dimerisation; secondary structures which decrease hybridisation efficiency. Sequences whose maximum Gibbs free energy (more negative ΔG) for these structures is lower (lesser magnitude) from the ΔG of the linear sequence are chosen as this implies a lower stoichiometric ratio of the presence of these hairpin loops in the final reaction mixture.

The primers chosen include primers for the PCR amplification of several fluorescent proteins from their commercial plasmid as well as a primer pairs designed to dimerise with sticky ends to generate the MCS required.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Taq DNA Polymerase
- Reaction buffer for Taq polymerase
- dNTPs

- Forward and reverse primers
- Target DNA
- Nuclease-free H₂O
- PCR tubes
- PCR thermocycler

Protocol

For a 50 µl Reaction.

1. Pipette 5 µl of 10x Taq polymerase buffer into a PCR reaction tube. Final concentration: 1x.
2. Pipette 1 µl of 10 mM dNTP mixture. Final concentration 200 µM.
3. Pipette 1 µl each of 10 µM forward and reverse primers
4. Pipette 1 µl of target DNA. Final concentration 1 ng
5. Quickly take enzyme from −20 °C storage and pipette 0.25 µl of stock (or dilution to this effect) into the reaction mixture.
6. Add nuclease-free H₂O to make up to 50 µl.
7. Incubate in a PCR thermocycler. Temperatures and incubation times need to be optimised given the melting and annealing temperatures for the primers used. For the primers used here the following 35x PCR loop was used:
 - Denature DNA at 90 °C for 2 minutes.
 - Allow primer annealing by decreasing temperature to 56 °C for 1 minute
 - Increase temperature to 72 °C for polymerase enzyme to extend primers
 - Repeat cycle 35 times.
8. After PCR, the reaction may be stored at −20 °C until needed, or the product purified, e.g.: by agarose electrophoresis followed by gel extraction.

DNA digestion and ligation

Visualisation of plasmid features such as open reading frames (ORFs) and restriction enzyme sites were performed with "A Plasmid Editor" ApE software. ApE is also able to generate *in silico* digestions of the target plasmid with

known restriction enzymes from its database or custom specified restriction cut sites. Furthermore these digestions can be visualised as bands on an electrophoretic gel simulation, separated by molecular weight. This tool allows the outcome of planned digestion experiments to be known in advance, and helps with troubleshooting. DNA digestions were usually performed with two different restriction enzymes (double digest) which were chosen to generate distinct 5' and 3' sticky end overhangs, in order to avoid concatamerisation or a reversed insertion of the excised fragment of interest in subsequent DNA ligations. To achieve a successful ligation, both the fragment to be inserted as well as the vector should be digested with the same enzyme pair. Different enzyme pairs can be used as long as the overhang left by these are still homologous and allow annealing, despite a different/longer enzyme binding site. DNA digestions and ligations were performed using a NEB protocol, empirically optimised for the constructs used.

Agarose electrophoresis and gel extraction

Agarose gel electrophoresis uses electric potential difference (voltage gradient) to pull negatively charged DNA from the cathode to the anode. The agarose gel provides an obstacle by which larger molecules can be slowed down relative to smaller molecules in reaching the anode, and thus it provides a way to separate DNA by molecular weight. DNA binding dyes in the gel allow the visualisation of the DNA bands after electrophoresis, and markers containing fragments of known molecular weight help to identify and estimate the molecular weight of the DNA. Furthermore, the brightness of the band also gives semi-quantitative evidence to the relative concentrations of each fragment's band. Different agarose concentrations serve to resolve different molecular weights. Gels made from 0.5-1% agarose can resolve fragments between 1-10kb, whereas gels made from 1-2% agarose are useful to resolve fragments between 100bp-1kb.

Besides its analytical/diagnostic use post digestion/ligation/PCR,

agarose electrophoresis also allows purification of individual fragments. If fragment purification is not required, gels are loaded with enough DNA to be detectable (~50 ng), nowadays through ultraviolet filters instead of radioactivity (Evenson et al., 2012). If purification of a particular band is desired, gel wells are loaded to maximum capacity. Gel extraction and purification have been performed using Qiagen gel extraction kit.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Agarose, analytical grade
- Weighing boat
- Balance
- Duran bottle (100 ml to 250 ml)
- TAE buffer, 0.5%
- Microwave
- Ethidium bromide (10 mg ml^{-1})
- Gel boat mould
- Gel well comb mould
- Electrophoresis tank with power supply
- DNA loading dye
- UV illuminator

Protocol

For 1% agarose.

1. Weigh 500 mg of agarose using a weighing boat and sensitive balance and transfer to 100 ml to 250 ml Duran bottle.
2. Pour in 50 ml of 0.5% TAE buffer.
3. Microwave for approximately 1 minute making sure that the solution does not boil over. Note, Duran bottle lid should be left on to prevent excessive evaporation, however **it must not be screwed tight as this will cause the bottle to explode**. During heating it is recommended that the bottle be picked up (using thermal resistant gloves) and swirled to mix. Dependent on microwave power or setting used heating might take longer than 1 minute. When all agarose is dissolved, stop.

4. Leave to cool at room temperature until the bottle is approximately 60 °C; still uncomfortable to hold with bare hands for prolonged periods of time.
5. Pipette 1 µl ethyidium bromide and mix by swirling. Ethyidium bromide is highly toxic and all equipment in contact with it (as well as used gels) should be properly disposed of or cleaned.
6. Pour agarose into a pre-prepared gel boat mould, removing as many bubbles as possible.
7. Add appropriate comb mould and let gel set and cool for at least 30 minutes. Gels should not be too warm when used as this temperature in combination with the heat from the electrodes can cause re-melting.
8. Load set gel in electrophoresis tank and fill reservoirs with 0.5% TAE buffer. Make sure wells are place on side closer to the cathode.
9. Pipette appropriate volume of DNA loading dye to the aliquot of DNA to be run on gel and mix by vortexing.
10. Add DNA and molecular weight ladders to individual wells on gel in tank submerged in buffer.
11. Place tank lid and turn on power. The recommended settings are 5 V cm⁻¹ for 1 hour or 0.2 V cm⁻¹ to 0.5 V cm⁻¹ overnight.
12. Once finished take gel and view in a UV illuminator.

2.2 Cell culture

Thawing, culturing and freezing cells

Perform all steps at room temperature unless specified. Use standard tissue culture techniques in a sterile laminar flow cabinet. The work in this investigation focuses on mammalian interphase chromatin. As such mammalian cells, both from human and mouse, were used. The primary motivation for the selection of cells was their relatively low profile against the coverslip. This flatter morphology increases the quality of SIM reconstructions and reduces adverse effects in the analysis from bleaching. Details on cells used in this work can be found in Appendix 9.5 and Table 9.6.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Culture media (DMEM, F-12, glutamax etc..)
- Cryogenic storage
- Heated water bath
- Trypsin solution
- Nunclon tissue culture dishes (10 ml)
- DMSO
- Falcon tube
- Ethanol
- Microcentrifuge
- Pump
- Tissue culture incubator

Protocol

1. Take cells from liquid nitrogen cryogenic storage (approx. -180°C) and warm up quickly to tissue culture temperature by submersion in heated water bath (37°C).
2. As soon as cell solution has thawed, remove cryogenic tube from water bath, sterilise with ethanol (70%) and pipette contents to 15 ml Falcon tube.
3. Immediately add ~10 ml of pre-warmed culture media. Cells should not be left thawed in the concentrated cryogenic solution for longer than required as the cryogenic solution may be cytotoxic when liquid.
4. Centrifuge solution gently at $300g$ for 3-5 minutes. Care must be taken not to disrupt cellular membranes by harsh centrifugation; rcf and length of centrifugation should be tailored for each cell type used.
5. After centrifugation aspirate dilute cryogenic medium with pump being careful not to disturb the cell-containing pellet.
6. Pipette 2 ml of pre-warmed culture media into tube and resuspend pellet by gentle agitation. Do not vortex or resuspend by pipetting as the shear stresses may rupture cells.
7. Once resuspended, add 8 ml more of pre-warmed culture media, mix gently by inverting tube and finally transfer cell suspension to a 10 ml tissue culture dish.

8. Cells should be cultured between 50-80% dish confluency in a humidified tissue culture incubator at 37 °C and 5% Carbon dioxide atmosphere.
9. When cells cover over 80% of the dish surface they may be “split” by aspiration of culture media and addition of 2 ml Trypsin solution.
10. Incubate from 2-5 minutes with Trypsin in the same incubator
11. Tap side of dish until cell monolayer has detached (this is clearly detectable by a change in the opacity of the dish bottom, however a low magnification microscope with phase contrast can be used to check if uncertain).
12. Add 8 ml of culture media to quench the trypsinisation and mix by gentle pipetting.
13. Aspirate required volume to discard for the splitting ratio desired (i.e: leave 1 ml or 2 ml of cell suspension for a 1:10 or 1:5 split respectively).
14. Top up with culture media up to 10 ml and place dish back in tissue culture incubator. Splitting can be repeated as often as required, dependent on the growing rates of the cell line in question.
15. For moving cells to cryogenic storage wait until cells are over 80% confluent and trypsinise as described previously.
16. Add 8 ml of culture media to quench after Trypsin incubation but do not aspirate.
17. Centrifuge for 5 minutes at 300g to generate a cell pellet and aspirate media solution.
18. Resuspend cells in 10 ml of culture media made up with 10% DMSO.
19. Aliquot into cryotubes and place on ice. Cool down tubes gradually by subsequent storage at –20 °C and –80 °C before finally storing permanently in liquid nitrogen at –180 °C.

Seeding cells and cell culture synchronisation

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Ethanol
- Hydrochloric acid
- H₂O, demineralised

- Nunclone 6 well dish
- High precision cover slips
- Stainless steel forceps
- PBS solution
- Thymidine (optional)

Protocol

1. Clean high precision cover slips in 1% hydrochloric acid bath for 1 hour followed by wash in demineralised H₂O and storage in ethanol absolute.
2. Do not flame sterilise cover slips when used for seeding cells for subsequent super resolution microscopy as this can warp the glass.
3. Use forceps to place a single cover slip at the bottom of each well to be used.
4. After growing and trypsinising cells as described previously, aliquot a volume of cell suspension that will give the desired splitting ratio when topped up with 2 ml of culture media. eg: 200 µl of cell suspension in 1.8 ml of culture media to achieve a 1:10 ratio. If cells must be synchronised seed at 20-30% confluency.
5. Add pre-warmed culture media to each well and agitate gently to spread cell suspension evenly over the cover slip.
6. Incubate in a humidified, temperature and carbon dioxide controlled incubator.
7. If cell culture must be synchronised use double thymidine block as described below: (Times here are for HeLa and C127 cells but may need to be optimised dependent on the growth rate of other cell lines used)
 - Add thymidine to 2ml of culture media. 50 µl-100 µl of 100 mM to final concentration of 2 mM-4 mM respectively.
 - Incubate cells for 18-20 hours in tissue culture incubator
 - Release block by washing off thymidine by aspiration and addition of 1x PBS solution. Aspirate PBS and add 2 ml culture media.
 - Incubate cells for 10-12 hours
 - Repeat thymidine block
 - Incubate cells for 18-20 hours.
 - Release second block by washing as before. The majority of the cell population should be at the G1-S phase transition ready for replication labelling experiments.
8. When cells reach desired confluency (generally 80-90% for imaging) progress to fixation and staining.

9. *Optional for perturbation of chromatin morphology before immunostaining. For chromatin decompaction add 600 nM of TSA to growth media or dilute growth media in 1 ml of dH₂O (to expose cells to hypo-osmotic conditions) per seeded well and incubate for 10 minutes before fixation. For chromatin compaction add 50 mM of NaN₃ or 200 µl of 10xPBS (to expose cells to hyper-osmotic conditions) to growth medium per seeded well and incubate for 10 minutes before fixation.*

Transient and stable cell line transfections

Transfection of mammalian cells was used to generate cell lines expressing a particular construct. In many cases these were only required transiently. However due to overexpression effects and other experimental limitations, stable cell lines were also generated. The protocol described below is an overview of the most general steps in transfection procedures. Nevertheless, individual commercial protocols can be used and if so, their unique protocol should be followed. In this investigation the following transfection reagents were used with their protocols: Lipofectamine 2000 (Invitrogen), Lipofectamine RNAi-MAX (Invitrogen), Fuse-It (Ibidi) and TransIT-X2 (Mirus, Caribou Biosciences).

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Seeded cells
- Commercial transfection reagents
- Sonicator (optional)
- DNA/RNA to transfect
- PBS
- Culture media (DMEM, F-12, glutamax etc..)
- Selection drugs (neomycin and puromycin)

Protocol

1. Seed cells on high-precision cover slips to ~80% confluency.
2. Mix Transfection reagents in quantities and concentrations as outlined by commercial protocols (generally the transfection reagent is diluted with reduced serum culture media, eg: Optimem).
3. Dilute DNA/RNA to concentrations required by the transfection reagent.
4. Add diluted DNA/RNA to transfection reagent mixture at a ratio outlined by commercial protocol.
5. Incubate mixtures to allow formation of transfection complexes (generally cationic lipid coating surrounding construct to transfect). Some protocols may also require sonication for thorough mix of these two components.
6. After incubation add mixture to cells at final concentration stated by individual protocols (generally ~100 μ l and ~1 μ g of DNA/ RNA per 1 ml of cells in tissue culture media).
7. Incubate cells with transfection complexes from 30 minutes to 2 hours.
8. Wash off media with transfection complexes with 1x PBS and replace with pre-warmed culture media.
9. Incubate cells after removal of transfection complexes over a period of at least 24 hours before continuing to visualisation (for transient transfection) or selection (for stable transfections).
10. For transient transfections 1-2 days of growth is sufficient, splitting may be required if initial confluency was high and cell growth rate was not significantly affected by transfection incubation.
11. For stable transfections after 1-2 days cells should be split and grown in normal culture medium with addition of drug for selection. In this investigation neomycin and puromycin were tested. The concentration of drug used is dependent on cell type. Here 1 μ g ml⁻¹ and 3 μ g ml⁻¹ of puromycin for RPE1 and C127 cell lines respectively.
12. As cells without stable integration of the construct conferring the drug resistance are killed the population of cells decreases significantly. After continued incubation with selection drug for a few days/weeks the cell population should now include exclusively those with stable integration of construct and can be split normally.
13. Dependent on the biochemical load that the expression of construct may have on cells, the growth rate of stably transfected cell lines may be slower than pre-transfection controls.
14. Stable cells may be grown without selection drug after a month of selection, but it is recommended that selection should be repeated periodically as cells without construct may still be present, or the construct lost from cells with previous integration.

15. Stable cell culture are still polyclonal and represent integration of construct at diverse genomic regions. These may lead to varied gene expression of the desired target. Single cells may be harvested after dilution of a trypsin spit and cultured to generate monoclonal cultures of homogeneous expression.

2.3 Cellular staining

This investigation was based on analysis from super-resolution fluorescence microscopy, which inherently relies on efficient labelling of structures and epitopes of interests within cells. Furthermore data obtained from super resolution micrographs has increased sensitivity to artefacts induced by sample preparation. As such, care must be taken when following the protocols described in this section. Details of the optical set up and fluorescence light paths used in this investigation can be found in Appendix 9.3. An in depth discussion on sample preparation for 3D SIM can be found in Kraus et al. (2017)

Cell fixation

Solutions and Equipment:

- Pipette
- Seeded cells
- Beaker (50 ml to 100 ml)
- Stainless steel fine tweezers
- Soft absorbent tissue
- 6-well plate
- Formaldehyde
- PBS
- PBST
- Desktop pump

Protocol

1. Cells should be seeded on high precision cover slips as described in §2.2 to the confluency required for the planned experiment.
2. Fill small glass beaker with 50 ml to 100 ml of 1x PBS.
3. Fold soft tissue lengthwise and place it on the bench next to the beaker.
4. For each coverslip fill (2 ml) of formaldehyde fixation solution (2% in 1x PBS) in any well of a new 6-well plate (alternatively 35 mm dish may be used).
5. Carefully pick up coverslip by its edge with fine tweezers; gently dab the side of coverslip on the tissue to remove excess medium and wash shortly by dipping 2–3 times in PBS.
6. Dab again and immediately transfer to the well with fixation solution. Washes can be performed by aspiration and immediate replacement of solution in the same well repeatedly; nevertheless the use of a dipping beaker is recommended as the excess volume of PBS (>50 ml) is more effective than that of a single well (2 ml to 5 ml).
7. Agitate gently by hand; then incubate for 10 min with closed lid under a hood at room temperature.
8. Aspirate solution with bench-top pump and simultaneously refill well with 2 ml to 5 ml of PBST avoiding drying of sample from formaldehyde evaporation (rapid evaporation will cause fixation artefacts apparent as shrivelled and creased nuclear outlines).
9. Repeat 2–3 times until fixation solution is fully exchanged with washing buffer. Samples can be stored in PBST overnight at 4 °C after fixation if required.

Post-fixation staining: immunofluorescence and counterstain

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5, 1.5, 2 ml)
- PBST
- Stainless steel fine tweezers
- Triton X-100
- Parafilm
- Maxblock

- RNaseI
- Humidified dark chamber
- Primary and secondary antibody
- Counterstain (DAPI, SYTOX)
- H₂O, demineralised
- Mounting media (Glycerol, Glycerol-DABCO, Vectashield, Prolong-gold/diamond, etc.)
- Nail varnish (clear)

Protocol

1. Follow protocol in §2.3.
2. Exchange PBST with 2 ml of 0.2% Triton X-100 in PBS. Agitate shortly, and then incubate for 10 min. Exchange with PBST.
3. Cut Parafilm to 6-well format and press it, to create a flat coating, on the top of the 6-well lid to be placed in a humidified dark chamber.
4. For each coverslip pipette 100 µl of blocking solution (Max Block), according to the well position onto the film. Pick up coverslip, dab on tissue and place on the drop with cell side down.
5. Incubate in a humidified dark chamber for 30 min.
6. Wash each coverslip carefully by repeated gentle dips in PBST solution and dabbing on tissue as described previously. A secondary pair of tweezers as backstop supports for careful lifting of coverslips from Parafilm allows for controlled lifting of the coverslip to avoid shearing cells from coverslips due to capillary surface tension created between the sample and the Parafilm.
7. Prepare a Parafilm layer as before and pipette a drop of primary antibody at the required concentration. Each antibody has an optimal concentration which can be derived empirically or sometimes from the datasheet provided by the supplier. Often this is around 1:500-1:1000 dilution in the same blocking media or in a commercial antibody-dilution mixture.
8. Place cells on antibody drops and incubate in a humidified dark chamber for 1 hour.
9. Wash in PBST as described, and prepare another Parafilm layer with drops of secondary antibody (fluorophore conjugated for detection).
10. Place cells on antibody drops and incubate in a humidified dark chamber for 1 hour.
11. Wash in PBST as described.

12. After secondary antibody incubation perform a post-fixation with 4 % formaldehyde/PBS
13. Wash in PBST
14. For counterstaining, add 100 μ l of DAPI solution (2 μ g/ml) or SYTOX Green solution (1 μ M) on Parafilm on lid and incubate in a humidified dark chamber for 10 min. Unlike DAPI, SYTOX Green does not have a bias towards binding AT-rich sequences (Appendix 9.1). However, it has a weak affinity to bind RNA at high concentrations. To avoid residual RNA binding, an incubation with 1–10 U RNaseI/PBS at 37 °C for at least 30 min is recommended prior to counterstaining (1 U RNaseI will degrade 100 ng of RNA per second in optimal condition).
15. Wash and dab coverslip as described previously in PBST as described above followed by a final wash in demin water to remove salts.
16. Immediately place on a 30 μ l drop of mounting medium on Parafilm for 2 min. The addition of this step between the last wash and mounting coverslips on microscopy slides ensures an excess of mounting medium over water in the final sealed volume of the sample. This equilibration step essentially reduces the dilution of mounting media and ensures the desired refractive index is maintained in the final sample for imaging.
17. Add 10 μ l of mounting medium in the centre of a pre-cleaned microscope slide. If using microscope slides with frosted coating for labelling it is recommended to mount on unfrosted side as frosting can create a small but noticeable tilt of the sample when mounted on the stage, leading to large focus point variations at opposing ends of the coverslip.
18. Pick up coverslip, dab excess dilute mounting medium on tissue and mount carefully onto the drop of mounting medium on the slide. Let the mounting medium settle for 2 min.
19. Excess mounting medium should be carefully removed without moving the coverslips. We recommend the use of fine tissue (or a single layer of double layered Kimwipes) placed evenly over the coverslip and removed vertically after absorption. This may need to be repeated once or twice with the application of soft pressure with the fingertip along the edges of the coverslip to ensure that all excess mounting medium has been absorbed before proceeding to sealing samples. Thorough removal of excess mounting medium will prevent the coverslip from moving during the sealing procedure.
20. When all excess has been removed, seal edges with appropriate amount of nail varnish. A small amount of nail polish applied with a single stroke along each side of the coverslip is usually sufficient for sealing, and can be repeated after the first coat has dried if necessary.
21. Sealed slides can be stored in slide-boxes at 4 °C.
22. Coverslips should be cleaned from residual medium and, after imaging, from immersion oil using 80% ethanol/H₂O demin. and fine tissue.

Directly before imaging, chloroform dipped cotton swabs may be used to remove residual dirt off the coverslip.

23. Note that repeated cleaning with ethanol or chloroform may dissolve nail polish. In this case old nail varnish may be peeled off using fine tweezers and new nail varnished reapplied.

2.4 Replication-based foci labelling

Here are described various replication (pulse) labelling methodologies compatible with 3D-SIM imaging for the subsequent segmentation of individual chromosome territories or replication domains in both live cell experiments and fixed mammalian cell samples. This is achieved through the incorporation of fluorescent or traceable thymidine analogues via nascent replication and fine-tuning of labelling pulse lengths or chase timing before imaging. Combinations of different labels at separate times allows the visualisation of the sequential nature of chromatin replication in 3D space. Moreover, simultaneous IF protocols targeting transcription factors, chromatin binding proteins and histone PTMs can help elucidate how active and repressive chromatin is organized and maintained in whole chromosomes or TADs. Whilst this protocol is concerned with optimization for super-resolution imaging, many steps also have general applicability to wide-field, confocal, and other fluorescent imaging methods.

This protocol can be applied to an asynchronous cell population. However synchronization is recommended in order to maximize labelling efficiency. The duration of incubation with marker depends on the S phase length of the respective cell line as some cells may experience some lag from block release.

Whole replicated genome and single chromosome territory stain

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)

- Seeded cells
- F-ara-EdU
- Culture media
- Buffers PBS and PBST
- Fixation solution
- Triton X-100
- PBS
- PBST
- Parafilm
- Humidified dark chamber
- H₂O, demin
- Tris-HCl buffer
- Sodium ascorbate
- Copper sulphate
- Fluorescent dye, azide tagged

Protocol

1. Cells should be seeded on high precision cover slips in 2 ml wells as described in §2.2 to the confluency required for the planned experiment and taking into account the time required for synchronisation.
2. Add F-ara-EdU to final concentration of 10 μ M. For labelling entire chromosomes incubate for at least 10 h to cover the length of an entire S-phase. Given the reduced toxicity of the F-ara-EdU compared to EdU, it is possible to incubate cells for over 12 h or overnight to ensure maximum coverage. The F-ara-EdU concentration may be reduced to 1 μ M if overnight incubation is required to further minimize adverse effects on particularly sensitive cell lines.
3. If only bulk chromatin labelling is required then fixation can be done at this point as described in §2.3 as this method is a suitable alternative for counterstaining with standard chromatin intercalating dyes.
4. If single chromosome territories are required then wash cells in 1xPBS and exchange with label-free media.
5. Incubate for at least 4 days to allow segregation of labelled and unlabelled chromosomes. As with labelling pulse length, chase length is dependent on the cell cycle of the respective cell line. Given the high label coverage throughout the genome, cytotoxic long-term effects cannot be ruled out. A chase time of four post-labelling cell division can be sufficient for segregation of individual territories in a per-cell basis.

6. Cells should be fixed as described in §2.3 leaving cells in 1xPBST after fixation.
7. Fill small glass beaker with 50–100 ml of 1x PBS.
8. Fold soft tissue lengthwise and place it on the bench next to the beaker.
9. Exchange PBST with 2 ml of 0.2% Triton X-100 in PBS. Agitate shortly, and then incubate for 10 min. Exchange with PBST.
10. Cut Parafilm to 6-well format and press it, to create a flat coating, on the top of the 6-well lid to be placed in a humidified dark chamber.
11. *Optional for additional immunostaining after click reaction: for each coverslip pipette 100 µl of blocking solution. Dab on tissue and place on the drop with cell side down. Incubate in a humidified dark chamber for 30 min*
12. Make Click reaction solution (per 100 µl): 55 µl demin H₂O, 10 µl Tris-HCl buffer (1 M, pH 7), 10 µl sodium ascorbate (500 mM), 5 µl copper sulphate (100 mM), 20 µl fluorescent azide dye (0.1 mM). It is recommended making this fresh every time. Stock solutions for Tris-HCl buffer, copper sulfate and fluorescent azide dyes can be kept for prolonged periods at 4 °C, However sodium ascorbate is a strong reducing agent and should optimally be made fresh every time. If stocks of sodium ascorbate are to be made it is recommended to storing it only for up to one month at 4 °C. Visual inspection should confirm deterioration from a light translucent yellow to a deep red amber solution.
13. Pipette a 100 µl drop of the click reaction mix for each coverslip onto Parafilm applied on a 6-well lid as described previously. Pick up coverslip with the tweezers, dab on tissue and place on the drop with cell side down. Incubate in a humidified dark chamber for 30 min.
14. Wash away click reaction by dipping coverslips in beakers of 1x PBST dabbing on soft tissue in between dips as described previously.
15. For a combined Click/immunofluorescence labelling it is possible to incorporate a primary and secondary antibody incubation protocol at this point, to label desired targets and their relative position to replicated chromatin. Follow the protocol described in §2.3. Note that the Click reaction can be detrimental to imaging GFP protein fusions. Kits with optimized buffer condition to avoid this are commercially available; alternatively one can use anti-GFP antibodies or nanobodies (GFP- booster) in a combined Click/immunofluorescence labelling protocol.
16. For optional counterstaining follow the protocol described in §2.3.
17. Wash in PBST followed by demineralised water and mount with mounting media before sealing sample as described in previous sections.

Fixed single replication foci labelling (and alignment slide for SIM)

Progression of DNA synthesis through S phase occurs in distinct stages. Early S phase short labelling pulse produces a punctate pattern throughout the nucleus, mid S phase labelling shows a dotted ring at the nuclear and nucleoli periphery, and late S phase pulses label regions of constitutive heterochromatin (such as mouse chromocenters). Incubation times for labelling can extend to 20–30 min and punctate pattern will still be conserved, nevertheless individual spots may contain multiple replication domains. Incubation should not last for less than 5 min, as this is the approximate lag time for entry of the dye molecule into cells. Incubations shorter than 5 min may label smaller genomic regions, but can compromise efficiency of label incorporation and density. EdU or F-ara-EdU can be used, as the absolute incorporation is far lower than for whole chromosome territories, reducing the chance of observing cytotoxic effects.

By amending the following protocol using multi-colour detection of replication foci a 3D calibration samples for alignment of 3D-SIM datasets can easily be generated. This part of my work is now published in Kraus et al. (2017). Multichannel 3D-SIM data, particularly if acquired from multiple cameras, requires post-acquisition colour channel alignment. Alignment slides for this calibration can be made by fixing multispectral beads on the plane of the coverslip surface or a 2D array of milled perforations on a mirror-mounted slide. For 2D datasets these slides can correct X-Y translations, rotations and zoom variability between detectors, but they fail to address the depth-dependent z-shift caused by chromatic dispersion present in multichannel 3D datasets. Some software packages avoid this by splitting alignments in two phases; first the automated 2D registration using 2D alignment slides, followed by a manual correction in z relying on visual inspection to generate Z-axis calibration parameters (Bratton and Shaevitz, 2015). For preparation of dual-layer bead slides for 3D-channel alignment see Demmerle et al. (2017). In this study 3D alignment slides were first made from beads and

later from EdU incorporation in replicating mammalian cells, followed by a “click” reaction to couple EdU residues to multiple azide-functionalized fluorophores. These slides have the same 3D nuclear foci pattern in all channels, instead of foci on a single z plane (occurring in beads). When coupled to open-source software capable of iterative 3D registration (e.g. Chromagnon: <https://github.com/macronucleus/chromagnon>, or similar software (Křížek et al., 2016), these 3D patterns allow the generation of alignment parameters in all spatial dimensions simultaneously. In this way, manual alignment is dispensable, increasing the efficiency and reproducibility of this step. The parameters can then be applied to any other datasets acquired under comparable conditions in a batch-processing mode.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Seeded cells
- EdU or F-ara-EdU
- Culture media
- Buffers PBS and PBST
- Fixation solution
- Triton X-100
- PBS
- PBST
- Parafilm
- Humidified dark chamber
- H₂O, demin
- Tris–HCl buffer
- Sodium ascorbate
- Copper sulphate
- Fluorescent dye, azide tagged

Protocol

1. Cells should be seeded on high precision cover slips as described in §2.2 section to the confluency required for the planned experiment and taking into account the time required for synchronisation.
2. If culture is synchronised wait until desired S phase stage can be labelled then incubate for 5 min with addition of 10 μ M EdU or of F-ara-EdU. Due to the short labelling pulse and chase the increased level of cytotoxicity from EdU as opposed to F-ara-EdU is not significant. Furthermore BrdU can also be used followed by anti-BrdU immunodetection as opposed to click reaction.
3. If culture is not synchronised then approximately 1/3 of cells in the sample should have undergone S phase during the labelling pulse. of this third there will be cells in all stages of S phase.
4. Exchange with label-free media
5. Cells should be fixed as described in §2.3 section.
6. Follow protocol in §2.4 after fixation step.
7. For generation of 3D calibration slide the click reaction is modified to have multiple fluorophores. The ratio of these must be empirically defined to compensate for rate of incorporation and brightness of fluorophore. In this study a simultaneous 3-colour click reaction at a 7:2:1 ratio of azide-tagged CF405M (blue), Alexa Fluor 594 (red) and Alexa Fluor 488 (green), was used respectively.
8. For acquisitions of 3D calibration slide, cells whose foci pattern is of mid-late S phase are recommended (Figure 3.2) as these foci are distinct, high-contrasted and distributed evenly throughout the entire nuclear periphery and around nucleoli. Early S phase patterns are still punctate but tend to result in a more diffuse, low-contrast signal throughout the nuclear volume, making them a more challenging sample for image registration.

Replication labelling for *in vivo* imaging: scratch labelling and nucleofection

Compatibility of labelling with cell permeable and impermeable dyes introduces temporal resolution for determining replication directionality in super-resolution imaging.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Nunclon, 6 cm dish
- Microlance needle
- Fluorescent conjugated dUTPs
- 4D Nucleofector (Lonza)
- 4D Nucleofection kit

Protocol

1. Cells should be seeded on high precision cover slips as described in §2.2 to the confluency required for the planned experiment and taking into account the time required for synchronisation.
2. If culture is synchronised wait until desired S phase stage can be labelled.
3. For scratch labelling, transfer coverslip to a 6 cm dish, and immediately cover cells (to avoid drying) with 20 μ l of fluorescent conjugated dUTPs (10 μ M to 20 μ M).
4. Scratch surface of coverslip thoroughly with light pressure using a Microlance needle. The scratching procedure leads to transient permeabilization of the membranes of cells along the scratch wound site due to mechanical shearing, allowing the uptake of cell impermeable dyes for a few seconds (Schermelleh et al., 2001).
5. Wait 2 min and return coverslip to media-filled well
6. Another method to deliver cell-impermeable dyes at higher efficiency is based on electroporation and requires the commercially available 4D Nucleofector (Lonza). Nucleofector and nucleofection kit buffers optimized for multiple cell lines. This requires harvesting cells in suspension or in a 24-well plate format, mixing with nucleofection buffer and 1 μ l of fluorescently labelled dUTP (for 20 μ l reaction mix), and selecting a shock program suitable for the chosen cell line. This procedure has the advantage of delivering nucleotides to the majority of the cell population, although incorporated only by cells currently in S phase. The procedure is typically more efficient (providing a more reproducible and higher fraction of labelled cells) than scratch labelling, which only affects cells along the scratch wound site. Over 85% of cells in S phase within an unsynchronized population can be efficiently labelled using this approach compared to EdU control.

2.5 TAD labelling methods

The denaturation step required for traditional FISH labelling methods alters chromatin ultrastructure. Furthermore it also denatures the proteome disrupting epitopes for targets *in situ*. Because of these restrictions traditional FISH has limited usability for single cell studies of the 3D epigenome in topologically associated domain (TAD) organisation with 3D-SIM. A flexible, non-denaturing labelling tool for super-resolution imaging of single loci, compatible with live cell and multi-colour experiments, would be needed to help answer some of the most pressing questions of chromosomal biology dogma today.

Dam-ID based LAD labelling

Dam-ID is a two component prokaryotic system that can discriminately label chromatin lamina associated domains (LADs) with the methyltransferase activity of the Dam enzyme after transient contact (Greil et al., 2006). A second construct able to detect the ectopic methylation (m6A) can be fused to GFP; this is known as the m6A tracer (Kind et al. (2013), Kind and van Steensel (2010)). In this study this combination is referred to as LAD-ID. The requirements are the Dam methylase as a fusion construct with a lamina component (Dam-LMNB1) and the methyl tracer (Dpn7) fused to GFP (eGFP-Dpn7). The LAD-ID system has an double inducible expression system requiring the addition of Shield to make Dam-LMNB1 stable, and the removal of doxycycline from the growth media for the production of the m6A tracer (Figure 7.2). By culturing without doxycycline the GFP based tracer is produced. However this only amounts to dim background signal. By addition of Shield for a specific period of time (e.g.: 30min) the Dam-LMNB1 construct is stabilised and can methylate DNA until Shield is removed and the construct degraded. By varying the length of Shield incubation the time-span of LAD-ID labelling can be regulated within a single G₁ stage of the cell cycle.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Seeded cells
- Culture media
- PBS
- Construct for transfection
- Doxycycline
- Shield

Protocol

1. Cells should be seeded on high precision cover slips as described in §2.2 to the confluency required for the planned experiment and taking into account the time required for synchronisation.
2. Transfections of these constructs were made following the protocol outlined in §2.2. Generally a good transfection efficiency can be achieved in C127 cells with these constructs (pTuner-DD-Dam-LMNB1, pL-EcoDAM-hLMNB1 and PLgw-CMV-Dam-LMNB1 plasmids). In half of the cells a rim is normally visible 24 hours after transfection (if continually induced).
3. Grow cells in optimum tissue culture media with addition of doxycycline as described in Kind et al. (2013).
4. Start live imaging or fix the cells within 24 hours after transient transfection.
5. Induce the system by washing cells in 1x PBS and exchanging with doxycycline-free medium before adding Shield at 500 nM final concentration.
6. A bright, easily detectable signal is obtained after induction from co-transfections. eGFP-Dpn7 single transfection gives poor signal. Probably due to higher protein stability of the eGFP-Dpn7 when bound to m6A-DNA.
7. Lamin B1 over-expression in transient transfected cells can lead to nuclear morphology phenotypes. Stable cell lines of Dam-LMNB1/m6ATracer have much lower Dam-LMNB1 levels than the co-transfected cells but the GFP signal is weaker.
8. After generation of stable cell lines an antibody based GFP booster was often used, following the protocol described in §2.3.

CRISPR labelling

The workflow for a CRISPR-based imaging modality used in this study is composed of four sections: sgRNA design, synthesis of labelled constructs, *in vitro* control of sgRNA:protein function and *in situ* delivery to live cells.

sgRNA design

Initial developments were made using a sgRNA against mouse major satellite repeats (MS-sgRNA), which were previously used to publish the first CRISPR labelling system and had thus been previously validated by FISH (Deng et al., 2015). Furthermore the genomic location of major satellite repeats at dense mouse chromocentres makes for a clear interpretation of positive results with just a chromatin counterstain without constant FISH validation for every step. This allowed multiple iterations of tests on label incorporation, labelling densities and delivery protocols with comparable data.

Synthesis of labelled constructs

Labelled sgRNAs were synthesised via the incorporation of NTP homologues such as UTP coupled with organic dyes (Alexa488, 564, Cy5 etc. Figure 7.4) for live or fixed immunofluorescence (IF) and Biotin coupled UTP or Ethynyl Uridine (EU) for fixed sample IF or click chemistry respectively. Stochastic label incorporation throughout the sgRNA via *in vitro* transcription was possible by designing a dsDNA oligo with a T7 RNA polymerase promoter upstream of the sgRNA coding sequence, described in Bassett et al. (2013). This strategy allows titration of the number of incorporated labelled UTPs. In initial tests this achieved a five-fold increase in label density per molecule compared to the dCas9-eGFP system. Given that *in vitro* labelling can be done for multiple sgRNA the timescale from oligo synthesis to purified labelled sgRNA is on the order of 1-2 weeks. Chemical synthesis of sgRNA could in theory be used as an alternative to *in vitro* transcription, and for incorporation of wider palette selection not currently amenable to the T7 RNA polymerase. For *in*

in vitro transcription of sgRNA follow commercial T7 RNA polymerase protocols. For purification of *in vitro* transcribed sgRNA follow commercial Illustra G-50 purification kit (GE Healthcare).

Control of *in vitro*-sgRNA:protein function

After *in vitro* transcribed incorporation of labels in sgRNA it is necessary to control that the activity of the Cas9-sgRNA complex is not affected. This was shown by an *in vitro* cleavage assays that is compatible with Cas9 binding and function. A plasmid of known molecular weight with a GFP sequence flanked by sgRNA recognition sites is incubated with an active Cas9 nuclease and the sgRNA to test. Cleavage of the plasmid, determined by agarose gel electrophoresis confirms that the *in vitro* transcribed sgRNA is still capable of recognising the target sequence, complexing with Cas9 and does not impede Cas9 binding and activity (Figure 7.5).

***in situ* delivery to live cells**

There are two main options for delivery of the CRISPR system to cells. Firstly make dCas9-sgRNA complexes *in vitro* and introduce whole complexes into cell via cationic lipid transfection; as described in §2.2. After multiple attempts many issues were encountered with this method of delivery of protein-RNA complex. As a second strategy, stable cell lines were generated expressing dCas9 or dCas9-eGFP for sgRNA transfection or nucleofection. These Cas9 cell lines were made as described in §2.2. Nucleofection was performed as described in §2.4. Results of this section is found in Figure 7.7.

FISH probes and Raser-FISH TAD labelling

Nick-translation for probe generation

DNA FISH probes are generated via nick-translation reaction originally described in Rigby et al. (1977) with the addition of a hapten or fluorophore

tagged uridine homolog. The main reagent for this method is purified DNA homologous to the target sequence. This can be from mini-, midi- or maxi-prep kits (Qiagen) for plasmid constructs. BACs should be purified by Large-construct purification kit (Qiagen). Concentration of purified DNA should be accurately checked by fluorometric quantitation using the commercial Qubit instrument. This is recommended over simple spectrophotometers, as the measured fluorescence is proportional only to the DNA concentration from the proprietary reagent.

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- RNaseI
- β Mercaptoethanol 100 mM
- DNase (2000 U/ml)
- DNA polymerase (10 000 U/ml)
- Ice bath
- $MgCl_2$, 50 mM
- BSA, 0.5 mg/ml
- dATP 0.5 mM
- dTTP 0.5 mM
- dCTP 0.5 mM
- Agarose gel
- Electrophoresis tank
- Illustra G-50 columns

Protocol

1. Take 1 μ g of DNA and add 200 ng of RNase and top up with H_2O to a volume of 28 μ l. Incubate for 30 minutes at 37 °C and then leave on ice. RNase treatment removes any residual RNA from the DNA purification step and generates cleaner probes. Do not perform RNase treatment for RNA-FISH probe preparation.
2. Make Nick-translation buffer (NTB):

- Tris-HCl, pH 8, 500 mM
 - MgCl₂, 50 mM
 - BSA, 0.5 mg/ml
3. Make d(AGC)TP solution in demin water:
 - dATP 0.5 mM
 - dTTP 0.5 mM
 - dCTP 0.5 mM
 4. Add 5 µl NTB, 5 µl βMercaptoethanol, 5 µl d(AGC)TP mix to DNA with 1 nM hapten or fluorophore labelled dUTP.
 5. Add 3 µl DNase and 2 µl DNA polymerase and mix by gentle agitation. DNA polymerase is susceptible to denaturation by vortexing.
 6. Incubate mixture for 2 hours at 16 °C. Leave mixture on ice to stop enzymatic reaction.
 7. After incubation use 3 µl of this mixture and run it on a 2% agarose gel as described in §2.1. If a clear smear is observed mostly below 500bp the mixture is ready for purification. If the smear is mostly above 500bp further nick translation is required. In such cases take mixture from ice back to incubate at 16 °C. The exact time required must be empirically defined for each batch of DNase and DNA polymerase to account for changes in processivity.
 8. When nick translation is complete, purify the FISH probes using the protocol from commercial Illustra G-50 columns.
 9. FISH probes can be stored at −20 °C.

FISH probe precipitation

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Cot1 DNA (m/hCot)
- Salmon sperm
- Sodium acetate
- Ethanol
- Microcentrifuge

Protocol

1. Pipette 100 ng of FISH probe DNA to 1.5 ml eppendorf tube and keep on ice.
2. Add 6 μ l of mouse or human Cot1 DNA (for experiments on mouse or human cell lines respectively).
3. Add 2 μ l of salmon sperm.
4. Add 10% per volume of sodium acetate.
5. Wash walls of tube with 2.5x per volume of ice cold 100% ethanol.
6. Centrifuge for 15 minutes at 4 °C at 13000 rpm on microcentrifuge.
7. Leave tube at –80 °C for at least one hour, preferably overnight, for DNA FISH probe to precipitate.
8. Probes can be left precipitated at –80 °C in storage.

Raser FISH

The following protocol for Raser-FISH was developed by the Veronica Buckle and Jill Brown at the Weatherall Institute of Molecular Medicine (WIMM).

Solutions and Equipment:

- Pipette
- Eppendorf tubes (0.5 ml to 2 ml)
- Seeded cells
- BrdU
- BrdC
- Precipitated FISH probes
- Microcentrifuge
- Ethanol
- TritonX-100
- Ammonium chloride
- PBS
- Hoechst stock: Use 100 μ l of 100 μ g/ml stock in 20 ml.
- Kreatech hybridisation buffer (KBI-FHB)

- Shaker incubator
- Exonuclease III (NEB MO206L)
- Exonuclease buffer 10x
- H₂O demin
- UV crosslinker
- Rubber cement
- SCC solution

Protocol

1. Cells should be seeded on high precision cover slips as described in §2.2 to the confluency required for the planned experiment.
2. To cells in 6 well dishes add BrdU-BrdC mixture (at 3:1 ratio respectively) at 10 M to 5 M overnight. There are cytotoxic effects from this incorporation so incubation should not last longer than 15 hours.
3. Take precipitated DNA FISH probes prepared as described in §2.5 and centrifuge for 15 minutes at 4 °C at 13000 rpm.
4. Cells should be fixed as described in §2.3.
5. Decant ethanol from centrifuged probe tube and wash with 400 µl of 70% room temperature ethanol. Centrifuge as before.
6. Permeabilise cells in TritonX-100 as described in §2.3.
7. Incubate cells for 10 minutes at room temperature in a solution of in 0.26% ammonium chloride. This quenches azide groups.
8. Take precipitated probe, decant ethanol and dissolve in 12 µl hybridisation buffer. Incubating for 10 minutes in heated shaker incubator at 37 °C followed by 20 minute incubation at rest also at 37 °C.
9. Wash cells in PBS as described in §2.3.
10. Make solution of Hoechst 33258 to 0.5 µg/ml in PBS and incubate cells for 15 minutes in the dark.
11. Wash cells in PBS as before and leave cells in PBS, exposing to UV light (260 nm) for 15 minutes at room temperature. Note that after UV exposure PBS solution would have heated up so PBS should be replaced with PBS at room temperature.
12. Prepare exonuclease mixture (for 6 coverslips) by adding:
 - Exonuclease III, 20 µl.
 - Exonuclease buffer 10x, 40 µl.
 - H₂O demin, 340 µl.

13. Dispense exonuclease mixture on flat Parafilm as described in §2.3 and incubate cells on exonuclease mixture for 10 minutes in a humidified dark chamber at 37 °C.
14. Wash cells in PBS as before.
15. Denature probes. For Bacs, oligos or core plasmids this is done by incubating for 10 minutes at 95 °C. BACs also require a following preannealing with at 37 °C for 20 minutes (for Cot1 to block repeats).
16. Dispense denatured and preannealed probes on clean microscope slides and invert coverslips on the probe solution.
17. Seal coverslip with rubber cement and incubate overnight horizontally on a slidebox floating in a waterbath at 37 °C overnight for the hybridisation to take place.
18. Following incubation remove rubber seal with fine pliers and wash coverslips in 4xSCC solution in the same fashion as PBS washes previously described.
19. At this point, epitope blocking, immunofluorescence detection, counterstaining, mounting and sealing can be performed as described in §2.3.

ChaiN - a high-throughput pipeline for super resolution image analysis

To investigate the relationship of individual histone modifications to the underlying chromatin architecture I have devised a conceptual semi-automated quantitative image analysis workflow: ChaiN , short for ChaiN analysis of the *in situ* nucleome (Figure 3.1). This methodology segments, annotates, and retrieves the 3D coordinates of immunofluorescent signals, and segments the intensity value of the chromatin stain as a proxy of chromatin density. It also defines a boundary between detectable chromatin environments and the non-chromatin "interchromatin compartment" (IC). The final output is useful for analysing how the nucleus is divided in terms of molecular biology as a function of density. The metrics to measure and compare chromatin environments are the surface area to volume ratio of the chromatin network, the relative occupancy of any epigenetic marker stain over the range of chromatin densities and absolute distance of the markers from the IC boundary.

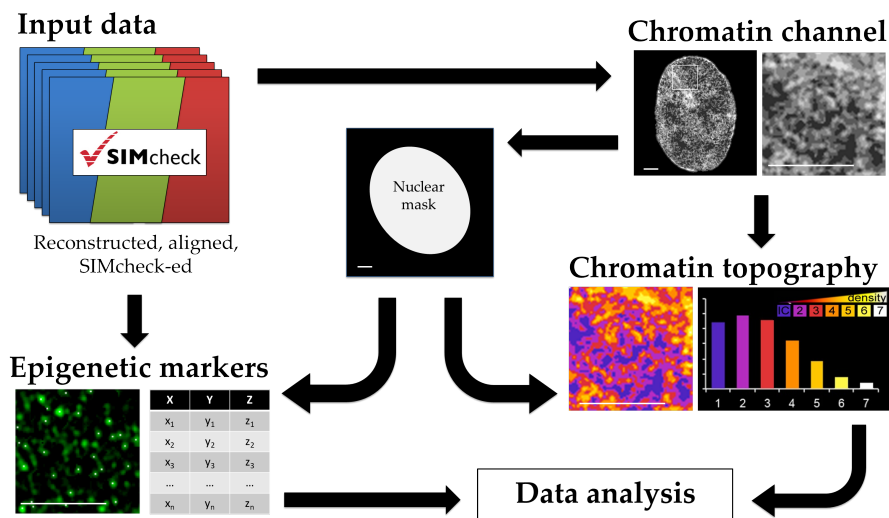


Figure 3.1: **Chain workflow.** Schematic representation of the Chain workflow. Input for Chain are images pre-processed, reconstructed, chromatically aligned and cleaned with SIMcheck (Ball et al., 2015). Different chromatic channels are segmented on the one hand from chromatin density and nuclear volume mask to the centroid position of IF spots on the other. Both of these are used alone and in conjunction with each other to analyse the 3D epigenome. Image analysis is performed in the Octave and R programming languages. Scale bar = 2.5 μm .

Chain requires the use of the programs listed in each section with their respective plugins/packages. These have been chosen, where possible, due to their wide distribution across operating systems and open source/free (*libre*) nature, for ease of accessibility, modification and configuration. The Chain analysis in this study was implemented on a GNU/Linux environment (Debian 8). A link to the source code for all can be found in the appendix (Appendix 9.1).

This section also covers the technical requirements necessary to take an IF stain from SIM acquisition to final normalisation of the aforementioned extracted metrics.

3.1 Acquisition and reconstruction

Acquisition, deconvolution and reconstruction

Images acquired for qualitative screening of antibody cross-reactivity or general experimental validity (not for subsequent image processing) were acquired by a widefield set up (API, DeltaVision Elite) and deconvolved with SoftWoRx (DeltaVision's acquisition and image analysis suite). Conservative parameter settings were chosen, sacrificing some of the potential gain in resolution to decrease the possible introduction of artefacts.

Images acquired via 3D SIM (API, OMX V3) require a larger range of parameters for proper acquisition and subsequent reconstruction. The optimisation of these parameters was determined empirically and each require preparation of dedicated calibration microscope slides. A brief description of each is given in this section; however a thorough and in-depth description of SIM calibration and artefacts to avoid can be found in Kraus et al. (2017) and Demmerle et al. (2017). Details of the 3D SIM optical set up and fluorescence light paths used in this investigation can be found in Appendix 9.3.

The main parameters required are: the optical transfer function (OTF) of the point spread function (PSF) of each wavelength, the K_0 angle of rotation of the SIM striped illumination and the top-phase and pitch of these stripes (described in Demmerle et al. (2017)). OTF generation or optimisation is wavelength specific and requires sparsely populated bright single foci from which a PSF can be acquired in 3D. K_0 , top-phase and pitch are also wavelength specific but are measured and optimised by a higher density concentration of fluorophore-labelled beads, enough to create homogenous puddles which allow the acquisition of high contrasted stripes over the entire illumination plane. It is worth noting that different microscopes have other parameter requirements for their reconstruction software.

PSF slide for OTF

Coverslips should be cleared of dust and small particles by multiple washes, first in 1% Hydrochloric acid solution for 10-30 minutes followed by rinsing in de-ionised H₂O and finally in 100% ethanol. Coverslips must then be left to dry vertically in a coverslip rack.

Commercial beads ("FluoSpheres", carboxylate-modified microspheres) come at high enough concentrations such that their undiluted use does not yield slides from which parameters can be calculated. To achieve the desired concentration (empirically defined subsequently) needed for PSF slides most commercial bead solutions should be diluted by a serial dilution to a final concentration of 1:1,000,000 to 1:100,000,000. These dilutions should be mixed in symmetric stoichiometry and kept refrigerated at 4 °C. The commercial beads used were:

- Red beads: 100 nm to 110 nm diameter, 580/605nm wavelength, Life Technologies F8801.
- Green beads: 100 nm to 110 nm diameter, 505/515nm wavelength, Life Technologies F8803.
- Blue beads: 100 nm to 110 nm diameter, 350/440nm wavelength, Life Technologies F8797. 170 nm diameter, 360/440nm wavelength, Life Technologies P7220.

For PSF slides a larger blue bead is used to compensate the lower brightness of the blue channel in the OMX system. The mixed beads solution is diluted 100 fold in 100% ethanol, using 5 µl of the dilution per cover slip. After ethanol has evaporated coverslips are mounted on high precision slides in 20 µl pure glycerol. Once glycerol has covered the volume encompassed between the coverslip and slide, the coverslip is sealed with clear nail varnish. It may be necessary to use a soft tissue to soak up any excess glycerol pooling on the edges of the coverslips (causing the coverslips to float and move) before applying varnish.

Puddle slide

The methodology here is similar to that for the PSF slide, save for requiring higher bead concentration, and the need to generate a separate slide for each wavelength. In PSF slides the beads are all labelled with individual fluorophores but separated far enough apart (10 nm to 20 nm between beads) therefore all wavelength channels can be represented in a single cover slip. For puddle slides the 100 nm to 110 nm diameter beads can be used for all colours, including blue, as the concentration of beads required to make dense puddles is also bright enough for detection. A more descriptive methodology and dilutions for all SIM optimisation slides can be found in Kraus et al. (2017) and Demmerle et al. (2017).

Alignment and software

The final step required to transform raw or reconstructed images into input for ChaiN is image registration. This is a requirement on all systems which have separate cameras and detectors for each individual wavelengths as different lightpaths shift the image projected onto each receptor relative to each other. The field of image alignment is many decades old and boast multiple methods for achieving similar results, with varying costs to either speed or accuracy dependent on sample and algorithm (reviewed in Oliveira and Tavares (2012)). Nowadays proprietary commercial image analysis packages have inbuilt tools for image registration. During the course of this investigation two alignment softwares were tried and compared, primarily: Softworx and OMX editor (API DeltaVision alignment software, and Chris Weisiger and John Sedat, UCSF, respectively). Of these, SoftWoRx currently lacks automatic alignment in the z plane which makes it not adequate for an image-analysis-based project of thousands of images. OMX editor includes this capability as well as being able to process large queues of images in batch alignment runs (provided all images require alignment by the same pre-acquired parameters).

Regardless of software used, all batch alignment programs require a set of alignment calibration parameters. These can't often be extracted from the actual sample to be aligned, as usually different entities are stained in different channels, with the occurrence of one signal bearing no geometric relation to a signal in another channel. Hence to measure this calibration parameters an image registration/alignment slide needs to be made.

Bead alignment

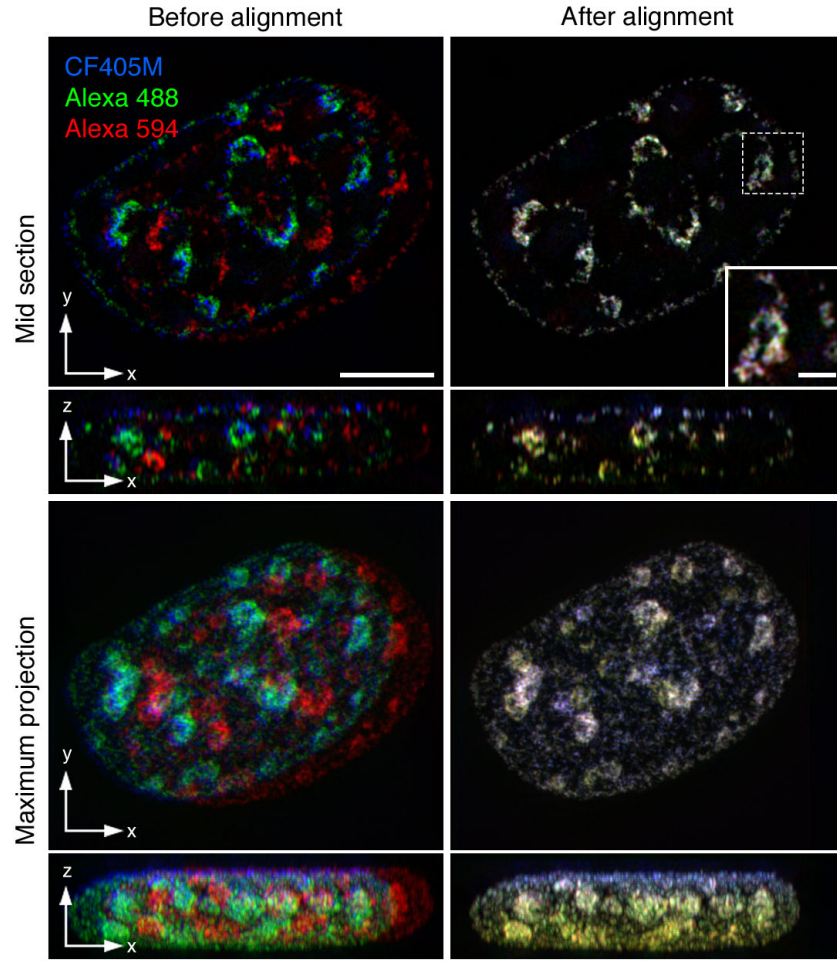
As with PSF and puddle slides, fluorescently labelled beads may be used in order to generate this calibration sample. Because precise alignment is required, these beads must be simultaneously labelled with a selection of fluorophores matching the excitation lines of the system. For the OMX V3, 200 nm diameter "Tetra-Speck" microspheres (Life Technologies T7280) were used. The methodology for making this slide is identical to that for making PSF slides outlined previously in §3.1, with 2 major exceptions. First although theoretically a single multi-spectral bead is enough to produce alignment parameters, it is often useful (for reasons described below) to have multiple beads per image, hence a higher concentration is required. Single beads are still desired over puddles so an dilution of the initial bead stock needs to be used, here 1:100,000 to 1:1,000,000 approximately (empirically defined for 1-10 beads per field of view). Second, application of this dilution over the cover slip only (as is the case for PSF slides) would yield a slide useful for alignment in 2D but one lacking any z component. This is because all beads will settle roughly within the same z plane (within the first tens of micrometres) near the cover-slip. To generate a slide for 3D alignment another drop must be dispensed on the slide too. Once both the coverslip and the slide have dried they may be mounted with glycerol between them.

Biological alignment

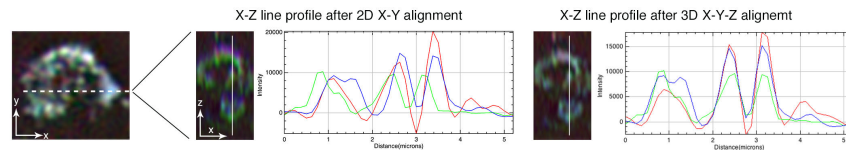
Registration calibration slides can also be made from biological samples. These have the advantage of being at hand from spare specimens. A single monolayer of tissue-cultured cells, i.e: mammalian primary cells or tumour cell lines can be used. As with any slide for registration parameters it is important that the same structure is labelled in all channels to be aligned. By taking advantage of synchronously firing origins of replication, incorporation of labelled nucleotides (eg: EdU) during S phase can be used for this task. After 10-20 minute incubation enough labelled nucleotide will be incorporated such that clear individual foci can be detected after fixation, i.e: via Click chemistry outlined in §2.4. The key step in generating these registration samples is to use a mix of all fluor label reaction. These fluorophores will attach to all sites with the labelled nucleotide and since these domains are highly abundant in labelled nucleotide the resulting stain has points labelled in all channels (Figure 3.2). Another advantage of biological samples over beads is the coverage of all z planes, identifying misalignments throughout the sample easier than from two 2D bead planes N nm apart. Note that the number of times a single alignment slide can be used depends on the durability of the biological sample when stored at 4 °C (cellular compartments degrade over time even when fixed and mounted).

Pre-processing**Pre-analysis data check**

Before committing files for analysis it is imperative to control for errors or data corruption which are detectable using image visualisation software, such as ImageJ. This is important when applying the ChaiN workflow in batch mode. It may no longer be practical to look at the output for hundreds of files at each junction on the analysis pipeline, but taking the time at the beginning to ensure that all files are correct will help avoid unexpected error in the analysis



(a) EdU quad staining



(b) Z alignmet

Figure 3.2: **Biological alignment sample.** (a) Quad-colour staining of EdU before and after alignment (left and right respectively) for a single z slice and maximum intensity projection of full z stack (top and bottom respectively). Note that forth colour in infra-red stain is missing from display. (b) Zoomed in chromocentre region viewed in x-y plane after 2D alignment (left) and comparing visual and intensity line profiles of alignment in z after 2D or 3D alignment algorithms (Softworx, OXM editor respectively). Adapted from Kraus et al. (2017). Scale bar = 5 μm , inset scale bar = 1 μm .

scripts which will otherwise halt progress and ultimately cost more in terms of time loss. Consulting a tool such as ImageJ's SIMcheck plugin provides a wide range of ways to examine the quality of raw and reconstructed SIM data and suggest optimal values for Wiener filter magnitude to refine data reconstruction before committing to analysis. An in depth description on how to use SIMcheck for SIM data inspection (required in ChaiN for clean-up) is covered in detail in Ball et al. (2015).

Furthermore one may still want to use image visualisation at key points in this analysis pipeline, such as for inspecting any tiff-formatted binary masks as well as foci and chromatin landscape segmentations on randomly selected files when processing in batch mode. This is enabled since intermediary steps in the ChaiN workflow are saved as control/log/quality image files.

Pre-analysis clean-up

One feature of the SIMcheck plugin which the ChaiN workflow relies on is the "Stripe contrast modulation" scoring. This makes use of *a priori* knowledge about how SIM data is acquired to display a voxel by voxel score of the quality of reconstruction at each voxel, its reconstruction fidelity (Figure 3.3). Although reconstruction fidelity and stripe brightness is mostly a positive linear correlation, this tool is more than a simple linear threshold of raw data intensity. Very bright regions may also cause low stripe modulation, due to saturation of the camera's bit-depth. Such regions show artefacts similar to regions of low stripe modulation caused by low signal intensity, generally hexagonal or parallel-striped high-frequency noise (Figure 3.3; inset, red arrowheads). To use this tool I have written an automated ImageJ macro (Appendix 9.6) which generates binary masks of "trustworthy" voxels which can then inform downstream ChaiN scripts if elements in that region are to be taken as real segmented structure or discarded. By careful adjustment of laser intensity and exposure during sample acquisition one can minimise the number of voxels with low modulation contrast which will otherwise be discarded

voxels, yielding higher statistical power per acquisition. No amount of clever and automated post-acquisition clean-up can truly be as good as starting with higher quality images.

3.2 Feature segmentation

Automation and requirements

The Chain workflow relies on two main programming languages for image and data analysis: GNU Octave and R. Both are open source, free (“libre”) software and were chosen because this makes the dependencies of Chain easily available with manuals for their use already accessible. Furthermore, both have large community-based forums where users who are not skilled programmers may ask for help if any issues are encountered. In the context of Chain, Octave is primarily used for image processing, such as the morphology operations needed to segment foci, described in the following section. Octave can be run from command line or from the newly introduced GUI.

For Chain, R must have the “*BioimagerTools*” and “*EBImage*” packages. R is used by Chain for image processing regarding chromatin landscape segmentation, however it is mostly used for numerical data processing and statistical analysis. Further packages for R are optional during manual analysis. “*ggplot2*”, is a tool for data visualisation and representation and was used throughout this project for graphs and plots in most figures.

To manage the multiple steps of batch processing of all the data, GNU’s “*Make*” was used. This is a reliable build system which is commonly used for compiling software. As with other build systems, Make can be programmed to keep track of all dependencies for every step within the network of scripts which make up the workflow. In Chain this is used to create a hierarchical tree of the inputs and outputs of every analysis script (Figure 3.4 and Appendix 9.6). Make is able to process all files of a particular type or format in an identical manner and naming each relevant script for processing these

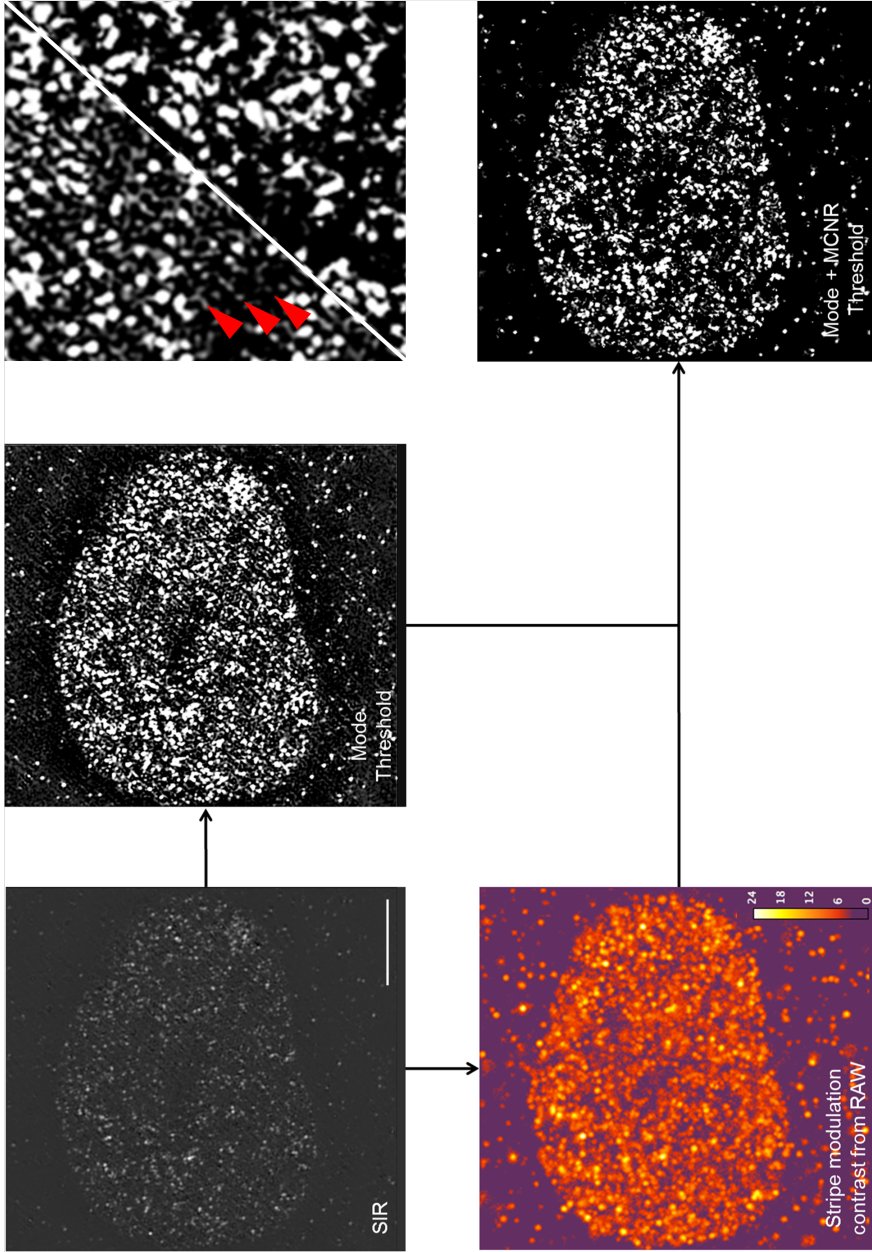


Figure 3.3: **Modulation contrast clean-up.** SIM single z plane image representing clean-up workflow from reconstructed and aligned image (SIR). A “mode threshold” is used to reset the pixel intensity values with 0 for voxels at or below the image mode (non-cell, grey background). SIMcheck is used to generate a stripe modulation contrast map (MCNR, bottom left) with a scale of reconstruction fidelity from 0-24 where <4, 4-8, 8-12 and >12 is poor, moderate, good and very good/excellent reconstruction fidelity respectively (Ball et al., 2015). This MCNR map is thresholded to adequate minimum (5) and made into an MCNR binary mask to complement the mode thresholded image (below a score of 5, reconstruction is not reliable for quantitation). In this example, increasing the contrast for both the mode threshold and mode threshold+MCNR mask (top and bottom right) reveals noise as many high-frequency stripes present in the former and removed in the latter (compare red arrowhead in inset, top right). Scale bar = 5 μm

files in the correct manner and order. Because Make can build up dependency trees, one is also able to configure Make to create a file which relies on input files which have not yet been made. In these cases, Make will work backwards to generate these inputs, and the required files for these (Figure 3.4). In this way the user may command the generation of the last output file in the pipeline, and all intermediate steps from the initial data onwards will be generated automatically, effectively “daisy-chaining” all scripts of the Chain workflow without user interference. Furthermore, flags can be used in command line for Make to run jobs in parallel which can significantly reduce the total computation time. An in-built help page which outlines the target of each Make command in Chain can be accessed by the “*make help*” command.

Masking

For any given image processed as part of this work, the biological region of interest only occupies a smaller volume than the total volume of voxels of the image. Since the size of the field of view is deliberately larger than the size of a cell’s nucleus. Importantly, the volume outside of the nucleus is never empty. Depending on the stain or antibody specificity/cross-reactivity, some images may harbour a low content of signal in the cytoplasmic or extra-cellular volume. A 3D binary mask including only the nuclear volume was generated to define the region of interest. When applied to subsequent scripts, this mask ensures that all non-nuclear data (false positives) are excluded; increasing the accuracy of the results and decreasing strain on computation. To achieve this, Chain employs a “*nuclear_mask.m*” Octave script (Appendix 9.6). The function of this script was originally written by David Pinto as part of a larger script subsequently edited by me.

IF foci centroid coordinates

Spot segmentation is required for the quantitation of punctate patterns from immuno-fluorescence targets within the counterstained nuclei. These spots

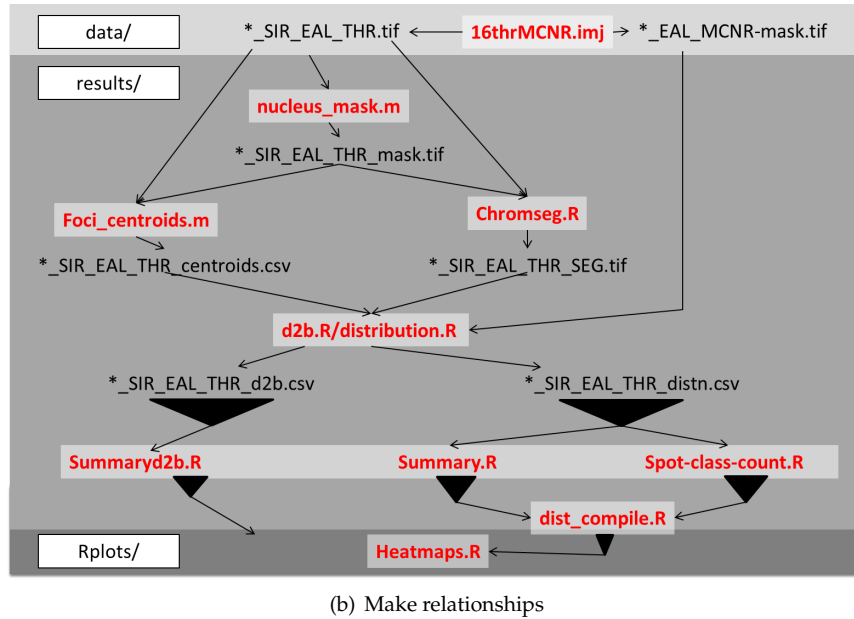
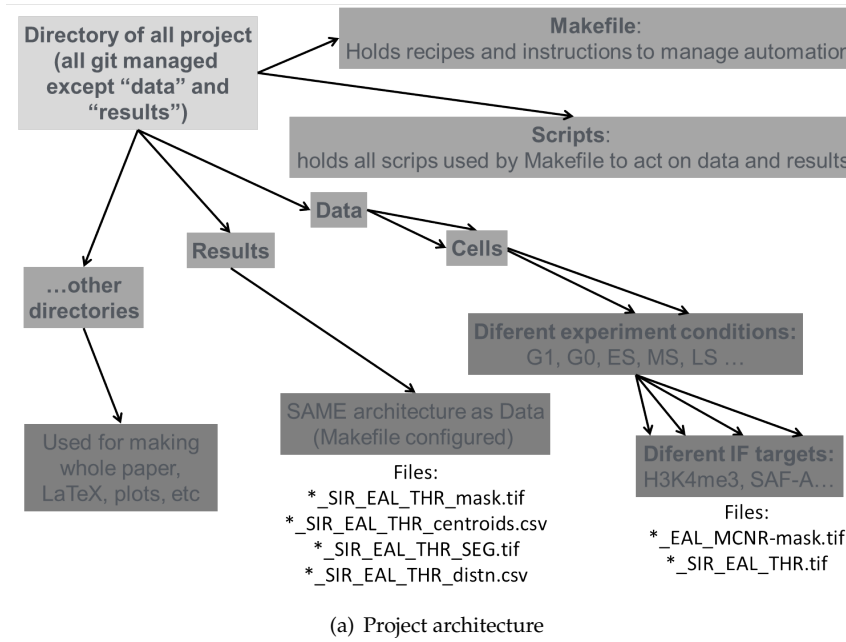


Figure 3.4: **Make workflow.** (a) Architecture of the ChaIN project allowing Make-control of automation by saving the mode threshold reconstructed image (ending with _SIR_EAL_THR.tif) and its MCNR mask (_EAL_MCNr-mask.tif) under a logical directory architecture which Make can replicate to save analysis outputs: *_SIR_EAL_THR_mask.tif, *_SIR_EAL_THR_centroids.csv, *_SIR_EAL_THR_SEG.tif and *_SIR_EAL_THR_distn.csv. (b) Graph of input dependencies coded in the Makefile. The Makefile is a file containing recipes outlining how the dependencies for each function/script relate between the data and results directories, scripts, and the automatic generation of file names for ease of data sorting and mining. The inputs required and outputs generated by each file are outlined in Appendix 9.6.

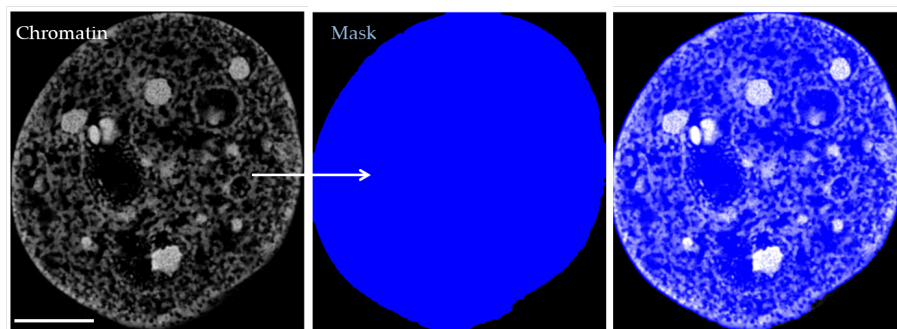


Figure 3.5: **Chromatin masking.** The “*nuclear_mask.m*” script uses the “image” package in Octave for the automatic generation of binary masks for the whole nucleus. The mask allows the exclusion of all non-nuclear voxels regardless of intensity and reconstruction quality, from the *a priori* knowledge that these are always false positives. Scale bar = 5 μm .

can be resolved by SIM, provided they are sufficiently spaced, to 100 nm in x-y dimensions and 300 nm in z (Demmerle et al., 2017). However the source of this point spread function (the excited fluorophore) is in the order of a few nm.

The “*foci_centroids.m*” script segments all puncta within the nuclear volume and assigns a weighted 3D centroid coordinate for each at sub-pixel resolution (Appendix 9.6). 3D watershedding followed by intensity thresholding implementing the “Otsu” thresholding algorithm, ensures that high density clusters of spots can be separated, rather than erroneously assumed to be a filamentous structure. A weighting factor is then implemented biasing the position of centroids towards the brightest pixels in a segmented foci. The resulting centroid position is a fairer representation of the real chromophore position than assignment by a simple geometric centre of mass as it uses shifts in signal intensity to estimate the chromophore position. The CSV output file contains four columns for X, Y, Z coordinates in nm and the number of centroid (Figure 3.6). The function of this script was originally written by David Pinto as part of a larger script subsequently edited by me.

As with all workflows there are limitations to this approach. The use of the Otsu algorithm for automated thresholding makes a good segmentation out-

put dependent on an high signal to noise ratio in the analysed images. Hence, visual inspection and validation using the log image files created by ChaiN is recommended, in particular when samples using new antibodies of unknown specificity. Furthermore, the basis of epigenetic marker detection via IF also has its limitations. First, it relies on indirect labelling. This means that the signal is only a proxy for the position of the target, provided that the antibody has high specificity for the target epitope and no cross-reactivity. High concentrations of antibody also reduce the efficacy of washing steps in the staining procedure, such that free-floating antibodies may be fixed, imaged and analysed as false positives. Second, antibodies are moieties of approximately 15 nm. Therefore the position of the actual target is never truly at the position of the fluorophore and this decreases the measurement accuracy Kraus et al. (2017). The use of a double antibody strategy only compounds these limitations. On the upside, antibody based detection is a reliable protocol which can be implemented with relative ease. Moreover, the modularity of primary and secondary antibody based detection allows many different primary antibodies to be analysed with the same secondary antibody (provided they are raised in a species producing cross-reactive IgG peptides). This aids the experimental design and increases the accuracy of comparison across samples, as no differences in the specificity of different secondary antibodies needs to be taken into account.

Chromatin topology

Segmentation is used for a simplified characterisation of the intricate topology of chromatin by the signal intensity of a chromatin stain as a proxy for chromatin density, resolved by super-resolution 3D SIM. In ChaiN , this is performed by the “*Chromseg.R*” script.

Chromatin signal intensities are normalised to the maximum intensity of chromatin in the image for comparison across samples. This is followed by linear thresholding the intensities into multiple bins, thereby segmenting

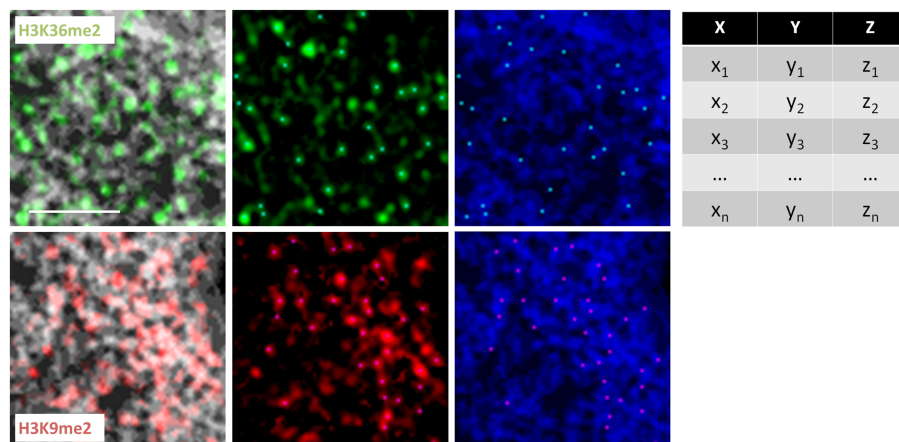


Figure 3.6: **Centroid segmentation.** The “Image” package in Octave was used for the localisation of the centre of individual IF spots. Here, for zoomed-in regions of H3K36me2 (top) and H3K9me3 (bottom). From left to right: composite of chromatin (grey) and IF (green or red), composite of marker and localised centres (cyan or magenta respectively), composite of chromatin (blue) with localised centroids. A CSV file with coordinate positions in x, y and z is saved (top right). Centroids are filtered by intensity and size threshold of IF spots and a 3D watershed function to split neighbouring spots whose PSF may be touching in 3D (appearing as red or green continuous filaments in middle composites). Not all IF spots show a centroid in this displayed region as these are in a different z plane not shown. Scale bar = 1 μm .

them into discrete “classes” based on intensity, which are subsequently refined through a Hidden Markov Model (HMM) as first outlined in Markaki et al. (2012). Implementation of the HMM may re-assign the “intensity class” of an individual voxel by taking into account the overall signal (or lack of) in surrounding voxels. This fine-tuning allows the capture of fine chromatin architecture on chromatin-sparse regions and of chromatin depleted channels within chromatin-dense structures, conversely (Figure 3.7). This segmentation also normalises the minima and maxima of the dynamic range between samples. The number of classes chosen was maintained at 7, as described in Markaki et al. (2012). Although partially arbitrary, this number is also informed by empirical observation that approximately a seventh of the dynamic range in a given micrograph contains all non-detectable chromatin environments. This volume in ChaiN is correlated to the IC, although some chromatin

may be found in this region as very small loops beyond the resolution of SIM. As the “*Chromseg.R*” script normalises the chromatin stain to the maximum intensity of chromatin in each image (and not to a hard cut-off) the definition of the IC as the first seventh bin of the intensity remains. Note that comparison across samples stained with DNA dyes of different characteristics or may have a different definition of the IC. A clearer definition of this space may in future be obtained by an exclusion staining independent of DNA. Possibly by fluorescently tagged moieties of different molecular weights. The regions excluding the largest moieties should be the chromatin network and by consequence, the rest of the nuclear volume can be defined as the IC.

Having 7 classes is also a good compromise between the different number of bins in which to segment the chromatin network. On the one hand, statistical rules to estimate the optimal number of classes solely on the resolution of distributions, such as Sturge’s, Scott’s or Freedman-Diaconis rules, can range from 17, 37 or 48 classes respectively (Figure 3.7). On the other hand, a harsh binary split of the network could also have been used with a max of 3 classes (one for the IC and two for chromatin). However any nuance or gradient in the distribution of markers would have been lost in this approach. By using 7 classes, 6 can be dedicated to the description of chromatin. This allows both a binary post-processing interpretation of low (2-4) and high (5-7) chromatin density, but also a softer interpretation of low (2-3), middle (4-5) and high (6-7) chromatin density, providing a middle ground for markers which may have more central/broad distributions. This is useful as genomic function may not be divided simply into two density environments. Having flexibility for interpretation post-processing could also be a case for using the higher number of bins as suggested by the statistical rules. However little is gained in from doubling or over six-fold increase in bin number (Figure 3.7) and in practice this makes downstream comparisons between datasets unwieldy.

The output generated by ChaiN therefore is a segmented image of 7 classes representing the original chromatin landscape with a high enough class num-

ber to assess subtleties in the distribution of markers and few enough classes not to subdivide the interchromatin compartment. All regions outside the nuclear mask will be in “class 0” which is not counted as a class for the purpose of analysis.

As there is no stain for nucleoli in this investigation (given the technical limitations of the number of lasers and dyes in the OMX SIM system used, see Appendix 9.3, Figure 9.4) it is not possible to disentangle the proportion of class 1 belonging to nucleoli vs the IC. Nevertheless, as the majority of this work focuses on describing the relative differences between markers in the same cell type, conclusions from such measurements do not suffer from a small systematic bias across all samples. Certainly, when comparing across different cell types, such as in §6.2 between C127 and IMR90 this limitation needs to be taken into consideration. Cell lines such as C127 with more/larger nucleoli will have a larger class 1 volume and so when normalising the distribution of markers to this class there may be a skew to less enrichment (or increased depletion) relative to cell types with fewer/smaller nucleoli. Further optimisation of ChaiN should take as input a nucleolar-specific stain and used an adapted version of “*nuclear_mask.m*” to mask-out the nucleolar volumes.

3.3 Measurements and statistics

Chromatin metrics

Exposing cells to a variety of chemicals or buffers may cause visible perturbations to the overall chromatin morphology (Talwar et al. (2013), Figure 6.2 Figure 6.4). Some of these, trigger reversible compaction or decompaction of the chromatin environment, eg: hyper/hypo-osmolar growth media (Irianto et al., 2013). Others conditions achieve the same result irreversibly, by ATP sequestration via sodium azide, or by global nucleosome deacetylation via trichostatin-A for compaction or decompaction respectively. The global ef-

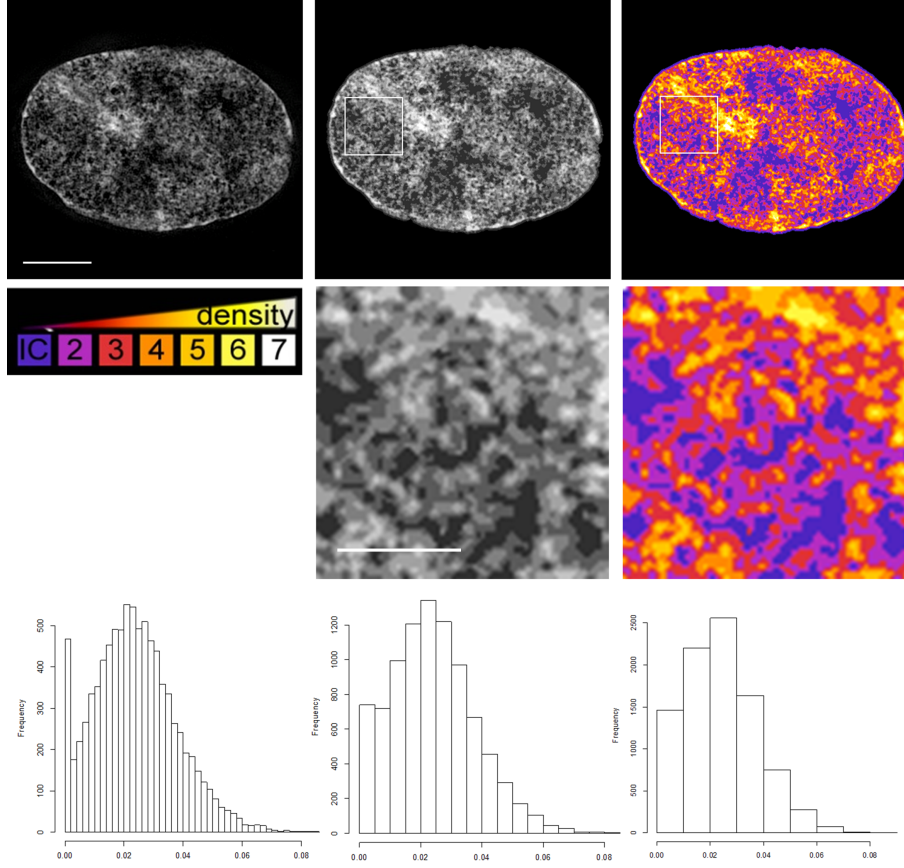


Figure 3.7: **Chromatin density classes segmentation.** Top: Chromatin is segmented by intensity, density proxy, into 7 classes (plus a 0^{th} class, outside the nucleus, top row gallery). The lowest class represents the inter-chromatin compartment (IC) which can be distinguished from all other classes in pseudo-coloured images (right, blue). Bottom: Binning into 7 classes is the best compromise between decreasing complexity of analysis whilst keeping analysis comparable to higher bin numbers [markaki 2012](#). In this example the distribution of centroids for the IF of histone modification H3K36me3 is plotted as a histogram of their absolute frequency (y axis) against the same chromatin stain segmented in various bin numbers (x axis). The same left-shifted distribution is observable if 7 bins are used (right) instead of suggested bin numbers by: Freedman–Diaconis’ (FD) rule (48 bins, left) or Sturge’s rule (17 bins, centre) for bin-number estimation. FD is optimal for random samples of normally distributed data and its less influenced by outlier values than Scott’s rule. Sturge’s rule assumes an approximate normal distribution and can perform poorly if $n < 30$, not true in this case with $n \approx 3.5$ million voxels analysed per image. Scale bar = $5 \mu\text{m}$, inset scale bar = $1 \mu\text{m}$.

fects of these perturbations have been described without the application of super resolution microscopy techniques. Nevertheless, a thorough metric for quantitation of these effects has been lacking in the field. With the use of SIM the rearrangements of chromatin can be resolved to yield images of the quality necessary for the extraction of measurements of chromatin anatomy at the nano-scale.

Chromatin volume

ChaiN uses the output of the chromatin segmentation to approximate the volume of the chromatin in the nucleus (Appendix 9.6). This makes use of the available resolution to estimate changes in genome size.

Because ChaiN has a previous chromatin segmentation step, it is able to find not just the ratio of chromatin to the nucleus, but also the ratio of different densities of one chromatin density class to the volume of all chromatin in any particular nucleus (Figure 4.2). This utilises the nuclear mask calculated by ChaiN previously (Appendix 9.6), but by replacing the whole-nucleus mask with specific sub-nuclear regions of interest (i.e: sites of nascent DNA replication) then the distribution of chromatin densities at these locations can also be known (Figure 5.2).

Chromatin surface area

Using ChaiN a boundary region between the interchromatin channels (spaces within the nucleus without detectable chromatin) and regions of chromatin detection (of all intensities) can be defined. Because this feature can be extracted as a numeric approximation of the number of voxels representing it in the 3D image, it can also be normalised to the previously extracted approximation of the voxels representing the chromatin volume from the landscape segmentation. Such normalisation is effectively a surface area to volume ratio metric. Surface:volume can be helpful in describing and comparing the magnitude by which the previously described perturbations effect the chro-

matin network, and is a metric which has not been possible to implement before. Networks with low complexity in 3D topology will be represented by lower values of surface:volume whilst increasing ratios are proportional to increase in topology complexity; more channels and convolutions of a 3D volume cause an increase in surface area for the same available volume (Figure 4.3). A single and perfectly spherical domain of chromatin would theoretically have the lowest topology complexity and thus the lowest ratio.

Chromatin network width

Another metric of the nature of chromatin networks is the description of the average width of the network filament. Using the segmented chromatin surface area, ChaiN is able to estimate the distances of all chromatin-containing voxels to the chromatin surface (Appendix 9.6) by calculating the distance of a high concentration of random spots (Figure 3.8(a), Figure 3.8(b)). The filamentous network was modelled as short tubular stretches of various widths but with an average diameter. A cross-section of the average tubular filament can be represented as a circle (C_1) of radius R (Figure 3.8(c)). When random spots are placed into the volume these will intersect randomly to the perpendicular plane formed from the cross-section of the filament, such that this can be represented also by random spots on the 2D cross-section. Taking the random distribution of these spots, 50% will be found in a smaller area, a circle (C_2) with radius r , such that $R = \delta + r$ where δ is the difference between these radii (Figure 3.8(c)). I calculated the ratio between δ and R as follows:

$$Area(C_1) = 2 * Area(C_2)$$

$$\pi R^2 = 2\pi r^2$$

$$R^2 = 2r^2$$

$$R = \sqrt{2}r$$

$$\delta + r = \sqrt{2}r$$

$$\delta = (\sqrt{2} - 1)r$$

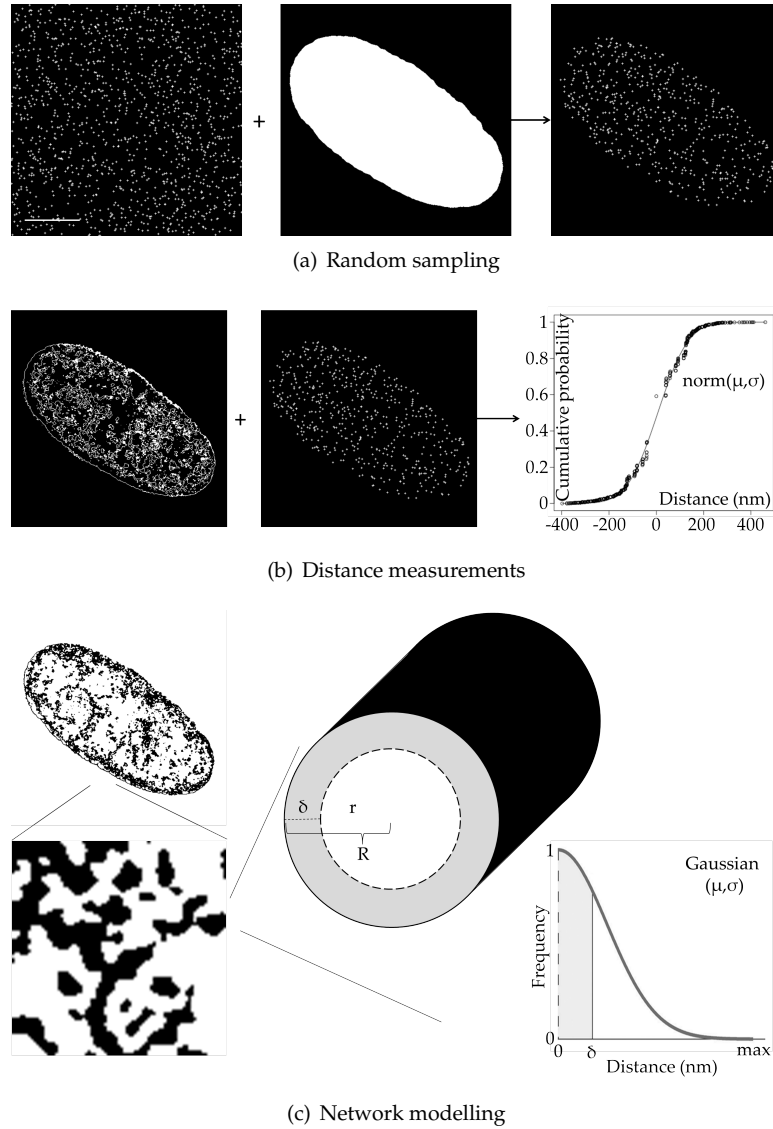


Figure 3.8: **Chromatin network dimension.** (a) A random sampling of distances to the network boundary can be obtained by generating a random scatter of coordinates *in silico* (left) restrained in the volume of the nuclear mask (middle) to give a pseudo-IF centroid coordinates (right) with no biological bias. (b) The distance of the network boundary (left) to the aforementioned random spots is calculated (middle). As this is a random distribution, plotting the cumulative probability of each centroid (right, y axis) to distance to the boundary (right, x axis) yields a symmetric curve (black circles) which can be described from fitting a cumulative normal distribution (black line). Fitting is achieved by the least square method. From this a mean distance (μ) and standard deviation can be obtained (σ). (c) From empirical observation of cells the chromatin network can be approximately modelled as a network of tubular filaments with a degree of variability in width but with an average radius of R (top left, bottom left, and centre tube). Plotting the average distance (those within the network) is possible by taking the average μ and σ fitted from all cells. The average radius R can be sub divided into a smaller radius “ r ” which encompasses the inner half of the volume of the tube (and therefore half the area of the cross-section) and δ , the difference between r and R . Scale bar = $5 \mu\text{m}$.

And therefore:

$$r = \frac{\delta}{\sqrt{2} - 1}$$

Given that $R = \delta + r$, it can be shown that:

$$R = \delta + \frac{\delta}{\sqrt{2} - 1}$$

$$R = (2 + \sqrt{2})\delta$$

$$R \approx 3.4142\delta$$

And therefore:

$$\delta = \frac{1}{2 + \sqrt{2}}R$$

$$\delta \approx 0.2929R$$

From the resulting estimated frequency of all distances, the distance between the 50% of spots closest and furthest away corresponds to the distance δ away from the surface. In other words, 50% of modelled particles will always be less than δ nm away from the surface and the other 50% of distances will always be more than δ nm away from the surface. By using the calculated ratio of $2 + \sqrt{2}$ we can calculate an estimate for the R radius of the average tubular filament (Figure 3.8(c)).

Epigenetic Marker metrics

Distribution relative to density

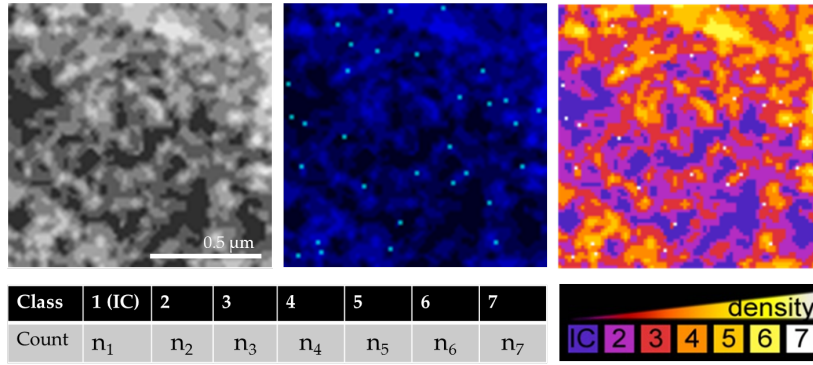
In the ChaiN workflow the 3D coordinates of the centroids of segmented IF spots are mapped on the segmented chromatin landscape (Appendix 9.6). The measurements obtained from this indexing of the distribution of nuclear elements as a function the underlying chromatin landscape allows the quantitation and comparison of any bias in distribution between markers of distinct biological function. The methodology to achieve this begins with the output CSV files of segmented centroid coordinates (Figure 3.6) as well as the segmented chromatin landscape image (Figure 3.7).

The segmented chromatin landscape image is indexed with the IF coordinates to find the corresponding value of the voxel on the image (i.e: the voxel where a centroid coordinate resides). A final sum of all unique values in this new dataset produces a table of the distribution of every centroid position per segmented chromatin class. This can be saved as an individual CSV (Figure 3.9). Each cell may have a different absolute volume of each chromatin intensity class represented, requiring normalisation of the calculated distributions. To normalise the absolute values of spots found in each class Chain divides each by the relative volume of their class; that is, the proportional number of voxels in the segmented image which belong to the same class. The rationale being that for a modelled stochastic spread of coordinates, the number of them caught in each class will increase proportional to the volume of the class (i.e: “a bigger bucket collects more rain drops” analogy). Normalisations as an enrichment or depletion over chance is best plotted as a log-fold change due to the symmetrical magnitude change for positive or negative log values (Figure 3.9(b)).

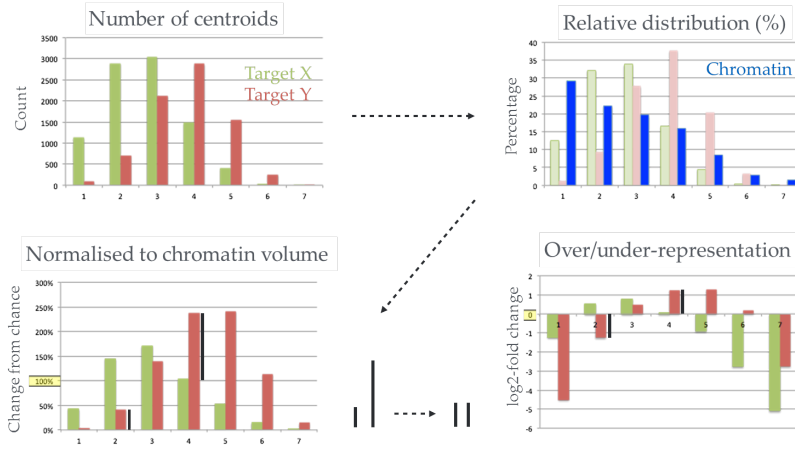
Normalised results are saved to CSVs for quality control of individual cells, but more importantly they are concatenated over the number of technical and biological repeats of each distinct marker required by the experiment. Scripts in Chain will automatically do the latter and calculate the mean of means for every marker as well as the standard deviation and the 95 percent confidence interval (also converted to positive and negative error bars in a log scale for data visualisation). Graphical representation of final outputs as log values provide an intuitive way of comparing by eye the magnitude of enrichment vs depletion by the simple length of bar graph (Figure 3.9(b), bottom right)

Distance to IC border

An important unknown of the epigenomic landscape is if chromatin accessibility (controlled by chromatin density) is a key regulator in the deposition of



(a) Count of marker occurrence in chromatin classes



(b) Normalisation of marker distribution

Figure 3.9: Count and normalisation of marker in chromatin classes. (a) The intensity value at any particular voxel of the 3 bit segmented chromatin landscape (left) can be indexed by the centroid coordinates (centre cyan dots, right white dots). Counting the number of times each chromatin class is indexed (right and bottom table) yields a distribution the most and least common classes where the IF marker is found. (b) The absolute count numbers (top left) are not useful for comparing between different target markers (X and Y, in green and red respectively). Their distribution is normalised to the total number of centroids for each marker in each cell (top right, green and red). As each chromatin class represents a different nuclear volume, a random scatter of centroids would yield an average distribution following the volume of each class (top right, blue) by random chance. The marker distribution is normalised to the random (bottom left) and can be displayed as enrichments ($>100\%$) or depletions ($<100\%$) of the values expected (100%, yellow) by chance Percentage is not ideal to represent change graphically when comparing both enrichment and depletion as depletions from 100-0 % is equal in magnitude to 100-infinity % enrichment. By plotting the log (base 2) of these deviations from chance (bottom right) equal magnitudes of change but in opposite directions (e.g: black bars) can easily be interpreted. Scale bar = $1 \mu\text{m}$.



Figure 3.10: **Distance to IC.** To calculate the average distance of a marker to the IC boundary the same method as that described in Figure 3.8(b) is used, replacing random generated centroids with those empirically calculated from IF and Chain segmentation. Scale bar = 5 μm .

functional epigenetic markers, such as histone modifications. Thus, a way to measure the distribution in distances between where such markers lie to the boundary of the interchromatin compartment would be a useful tool to test this hypothesis.

Chain calculates the distance between the sub-voxel centroid coordinates previously segmented to the nearest voxel of the IC boundary (Figure 3.10 and Appendix 9.6). The mean absolute values in these distributions shed light on the 3D compartmentalisation and deposition of each marker with relation to the chromatin topology. Furthermore, these distance distribution may also be normalised to take into account the probability that such distances could occur by chance alone. This can be done by measuring the nearest neighbour distance between every voxel in the image and the voxels in the IC boundary. However the Chain workflow generates a random set of coordinates indexing any voxel in the image but with a total sample number in the same order of magnitude as the number of marker centroid coordinates. The rationale here being that a stochastic distribution of spots with no biological bias can be used to sample the true number of available distances and so estimate the true statistical nature of said distribution (Figure 3.10) while reducing the magnitude of computations and thus computational time required.

Both the marker distance distribution and random distance distribution

are then fitted to a unimodal density Gaussian function normalised to unity iteratively until optimal parameters are found for each. Using the μ and σ calculated, the expected distribution is fitted between ± 400 nm from the boundary as empirically this range covers the majority of marker coordinates. Finally this continuous function is made discrete by binning into single nm increments.

G_1 3D topology of the genome and epigenome

4.1 Interphase chromatin topology

In order to match the epigenome to the underlying genome, ChaiN must first segment the intricate network of chromatin. As outlined previously, chromatin is stained by SYTOX green after RNase treatment and acquired by 3D SIM during G_1 as this provides a defined window for focusing on epigenetics when most genomic functions are in their relevant 3D physiological state, eg: during metaphase chromatin compaction and loss of the nuclear envelop disrupt normal genomic functions. Cells in G_1 can be selected as opposed to cells in S phase by the absence of incorporated labelled nucleotides, to cells in G_0 by Ki67 stain and to cells in G_2 by their smaller size. Metaphase cells are also easy to select. The ChaiN workflow can, for the first time, generate metrics for a quantitative analysis of the chromatin network environment. These can be broken down to measurements in densities, volumes, surface areas and filament dimensions; descriptions of the network which can later be used to compare chromatin after remodelling by endogenous or exogenous factors.

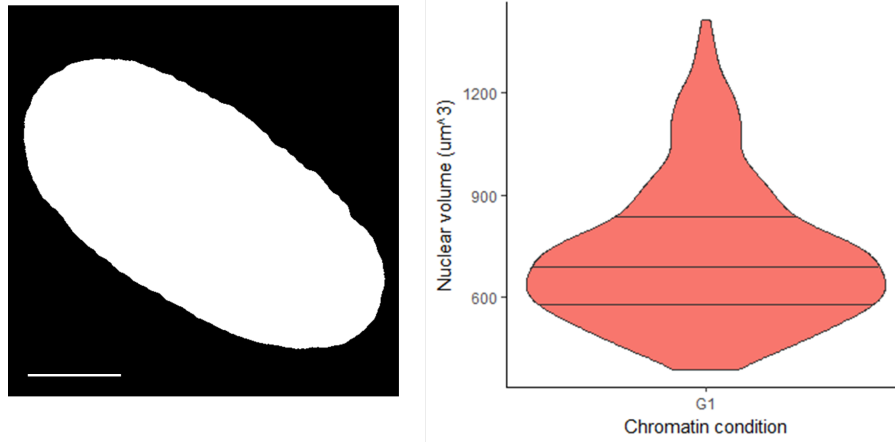


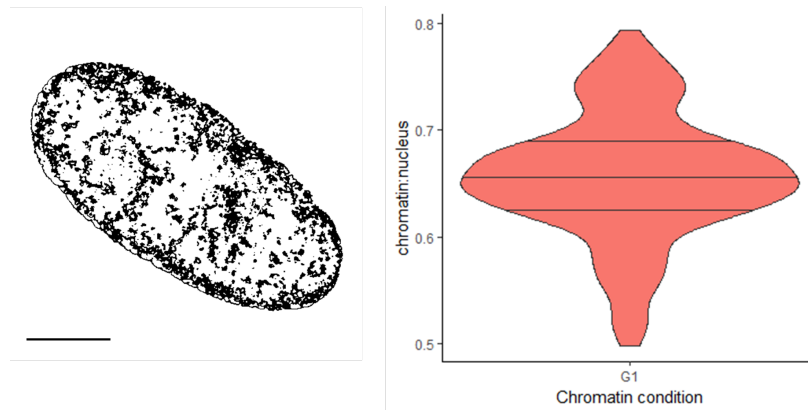
Figure 4.1: **Chromatin volume.** Absolute nuclear volume as voxel counts from the binary nuclear mask (left). Violin plot showing quartiles as horizontal lines for non-treated G_1 C127 nuclei. Scale bar = $5 \mu\text{m}$, $N_{\text{total}} = 320$ cells.

Volumes

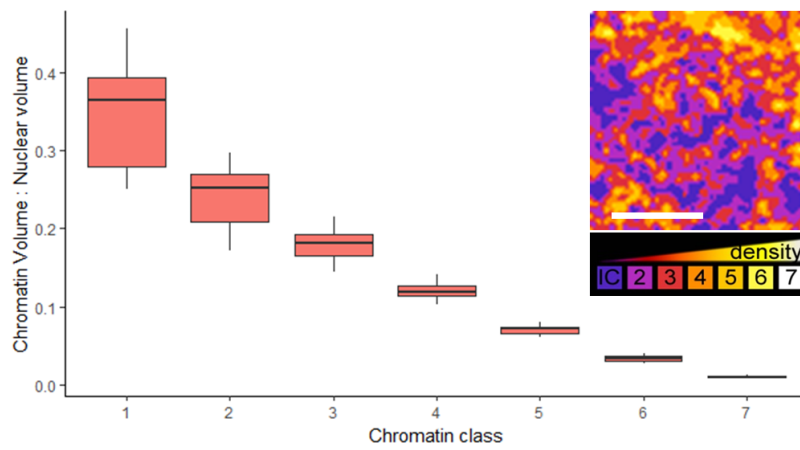
After the advent of deconvolution algorithms, followed by super resolution microscopy, the nucleus has been known to consist of a heterogeneous chromatin environment. Using ChaiN the total volume of the nucleus (Figure 4.1) and the volume dedicated to chromatin can be calculated (Figure 4.2(a)). For the C127 cell line approximately 65% of the nuclear volume is comprised of chromatin, with a range between 50% and 80%.

Density

The volume of chromatin is also heterogonous with respect to chromatin density with some percentage volume of the total chromatin dedicated to various amounts of chromatin compaction. Using a 7 bin classification outlined in the ChaiN methodology (3.2) a negative correlation can be clearly seen between increased density of chromatin and its proportion of nuclear volume (Figure 4.2(b)). This points to a significant reduction in volume occupied by increasing densities of chromatin, as would be expected from a network composed of a gradient of densities with most of its volume dedicated to a decondensed surface and a smaller core with condensed chromatin (Figure 1.13).



(a) chromatin:nucleus



(b) Proportion by density class

Figure 4.2: **Chromatin density.** (a) Ratio of chromatin volume (represented as a binary image of density classes 2-7, left) to whole nuclear volume plotted as a violin plot of distribution for untreated G₁ C127 nuclei (right). Scale bar = 5 μ m. (b) Distribution of (a) sub divided by chromatin density class, plotted as boxplot. Scale bar = 500 nm, N_{total} = 320 cells.

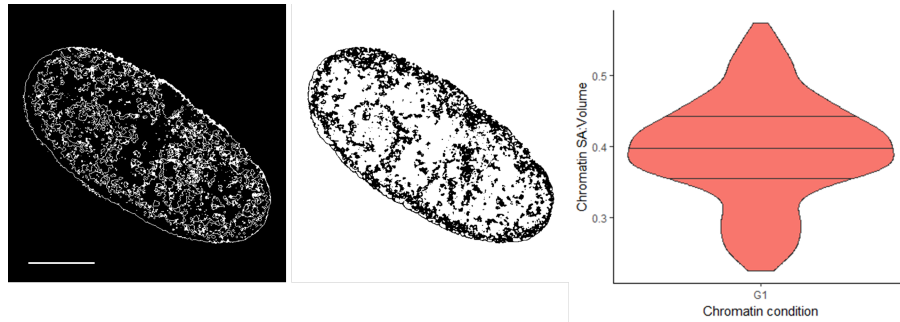


Figure 4.3: **Chromatin Surface area.** Surface area to volume ratio calculated from the voxel count of boundary voxels between class 1 and 2 (left) and the chromatin voxels of classes 2-7 (centre). Scale bar = 5 μm , N_{total} = 320 cells.

Surface area

From the segmentation of the network surface between chromatin and non-chromatin it is also possible to calculate a surface area to chromatin volume ratio. A range of ratios from approximately 0.2-0.6 is found, with a physiological average of 0.4 (Figure 4.3). Although these values are relative and don't mean much for a single experiment type, this metric is a useful tool to understand how complex and convoluted the network is, and to compare how this complexity changes from factors which can remodel chromatin. This is important for the accessibility-based regulation hypothesis as the surface area would be the site of interaction between machinery in the IC and chromatin to be acted upon.

Network dimension

Finally an absolute measurement of the network's filaments can be extracted by considering the distance between a set of random points within chromatin and the segmented surface. The average dimensions can be obtained by modelling random distances as a decay from the surface (assuming that the average filament can be model as cylindrical) and multiplying by a correction factor. This is outlined in the ChaiN methodology (§3.3). The average filament radius calculated is 263 nm, giving a filament width of 526 nm (Figure 4.4).

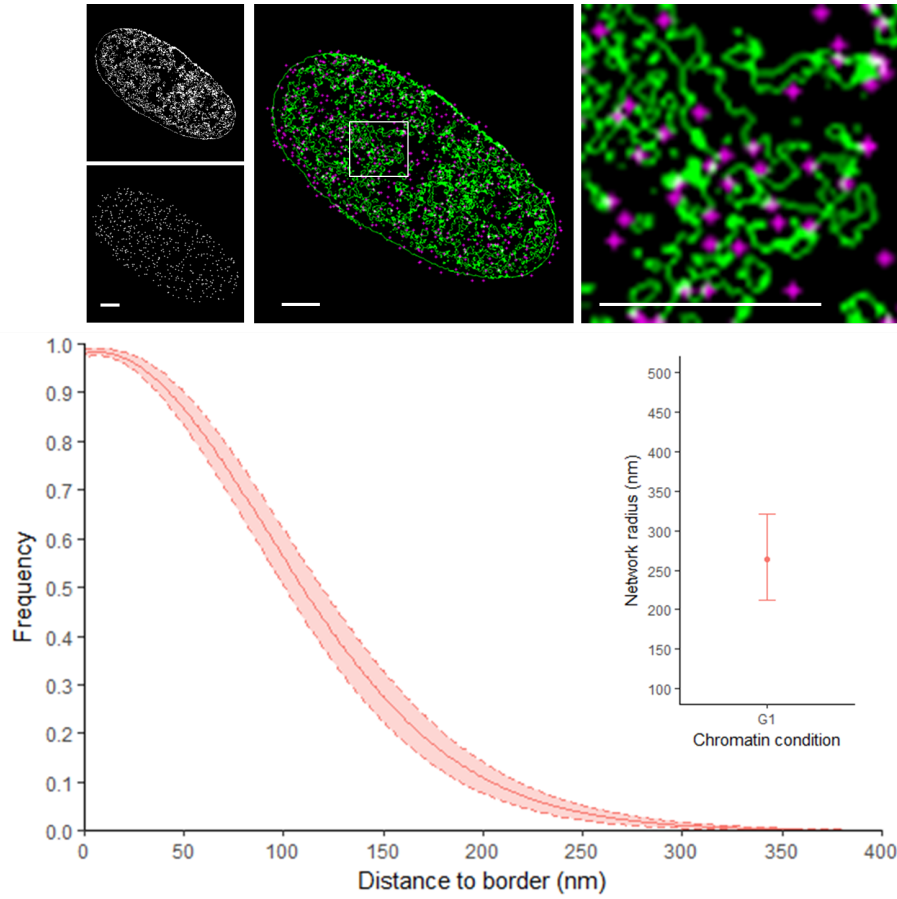


Figure 4.4: **Chromatin Network.** Top gallery: Dimensions of chromatin network calculated from the distances of random sampled voxels (magenta) to the network boundary (green) as outlined in §3.3. Bottom: The average of the modelled frequencies fitted for each cell at each nm distance to the IC boundary can be plotted (red line) with its 95% confidence (ribbon between dashed lines). Modelling the network as tubular filaments as described in §3.3 allows the calculation of the average network filament radius at 263 nm, giving a filament width of 526 nm. Scale bars = 2.5 μm , N_{total} = 960 cells.

4.2 The 3D epigenome

Correlating the spatial distribution of epigenetic markers as identified by IF to the density of underlying chromatin in C127 mammalian G_1 cells as described in §3.3 yields a rich dataset of either enrichment or depletion of the abundance of these IF signals for each binned chromatin density class. By plotting this dataset as a heatmap of log2-fold enrichment and grouping markers by *a priori* functional knowledge a clear distinction of functional compartmentalisation can be observed (Figure 4.5). Generally, the distribution of structural markers and markers of transcriptional activity are found in the lower density classes while the inverse is true for markers of genomic inactivity. It is worth noting that throughout the profiles of all markers analysed, class 1 (defined as the chromatin devoid IC) is evidently depleted of all chromatin associating markers (Figure 4.5), making it a good internal control for the intensity binning classification (§3.2). The group of markers most enriched at decondensed chromatin classes (the structural elements and the markers for active transcription) are still under-represented in the IC (albeit to a lesser degree than depletion observed for markers of heterochromatin in the same class 1).

Even for structural and active markers a gradient of correlation with IC occupancy can be seen when sorted by the degree of chromatin interaction. Histones with post translational modifications show the greatest depletion in the IC, followed by the non-histone transcription complex RNAP, then by non-histone chromatin architecture factors, and finally non-histones with chromatin and RNA association. Furthermore plotting the relationship between the size of chromatin remodellers and the chromatin class where the marker deposited by each remodeller is most highly enriched gives a strong positive correlation (Figure 4.6). Individual distributions will be analysed in detail for each marker of each functional sub group subsequently.

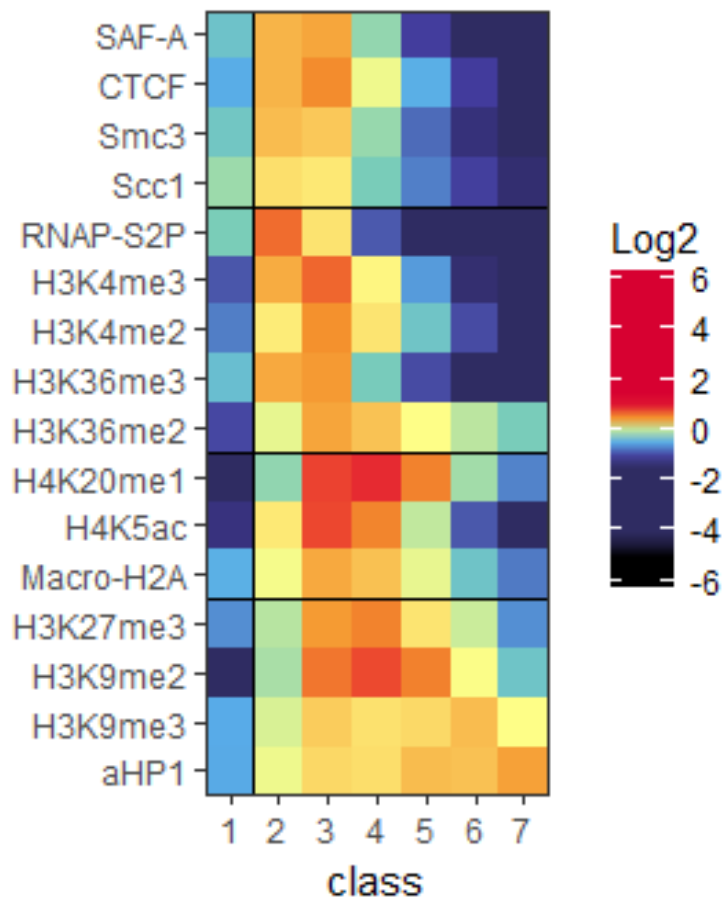


Figure 4.5: **Active marker distribution.** Overview heatmaps of log2 fold enrichment or depletion of marker IF abundance relative to chance in each chromatin class for all markers. Markers are grouped by function from top to bottom: Structural, Active, Mid, Inactive. $N_{total}=320$ cells.

Decreased chromatin density correlates with transcriptional active environments

The clearest pattern to emerge is the relationship between markers of transcriptional activity and regions of low chromatin density (Figure 4.7). The serine-2-phosphorylated modified RNAPII(-S2P) being a marker for transcriptional elongation (Ahn et al., 2004) along with histone modifications associated with active promoters (histone 3 lysine 4 di/tri methylated, H3K4me2/3, Ruthenburg et al. (2007)) and active gene bodies (H3K36me2/3,

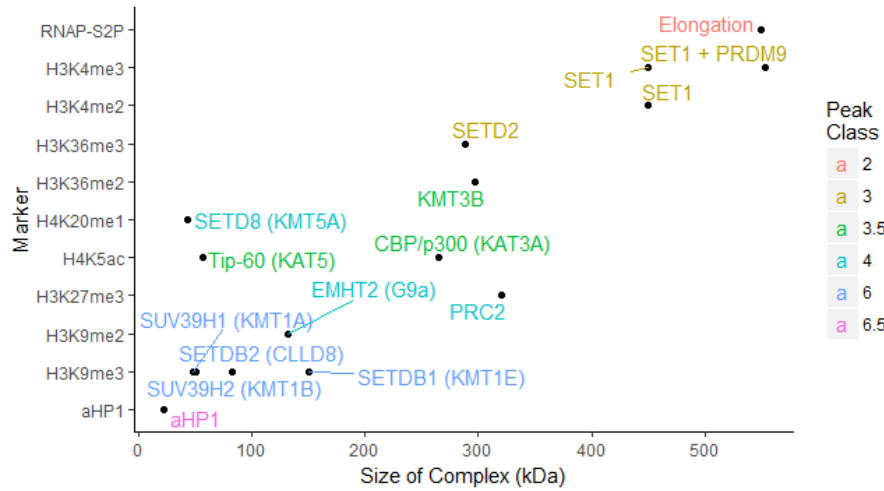


Figure 4.6: **Chromatin remodellers.** Scatter plot of the approximate sizes for chromatin remodeller complexes (individually labelled) responsible for the deposition of the chromatin markers in the y axis. Colour indicates the segmented chromatin class where the marker is found at highest enrichment (decimals used to extrapolate the mean between two classes if both show comparable enrichment).

Wagner and Carpenter (2012)) were found to be enriched primarily between classes 2-4 and massively under-represented in classes 5-7. The under-representation in these denser classes can reach up to a factor of 10 times lower than expected from chance alone.

The use of thousands of immunofluorescence spots for measurements and replicates over tens of cells gives high enough statistical power to decrease the range of confidence intervals to show clear separation between bars belonging to the same marker across densities and bars of the same chromatin density compared across markers. Whilst the general pattern for “active” function is not surprising, it is worth noting that for the first time, a quantitative analysis can be performed with sufficient statistical power to resolve the difference between markers of the same function. That is to say that the distribution of specific markers can be described, and not just the general pattern of a certain genomic activity.

From this dataset it can be observed that the epigenetic marker with least

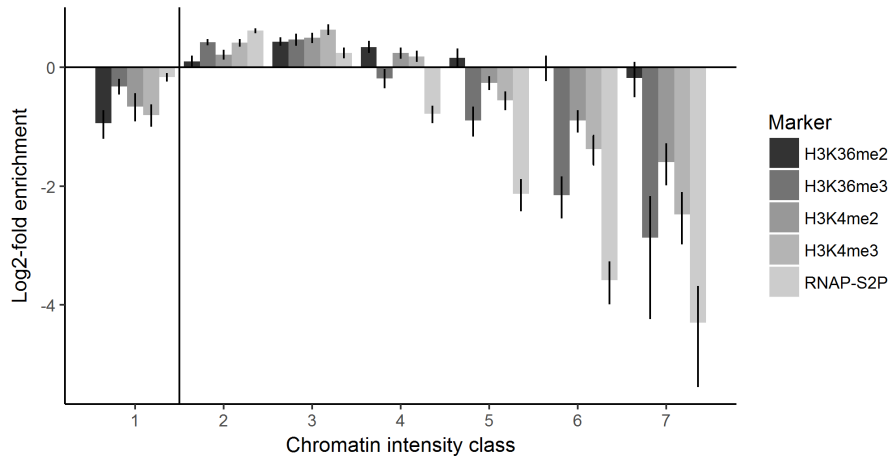


Figure 4.7: **Active marker distribution.** log2 fold enrichment and depletion of IF marker abundance for active genome function relative to chance in each chromatin class. Error bars = 95% Confidence interval, $N_{total}=100$ cells.

depletion in class 1, highest enrichment in the lowest densities (classes 2-3) and strongest depletion in classes 4-7 is the RNA polymerase complex (Figure 4.7). This agrees with the model, being a marker which by definition must be associated with transcription but also being the only marker of the five analysed not to be part of chromatin. One would expect it more than the rest in class 1, the IC, by chance alone, and one would expect more nucleosomes, regardless of modification as the density of chromatin increases. Furthermore, its large molecular weight (12 subunits comprising 550 kDa) makes it physically adverse to penetrate high density chromatin as hypothesised by the proposed model. RNAPII-S2P is in fact the benchmark used here for defining without histone proxy what chromatin density environment is the one in which active transcriptional elongation takes place, nominally classes 2-3.

Two further patterns of interest that the ChaiN workflow can extract are the striking difference between methylated states of histone modifications and also the difference between modifications on promoters vs gene bodies (H3K4 and H3K36 methylation, respectively). Following the former statement both tri-methylated states of either lysine 4 or 36 are the ones most depleted in dense chromatin compared to their di-methyl counterparts. This is

also in line with the accessibility model in that the deposition of such methyl groups is accomplished via enzymatic reactions carried out by large methyltransferase complexes, with the tri-methyl state proceeding only after mono- and di-methylation. For H3K4, methylation takes place by complexes containing the SET domain (Gu and Lee, 2013) such as the SET1 complex (450 kDa) and PRDM9 (103 kDa); though smaller, PRDM9 cannot initiate mono- or di-methylation but only performs tri-methylation on di-methyl sites, making the size of the SET1 complex the limiting factor for H3K4me3 deposition.

For H3K36 the KMT3B (297 kDa) and SETD2 complex (288 kDa) are responsible for its methylation. As stated this increased depletion with additional methyl group follows the predictions of the model as tri-methylation requires prior methylation to occur, and increased density would decrease accessibility of the large methyltransferase complexes.

Furthermore, it is known that H3K36 methylation distribution correlates with the anatomy of transcribed genes, with mono-methylation at promoters and increasing levels of methylation from 5' to the 3' end (Barski et al., 2007). From the ChaiN analysis results, H3K36me2 decreases in enrichment back to levels close to chance alone as it moves from classes 3-7, yet H3K36me3 is the highest depleted histone modification in classes 5-7, second only to RNAPII-S2P. This extreme depletion pattern mirroring RNAPII-S2P is backed up by results from published data in yeast, (Carrozza et al., 2005; Joshi and Struhl, 2005) describing that the deposition of H3K36 methylation occurs on nucleosomes after they have been displaced by the transcription machinery, with evidence to the same mechanism in mammals. Hence, even if the extreme low levels of RNAPII-S2P in denser chromatin may produce some leaky levels of transcription initiation from the also few actively marked promoters, the low local concentration of enzyme and substrate, coupled with high viscosity of dense chromatin may ensure that transcription stalls. This would prevent RNAPII-S2P from reaching the 3' end of the gene body more than the 5' decreasing tri-methylation with respect to di-methylation of H3K36, as observed

in this analysis.

Finally the differences observed between H3K4 and H3K36 may also be explained in terms of methyltransferase accessibility and the order of transcription. H3K4me2/3 are both depleted more than H3K36me2 as chromatin density increases, but they are also deposited by a complex of almost twice the molecular weight, which would have a decreased accessibility. On the other hand, they are depleted less than H3K36me3. But as shown previously the extreme depletion of H3K36me3 is a possible compounded effect of the extreme depletion of RNAPII-S2P at these densities, a complex even larger than H3K4me2/3 methyltransferases.

Transcriptional inactivity has a non-linear positive correlation with increased chromatin density

In contrast to the pattern described in the previous section, markers of silenced chromatin (H3K9me2/3, H3K27me3 and alpha-Heterochromatin Protein 1 - HP1A) (Lehnertz et al., 2003; Barski et al., 2007) are enriched between classes of denser chromatin, 3-6 (Figure 4.8). At first glance this might not appear novel, given that traditionally silenced chromatin has been characterised by higher density chromatin, and termed heterochromatin. It is thus a logical conclusion to find HP1A in traditionally defined heterochromatin. After all as its name suggest the main function of HP1A is transcriptional repression and maintenance of constitutive heterochromatin environments. According to the accessibility model too, HP1A a 22 kDa protein should indeed have far higher access to dense chromatin than complexes ten times its size or more.

However, with the exception of HP1A, the other markers for inactivity analysed here are not massively enriched at the higher density chromatin but are in fact mostly under-represented in class 7, albeit to a much lesser degree than the depletion of active histone markers and RNAPII-S2P discussed previously. This suggests that despite an expected depletion of transcriptional activity markers and an expected enrichment of silencing markers with in-

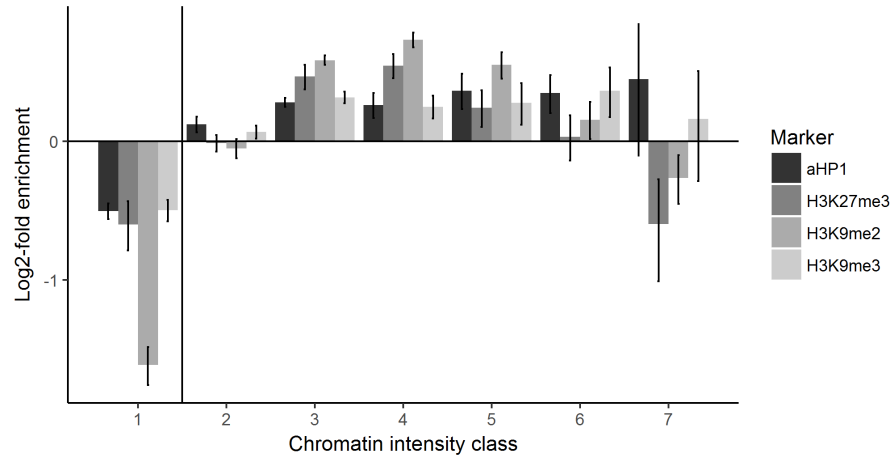


Figure 4.8: **Inactive marker distribution.** log2 fold enrichment and depletion of IF marker abundance for inactive genome function relative to chance in each chromatin class. Error bars = 95% Confidence interval, $N_{total} = 80$ cells.

creased density, the regions of densest chromatin (class 7) has fewer markers for inactivity than predicted. It is only in this study at the limits of optical super resolution and with high throughput image analysis that such a quantitation has been possible for the first time. One could envision that at the highest density, chromatin might be mostly inaccessible to immuno-detection protocols. This however is countered by the distribution profile of HP1A, which shows over most of its class 7 distribution a clear enrichment of almost up to 2x more than by chance alone (with a mean of at least 50% increase approximately). Given that any variation in the molecular weight of IgG antibodies is only in the order of tens of Daltons this marker works as a good internal control for accessibility of antibodies in dense chromatin and therefore suggests that for the other markers there is indeed a true depletion of epitopes in class 7 than expected from chance or predicted by previous studies.

For the histone modifications, this backs up the hypothesis that the individual methyltransferases responsible for the deposition of such marks are unable to or at the very least strongly hindered in their ability to access this chromatin environment. The methyltransferase EMHT2 (G9a, 132 kDa) catalyses the mono- and di-methylation of H3K9, followed by either set of

complexes: SETDB1/2 (KMT1E,CLLD8; 150 kDa, 82 kDa respectively) and SUV39H1/2 (KMT1A/KMT1B; 48 kDa, 50 kDa respectively) catalysing the trimethyl state. Most of these complexes are at least half if not 4 times smaller than the size of those for active markers, adding weight to the accessibility hypothesis. On top of this, the smaller sized complexes within this group also match the highest enrichment (or least levels of depletion) at denser chromatin levels. Furthermore, it can be observed that the density distribution profile of H3K9me3 is the only one which consistently and robustly mimics that of HP1A in all classes. This could be attributed to the fact that H3K9me3 is known to bind HP1A (Lehnertz et al., 2003) in constitutive heterochromatin.

H3K9me2 is also a known marker for the inactive X chromosome (Xi). Recent studies with super resolution microscopy have found that markers of Xi repression act not on the densest part of the chromosome but as a coating/surfactant on mid level chromatin density (Smeets et al., 2014). In fact, most of the Xi has been reported to be less dense than originally hypothesised (Smeets et al., 2014). The interactions with HP1A or with the Xi for H3K9me3 or -me2 respectively might be contributing factors in the differences between the distribution of these markers.

The tri-methylation of H3K27 is only known to be brought about by the EZH2 protein (KMT6A, 85 kDa)(Kuzmichev et al., 2002). According to the model, a protein of this size could penetrate to deep chromatin yielding a density distribution closer to that of H3K9me3. H3K27me3 however is actually deposited by EZH2 as part of a 4 core subunit complex: the Polycomb Repressive Complex 2 (PRC2) at a molecular weight of 320 kDa, and with interaction to many other subunits (Margueron and Reinberg, 2011; Tie et al., 2003). The PRC2 has also been characterised from precipitation as a functional dimer, and *in vitro* reconstitutions have confirmed this dimerisation (plus additional subunits) to a complex of 670 kDa in size (Davidovich et al., 2014). Clearly the large size of the functional PRC2 falls within the range of complexes hypothesised to suffer a large decreased accessibility to high density

chromatin, the observed distribution from the ChaiN

Moreover the fact that the H3K27me3 distribution mimics the active markers more than repressive ones may also have to do with the genomic target of the H3K27me3 induced silencing. This mark is deposited at gene promoters, as opposed to H3K9me3 which can also be heavily enriched at non-genic regions (Barski et al., 2007). Current literature also describes the existence of "bivalent domains". These are sites where H3K27me3 is found on the same promoters as active marks, such as H3K4me3. Bivalent domains often encompass key developmental genes (Bernstein et al., 2006). Whilst there is still much debate as to the functional significance of this bivalency, it is thought that such regions may aid in keeping important developmental promoters "poised" for transcriptional activity, while still keeping them silent (Voigt et al., 2013). This hypothesis is in line with the model proposed in this study. The transcriptional environment of genes seem to be restricted between classes 2 and 3 (RNAPII-S2P enrichment). If the role of H3K27me3 is to silence promoters from RNAP processivity, then H3K27 is only required in the chromatin density environments where transcription may still take place. H3K27me3 does not need to be present in classes 6 or 7 as the replication machinery is mostly unable to traverse any density equal or higher than class 4.

Mid-density chromatin correlates with environments of other non-transcriptional functions

Interestingly, markers of multiple function (transcription state plus other functions not directly associated with ongoing transcription, such as bookkeeping of cellular-state-specific active genes or DNA repair) showed a broad distribution profile. These have a range of enrichment between classes 2-5 overlapping most of the range of both active and inactive transcription markers (Figure 4.9). Their average peak enrichment is at class 3, close to the primary markers of active promoters and gene bodies, H3K4me2 and H3K36me2 respectively. Yet, class 2 seems almost non-enriched in comparison. This obser-

vation may be used together with previous profiles to suggest that the majority of gene-based activities (taking place on either active or inactive genes regardless of transcriptional state), generally occurs in classes 2-5. Using the markers discussed in previous sections, the spectrum within this density range changes from active transcription towards the lower density classes 2-4 and bookmark/facultative inactive genes or chromatin repair generally in the mid-range, in classes 3-5 and 3-6 respectively.

Regarding the accessibility paradigm the complexes depositing the markers are SETD8 only (KMT5A, 43 kDa) for H4K20me1, and either Tip-60 (KAT5, 56 kDa) or CBP/p300 (KAT3A, 265 kDa) for H4K5ac (Beck et al., 2012; Kimura and Horikoshi, 1998; Schiltz et al., 1999). These complexes, with the exception of CBP, are small enough according to the model to extend past markers for activation and into denser chromatin, as observed. Given the small size one might expect that the distribution would be further skewed to dense chromatin than what is observed. If the function of these markers were unique to non-transcriptional processes, that might indeed be the case. However, many of these markers have certain ambiguity in their use as histone code. For example H4K20me1 is both associated with the promoters of the most highly transcribed group of genes (Wang et al., 2008) but also acting cooperatively with H3K9 to demarcate silent genome regions (Sims et al., 2006). As with most acetylations, H4K5ac can be interpreted as a sign for open euchromatin. H4K5 has in fact no other modification save its acetylation, and CBP is a known transcriptional activator. Yet, despite this clear evidence, H4K5ac has also been reported to serve as an epigenetic bookmark together with HK12ac for priming transcription after cell division (Zhao et al., 2011). This marker then, whilst strongly associated with transcription at some stages, is also a crucial bookmark of genes in denser chromatin, possibly giving rise to its observable distribution here.

MacroH2A is not a histone PTM but a histone variant encoded by the H2AFY gene. In this study, MacroH2A displays low depletion in classes 6-7

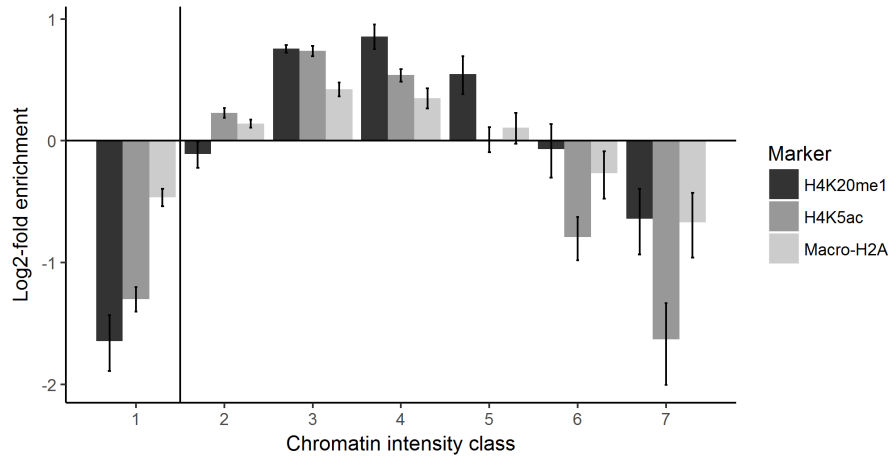


Figure 4.9: **Mid marker distribution.** log2 fold enrichment and depletion of IF marker abundance for multiple non-transcriptional exclusive genome function relative to chance in each chromatin class. Error bars = 95% Confidence interval, $N_{total}=60$ cells.

and detectable but low enrichment in classes 3-4. MacroH2A's first known function was suspected from early fluorescence microscopy, when it was shown to have an enrichment in the inactive X (Costanzi and Pehrson, 1998). Since, it has also been shown to have generally repressive effects on chromatin environments outside the Xi. This is in line with its distribution, being found in both male and females to the same degree as well as in other vertebrates that lack establishment of an Xi (Rasmussen et al., 1999; Buschbeck and Di Croce, 2010). There are however cell-differentiation specific genes which show positive upregulation with macroH2A interaction (Gamble and Kraus, 2010). With much of its mechanism still unknown, and with evidence to support the involvement of macroH2A in functions whose environments are characterised on either side of the density spectrum; it is possible to interpret the symmetrical distribution of this marker between classes 2-5 as a combination of sub populations of macroH2A. Some partly involved in Xi and autosomal chromosome repression between classes 3-5, and others in gene specific activation between classes 2-3.

Key genome architectural proteins correlate to chromatin environments of active transcription

One of the most remarkable observations to come from this analysis are the density distributions for the structural proteins of genome architecture. Protein subunits of cohesin (Smc3, Scc1), the insulator CTCF and the nuclear scaffold protein SAF-A (heterogeneous nuclear ribonucleoprotein U, hnRNP-U) showed a clear if subtle over-representation between classes 2-3 (Figure 4.10) coupled to a sharp under-representation correlated with increased density from classes 4-7. Of all markers analysed this distribution is only matched in magnitude and pattern by RNAPII-S2P and the H3K4/H3K36 methyl modifications, all the markers for active transcription. Both cohesin and CTCF have been repeatedly shown to be key regulators of TAD structures (Ji et al., 2016). Some such studies have suggested models whereby a combination of cohesin and CTCF may act on these structures by a pulling/cohesive forces holding chromatin domains from within, whilst others picture cohesin and CTCF acting as single “staples” joining the two ends of a TAD from the outside (like a tied balloon). The results from this investigation seem mostly in agreement with the latter hypothesis. Some cohesin and CTCFs may be found within chromatin of higher density, yet their environment of peak enrichment lies predominantly at less dense regions, overlapping with active transcription. The striking under-representation of both cohesin and CTCF in chromatin past halfway in the density scale, beyond class 4, points to alternative mechanisms of chromatin architecture maintenance for chromatin in dense environments. The data seems suggestive that if such mechanisms exist, they may likely be linked to HP1A or its binding partners, given its demonstrated enrichment distribution in class 7. A definitive proof of a cohesin and CTCF independent TAD maintenance in single cells would need to be assayed through ChaiN by correlative single-TAD FISH and respective IF stains; solely for TADs inhabiting classes 6-7.

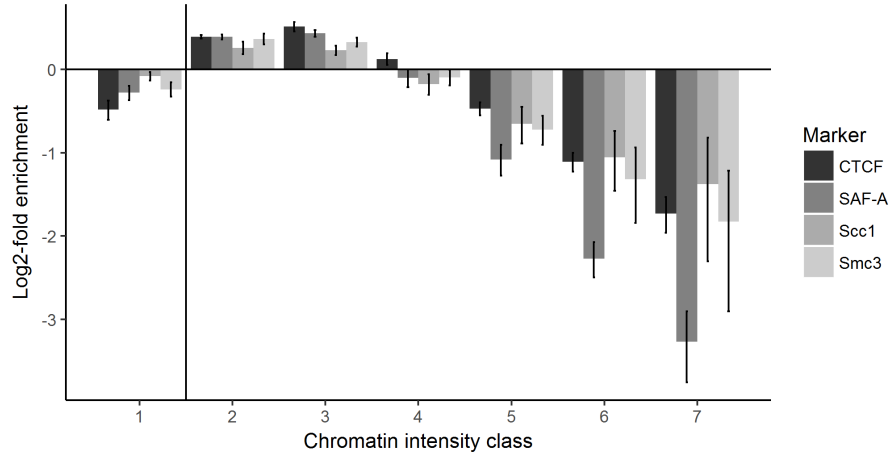


Figure 4.10: **Structural marker distribution.** log2 fold enrichment and depletion of IF marker abundance for genome structural proteins relative to chance in each chromatin class. Error bars = 95% Confidence interval, $N_{total}=80$ cells.

The same pattern described for cohesin is also observed for SAF-A. While SAF-A shows a similar level of enrichment in classes 2 and 3, it is clearly the most depleted of all the structural proteins assayed here for density classes 5-7. In almost all these cases SAF-A shows approximately twice the level of depletion as the average of the others combined, being twice, 4 times or 8 times less populous in classes 5, 6, and 7 respectively, than predicted from chance alone. As previously mentioned SAF-A is a nuclear scaffold protein (Romig et al., 1992). In this study SAF-A is predominantly found at the perichromatin surface, classes 2 and 3, the interphase state of chromatin between the chromatin network and the IC channels. This characterisation is in agreement with its known interaction to H3K4me2/3 and H3K36me2/3 marked chromatin (Puvvula et al., 2014) as well as its role in mediating post-transcriptional mRNA splicing (Xiao et al., 2012). Given that transcription of RNA occurs between classes 2-3 and the region of H3K4 and H3K36 di- and tri-methylation enrichment are also between classes 2-3 it is logical to find SAF-A associated with this environment. SAF-A is of course less depleted in the non-chromatin class 1 as it is not by definition part of chromatin, only associated with some of its functions, and it is also involved in RNA process-

ing. RNAs have been shown to be enriched primarily in the IC, this is also expected as the IC is where RNA processing machineries reside, as well as providing chromatin devoid channels to the nuclear pore complexes for RNA export (Markaki et al., 2010).

Depth of epigenome marker deposition

Using the segmented chromatin boundary between classes 1 and 2 and the sub-voxel position of centroids for each marker, ChaiN is able to output statistics on the average distance between where markers are enriched and the chromatin network surface (Figure 4.11). The results from this analysis are not as clearly defined as the correlation of markers with the chromatin density in their environment. Despite a sample of 20 cells analysed per marker, or 320 cells in total, with an average spot number of 7300 per cell, the CI95 error bars are still wide and overlapping. Nevertheless, what can be seen is that for all markers analysed, 95% of their distribution lie somewhere between -5.8 nm and 57.7 nm from the surface (taking all CI95s on average). In other words, of a network with average filaments measuring 526 nm $\pm$ ~ 100 nm, 95% of observations lie from ~ 6 nm outside the network to a limited ~ 60 nm inside the surface.

Another interesting observation can be obtained by using *a priori* knowledge, grouping results by function and taking their average. From this, structural markers show an average of 18 nm, active markers would show 23 nm, and both ambiguous markers and markers of inactivity would show 33 nm; from within the chromatin network to the surface. While the increase in distance for these functional groups is positively correlated with the increased density distribution for these groups, little else can be said given the current statistical power.

This is as much a limitation of the sample size, as a technical barrier. As described in Appendix 9.2, using SIM with a $60\times$ magnification as in this study, the pixel size accounts for 41 nm in both x and y and 125 nm in the z dimen-

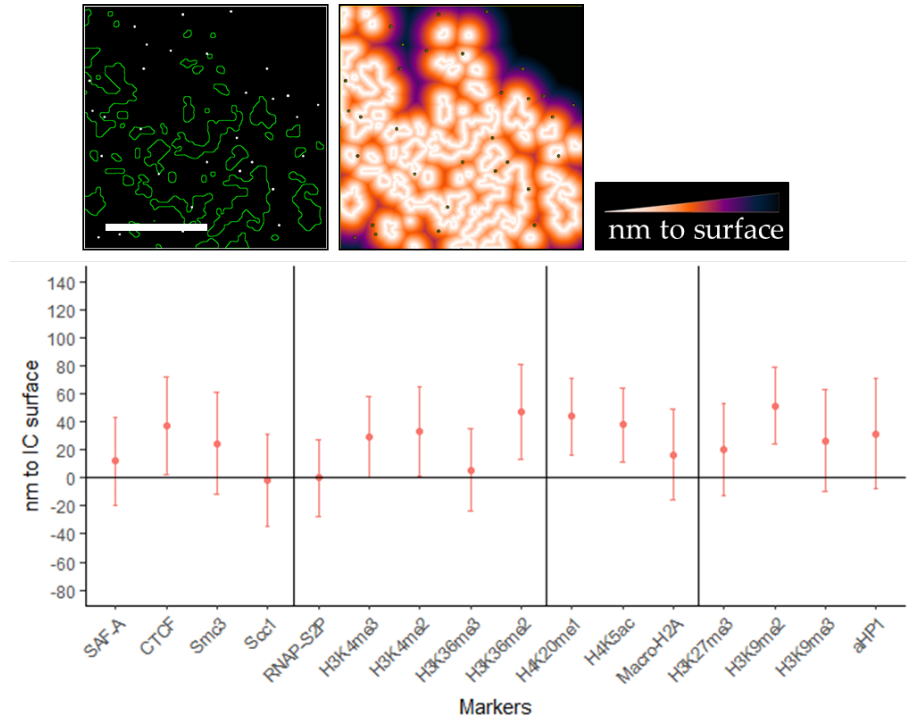


Figure 4.11: **Distance to IC.** Calculated mean sub-voxel distance of all IF markers to the IC. Positive distance is equivalent to marks within the chromatin network. Markers are grouped by function from left to right: Structural, Active, Mid, Inactive. Bars show 95% confidence, scale bar = 5 μm , $N_{\text{total}} = 320$ cells.

sion, giving a maximum voxel size uncertainty of 137.8 nm, and an error of half this value. With such small distances and differences between these, and comparably large voxel dimensions it is in fact remarkable that the CI95 for these results is only on average ± 31.8 nm with only 20 cells for each marker (Figure 4.11). This shows that measurements between cells of the same marker must have a close clustering around the mean value of the marker, yielding a small standard error (SE) when derived (Appendix 9.2). Increasing the sample size would therefore be a valid strategy to decrease the CI95.

Replication-induced local perturbations in S phase

Replication induces local 3D genomic rearrangements

The process of replication of the genomic content of a cell, like transcription and other chromatin associated processes, requires the formation of large complexes. According to the accessibility model explored so far, replication might be expected to suffer the same reduced accessibility to dense chromatin as with these other genome functions. This imposed restriction may be true to a certain extent but must also be overcome eventually.

It is known from studies using conventional microscopy that replication in single cells takes place in a spatially segregated order. During S phase, the first regions of the genome to be replicated are those which are away from both the nuclear periphery and chromocentres (Figure 5.1)(Schermelleh et al., 2001). These regions, at the confocal microscopy level, are co-localised with an enrichment of active chromatin (euchromatin), such as H3K4me3. As S phase progresses to its middle stages, sites of replication become enriched at the nuclear and nucleolar periphery (Figure 5.1, mid), and towards the end of S

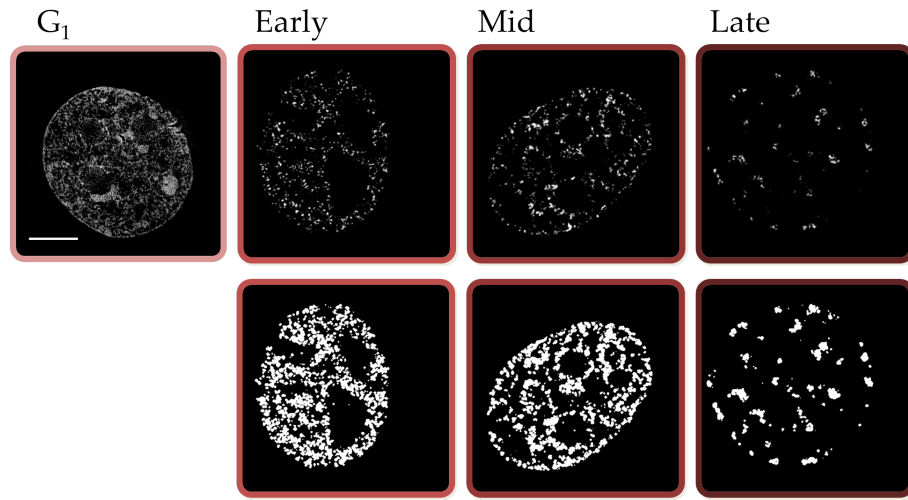


Figure 5.1: **Sphase stages.** Top row: Patterns of S phase progression from early mid and late stages of S phase (labelled from EdU incorporation) compared to whole chromatin in G_1 (counterstained with Sytox, left). Lower row: Representative segmentation masks.

phase the densest parts of chromocentres are replicated (Figure 5.1, late). This order of replication starts with the least condensed chromatin and progresses through with increased density. In fact even in late S phase, chromocentres can be observed to be replicated from their outside first.

Though replication may take advantage of readily available origins of replication found in pre-decondensed chromatin (from transcriptionally active environments), it does not stop solely at these regions but continues on into all available chromatin. This is particularly interesting in traditionally defined "euchromatin" regions, away from the nuclear and nucleolar periphery. In any given cell, there is not enough surface area at these peripheries (or volume in the chromocentres) to hide all inactive chromatin. Furthermore studies using chromosome painting reveals that despite some preferences, any chromosome may find itself closer or further away from any of these environments in any given cell (Croft et al., 1999). For any given cell, chromosomes which are not next to these macro heterochromatic structures, but which have inactive chromatin to pack away, must still regulate how ac-

cessible their chromatin is in a generally euchromatic environment. From the proposed accessibility model this can be achieved by protecting sequences from unwanted transcription by surrounding them with chromatin whose transcription is preferable.

What this means in terms of replication, is that at the point where the replication machinery progresses away from the low density origins, it will plough through this established packaging order. Thus, instead of density and accessibility regulating what can be replicated, the act of replication regulates the density of the chromatin surrounding the process. What is hypothesised from such mechanism would be a local chromatin decompaction surrounding foci of active replication, occurring at all stages of S phase.

5.1 Local chromatin decompaction

In this study ChaiN was applied to characterise the local chromatin decompaction around active DNA replication if any, and if so its magnitude with respect to cell cycle stage. For this we utilised a third staining of DNA, polymerase-incorporated ethynyl deoxyuridine (EdU) subsequently tagged via Click chemistry to robust organic azide-coupled fluorophores (Kolb et al., 2001). The methodology is described in detail in Miron et al. (2016) as well as in §2.4. This was made into a 3D binary S phase region-of-origin (ROI) mask, narrowing the screen volume to the immediate chromatin neighbourhood surrounding regions of active replication.

The results clearly back the known progression of replication from generally less dense chromatin to denser chromatin environments, but this time with quantifiable data. The described progression is evident from the dramatic shift in the peak abundant class from G_1 through to early, mid and late S phase (ES, MS, LS, respectively) shown in Figure 5.2 as a proportion of the chromatin density class over the masked environment. This normalisation is done on a per-cell basis as each cell will have a different total volume of chro-

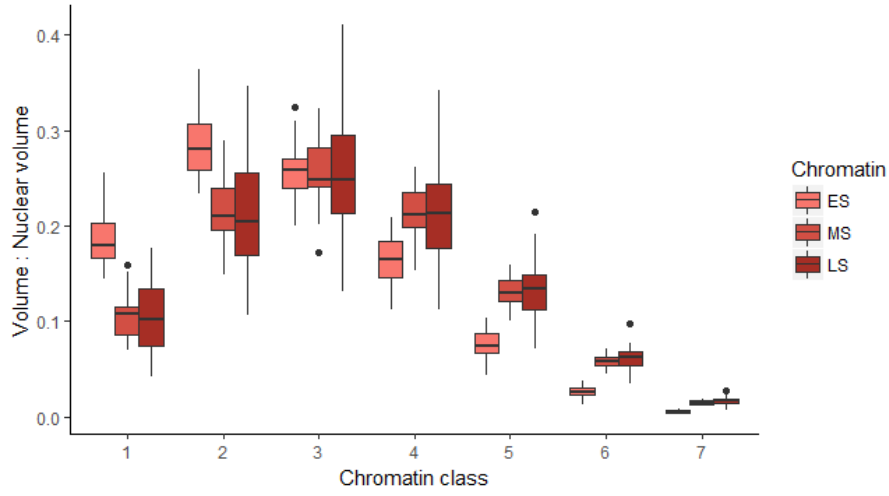


Figure 5.2: **S phase density per class.** Proportion of chromatin density represented at each stage in S phase for each chromatin class. $N_{total}=800$ cells.

matin in the masked regions. In G_1 the mode class is the IC, class 1, with 37% of the nuclear volume, the proportion of the nuclear volume take by any other class decreases linearly with higher class number

In early S phase this distribution is already changed. In the chromatin surrounding replication in ES there is a significant drop in the proportion of class 1, and a new peak at class 2 is established. This is closely followed by class 3, with significant elevated distribution above G_1 , and it is not until class 5 that the proportion drops to that observed from G_1 proportions. This agrees with the view that the most readily available origins of replication are available because they happen to be in the chromatin density environment of choice for accessibility by transcription machinery (classes 2-3). The drop in class 1 is explainable by the very definition of the class having no detectable chromatin, and as such replication would not be present in regions of the nucleus with chromatin to replicate.

During mid S phase the peak shifts even further to the right, to class 3, and the deviation above G_1 remains significantly large for classes 4-6. Already at MS, the proportion of class 1 is the lowest distribution achievable (tied with LS). The distribution for LS mimics MS closely with a peak at class 3 and with

some slight increase from MS in classes 5-7. However due to the many fewer foci in LS than MS (with MS already fewer in number than ES) the width of the variance in the population becomes too large to tell much apart between LS and MS.

Ultimately, the rearrangements caused by the replication machinery at any stage of S phase should be returned to the status quo of chromatin compaction. This assumption is shown valid by prolonging the time between the incorporation of the EdU and fixation Appendix 9.1.

5.2 Rearrangement of the 3D epigenome

General effects of replication on local 3D epigenome

When considering the analysed distribution changes for all markers, regardless of their individual epigenetic function, it is possible to extract a general order for the perturbation caused to the 3D epigenome through replication progression. Grouping markers by general function still provides insight into how each function is affected by this general and ordered change (Figure 5.3).

During ES the majority of the disruption and therefore change observed occurs for the lower chromatin density classes (1-4). This has the highest impact on the markers and functions which in G_1 had their enrichment peaks within these classes, generally, the markers for active transcription and the proteins involved in genome architecture. There is an observable dampening of enrichment for all markers, including markers with multiple undefined functions and for markers of transcription inactivation or silenced chromatin. Nevertheless unlike active and structural markers, which also show some dampening combined with a left-shifted peak enrichment, the markers of inactivity have for the most part an unperturbed right skewed distribution in ES. This phenomenon is clear when looking at the overall patterns for H4K20me1, H3K27me3 and H3K9me2/3, where between G_1 and ES the peak enrichment classes do not change (Figure 5.3(b)).

One exception to this dichotomy is the large effect ES has on the HP1A marker. The new ES profile for HP1A is strikingly similar to those of most architectural and active markers with the exceptions of both H3K4me2/3 and H3K36me2. These mentioned markers of active promoters do not show an enriched peak at class 1 in ES, where the structural markers and RNAPII-S2P do. The only marker which suffers from neither a very strong dampening of enrichment or a noticeable distribution peak shift during ES is H4K20me1. Perhaps due to its dual purpose as both a marker of transcriptionally open state in the highest transcribed promoters, but also its association with H3K9 in silencing.

When looking at regions with recent replication in MS the pattern of change has a more uniform effect on all markers. In MS it can clearly be seen that regardless the function of the marker, no enrichment is detected beyond class 5 (Figure 5.3(c)). There is still heterogeneity comparing the enrichment levels of all markers in these higher density classes with the majority of the highest depletion being found in all the markers of active transcription, the non-cohesin structural proteins, the mixed-function markers (H4K20me1 and H4K5ac), and HP1A. All other markers are not enriched here, but their depletion is not severe. It is worth noting that although both H4K20me1 and H4K5ac show large depletion in classes 6-7, they (alongside H3K9me2) do not show a peak enrichment in class 1 or even class 2, but have the most right-shifted peak of all markers, in class 3. For the most part, all lower density chromatin is now enriched relative to the higher density for each marker. Most of the markers now exhibit a very prominent enrichment in class 1 and 2. Curiously, the markers with the largest enrichment in these classes are also those mentioned as having the lowest depletion in high chromatin classes during MS. Finally during LS the high enrichment pattern for lower density chromatin classes becomes more pronounced. This is both as a result of a further increase in the classes already enriched in MS, but new enrichment in these classes for other markers, such as H3K36me2, H4K20me1, Macro-H2A,

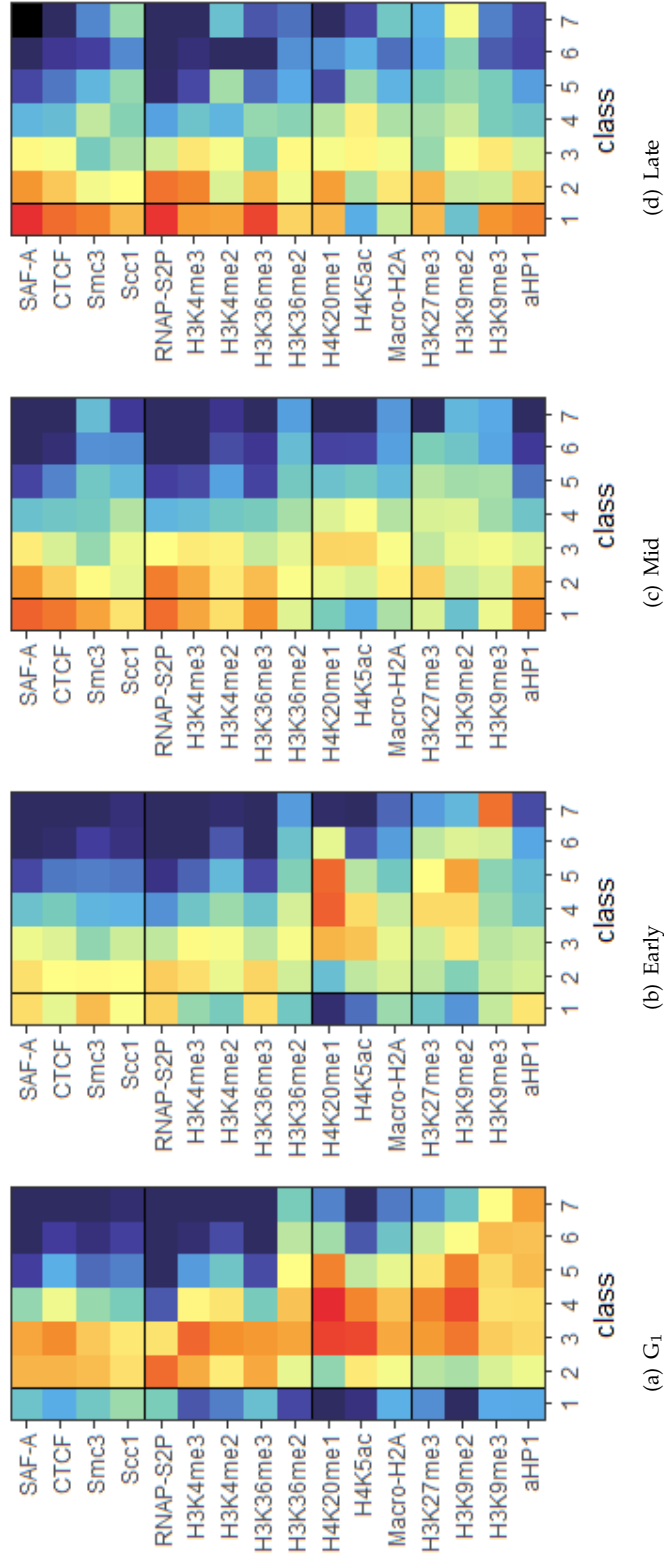


Figure 5.3: **Sphase changes to all density distributions.** Overview heatmaps of enrichment or depletion per chromatin class for all markers comparing G₁ to S phase stages. Markers are grouped by function from top to bottom: Structural, Active, Mid, Inactive. For details of individual markers see Appendix 9.4, Figure 9.6. Profiles with error bars can be found in Figure 9.6. $N_{total} = 800$ cells.

H3K27me3 and H3K9me3 (Figure 5.3(d)).

The described observations can be made sense-of when considering that for many of the regions perturbed in MS, and for nearly all those in LS, are mostly the highest density regions of the genome. The environment around these foci would only have a small fraction of the volume as class 1 or even class 2, hence any detection of markers in these regions would be much higher than what would be expected from chance alone, giving rise to the observed enrichment. LS would be expected to have even fewer voxels in class 1 or 2 than MS, and thus its enrichment in these classes is higher than MS.

Subsequently the general trends outlined here will be examined in relation to selected markers in each of the major functional categories: Activity, mixed activity and bookmarking, repression and chromatin structure.

Active markers

The markers of active transcription all follow a similar pattern of change in response to replication. The pattern of enrichment of each marker shifts left as replication starts. For each marker, the top row of the heatmap shows the mean value of each marker at each chromatin class for G_1 , i.e: the value of the $\log_2$ -fold change bar graphs in §4.2 (Figure 4.7) represented as a colour scale for ease of comparison in a 2D matrix of results.

For RNAPII-S2P and H3K36me3 the ES chromatin condition is enough to shift the new peak enrichment from classes 2-3 to classes 1-2 (Figure 5.4) (Figure 5.5). Although H3K36me2 and H3K4me2/3 still have their enrichment within classes 2-3 they, along with all others beginning to show further depletion in classes 4-7 (see Figure 5.3) By the MS condition all markers have clearly shifted to classes 1-2 (with the exception of H3K4me2 which has a broader signature spanning 1-3). Finally when analysing LS chromatin condition the pattern has truly shifted left for all with the highest peak being class 1.

An intriguing observation is not that the overall pattern shifts left, but that the shift becomes progressively stronger from ES, MS and finally LS. RNAPII-

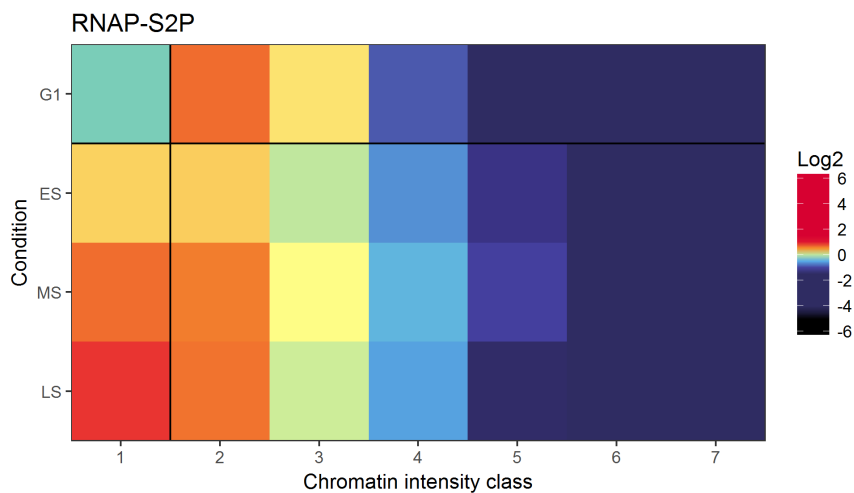


Figure 5.4: **Sphase changes to RNAP-S2P distribution.** Heatmap of enrichment or depletion per chromatin class for RNAP-S2P comparing G₁ to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total}=65$ cells.

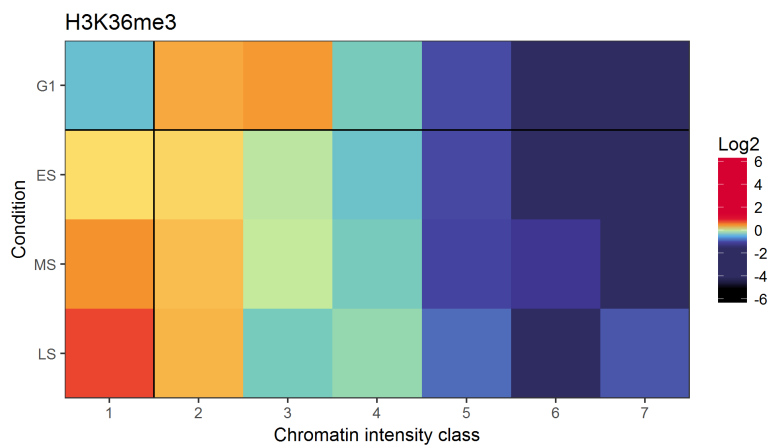


Figure 5.5: **Sphase changes to H3K36me3 distribution.** Heatmap of enrichment or depletion per chromatin class for H3K36me3 comparing G₁ to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total}=130$ cells.

S2P and H3K36me3, are the markers of transcriptional elongation. H3K36me3 being deposited towards the 3' end of actively transcribed genes, i.e: those fully transcribed. The faster shift of these markers to lower density followed by markers of promoters is suggestive of a disruption primarily of the lowest density chromatin (Figure 5.5). This is because regions of active transcription elongation would be the most accessible followed by active promoters some which may be transcribed but others marked and poised to be. Such explanation also agrees with the known general pattern of 3D replication discussed earlier, so the shift is expected.

The reasons for the increasing strength of the shift is in part due to the proportion of chromatin being sampled by the mask around replication foci. The total voxel count for each class type in all these masked regions is used to normalise the absolute count of marker occurrence to their bin volume, producing an enrichment/depletion value from what would be expected from chance. As replication moves to MS and finally LS the proportion of higher chromatin density increases and lower density chromatin decreases (Figure 5.2). This is due to the disruptive local effects of the DNA polymerase when replicating. However on top of a decreased volume of chromatin classes 1-3, some of the chromatin and thus the markers that used to be at class 2, become class 1 as it decondenses, class 3 becomes class 2, and class 4 becomes class 3. However from G_1 it is clear that none of the markers are enriched beyond class 4, so any further decondensation of denser classes (eg class 6 to class 5) cannot lead to the replenishment of the markers that have moved from class 5 to class 4, given that class 6 has fewer markers than class 5 to begin with.

Inactive markers

Markers for transcriptional inactivity show certain similarities in their change with respect to S phase as shown by the markers for activity, with some subtle differences discussed hereafter. Of all the markers, H3K27 has its G_1 distribution most closely resembling that of active markers, the reason for which

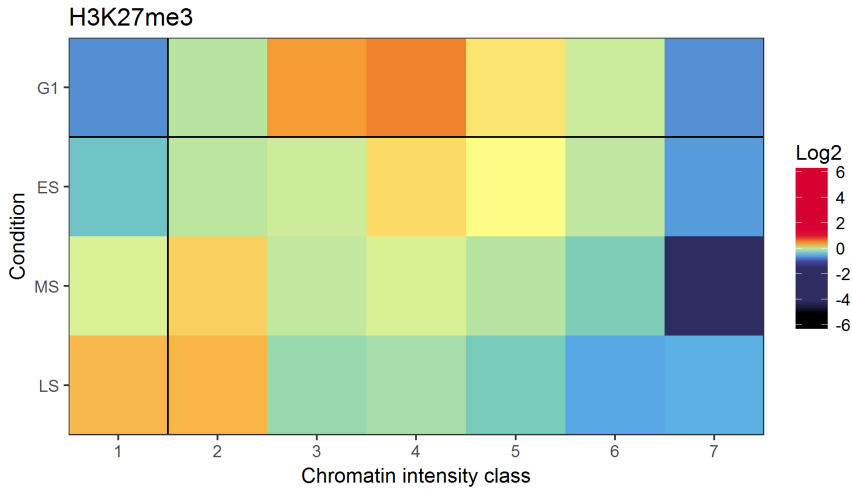


Figure 5.6: **Sphase changes to H3K27me3 distribution.** Heatmap of enrichment or depletion per chromatin class for H3K27me3 comparing G₁ to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total}=65$ cells.

are discussed briefly in §4.2 and more extensively in §8.2. Unlike active markers whose distributions are quickly shifted in ES, H3K27me3 shows almost no distinguishable shift in its class of peak enrichment from G₁ to ES (Figure 5.6). The only difference is a general flattening of the distribution, with regions previously low in marker getting enriched (e.g: class 1 and class 7) and regions originally enriched becoming slightly more depleted (classes 3-5). As S phase progresses the distribution of H3K27me3 eventually follows that of active markers. In MS it becomes more depleted than ES in classes 3-7 and enriched in classes 1-2 (with its peak shifting from class 4 to 2). In LS the peak in class 2 becomes stronger and is approximately equal to the enrichment in class 1. Classes 3-7 all become further reduced in marker but only down close to a log fold change of 0, or no change from chance alone.

The pattern of change for HP1A shows a very clear shift straight from ES. During ES the previous right skewed distribution with peak at class 7 becomes totally inversed, peaking at class 1 and showing negative correlation between marker enrichment and chromatin density (Figure 5.7). As with the other markers, this shift only becomes stronger and narrower in distribution

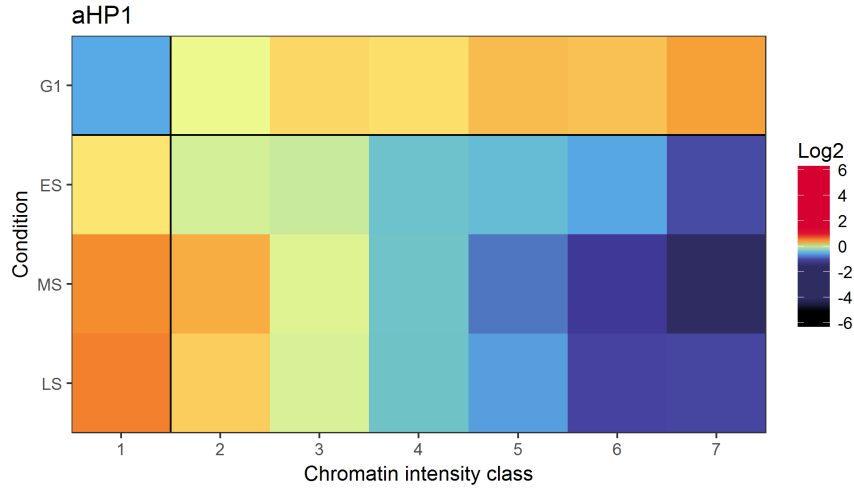


Figure 5.7: **Sphase changes to HP1A distribution.** Heatmap of enrichment or depletion per chromatin class for HP1A comparing G₁ to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total} = 65$ cells.

between classes as replication progresses. Interestingly, HP1A suffers a severe strong depletion in classes 5-7 from MS and LS, where for other markers the depletion here would be far more subtle (if any).

The clear strengthening of peak enrichment at lower density chromatin from ES to MS and LS are indicative of 2 main events. First, at ES where mostly lower density chromatin is perturbed, there are fewer of the inactivity markers to begin with. Thus when this region is perturbed it is only classes 1-3 that see any major changes. The distribution of higher density chromatin remains mostly untouched. After ES, when denser chromatin begins being segmented is when the true environments holding these markers becomes remodelled and so invariably all markers show a shift in peak distributions to lesser density chromatin. The same is true for LS.

The second point of interest is class 1. When considering that this class is mostly undetectable chromatin, having an enrichment here means that during MS and LS dense chromatin becomes so remodelled that markers which occupied inactive regions are pulled away from chromatin and can be spotted in the IC. It is worth noting that the same is not true for ES, both because it

perturbs chromatin of low density already but also because the sites of highest enrichment for some of the inactive markers also concur with those of replication at these particular S phase stages (nuclear envelope, nucleolar periphery and chromocenters).

Transcriptionally indistinct markers

The marker H4K20me1 does not show a left shift of peak enrichment from G₁ to ES but rather a subtle right shift to chromatin of denser nature (Figure 5.8). The peak enrichment is still class 4, however the enrichment profiles of classes 3 and 5 seem swapped over, class 2 is now depleted compared to G₁, but class 6 has increased enrichment. What is more, unlike most other markers it is also not as greatly dampened by ES in the levels of over or under representation of all classes. During MS and then LS the unique combination of right shift and decreased dampening is lost. The marker now shows a loss of enrichment and depletion, flattening out during MS, but establishing a new enrichment at class 2 in LS. To coincide with this the peak enriched class migrates from class 4 in ES to class 3 in MS (subtle as described) and finally to class 2, with a prominent enrichment at class 1 from MS to LS. Class 1 being under-represented heavily in both G₁ and ES, and more subtly but still depleted in MS.

Structural markers

Epigenetic markers that represent genomic architecture exhibit two distinguishable trends in their profile change with replication, though a general trend is also true for all. In the first instance CTCF and SAF-A behave similarly, both to each other and also to the distribution described for the stronger active markers, such as RNAPII-S2P. From the onset of ES there is already a visible and marked left shift of the peak enriched classes from 2-3 to classes 1-2 for both (Figure 5.9 and Figure 5.10). There is, as with most markers also an element of reduced enrichment for all previously enriched classes, which is

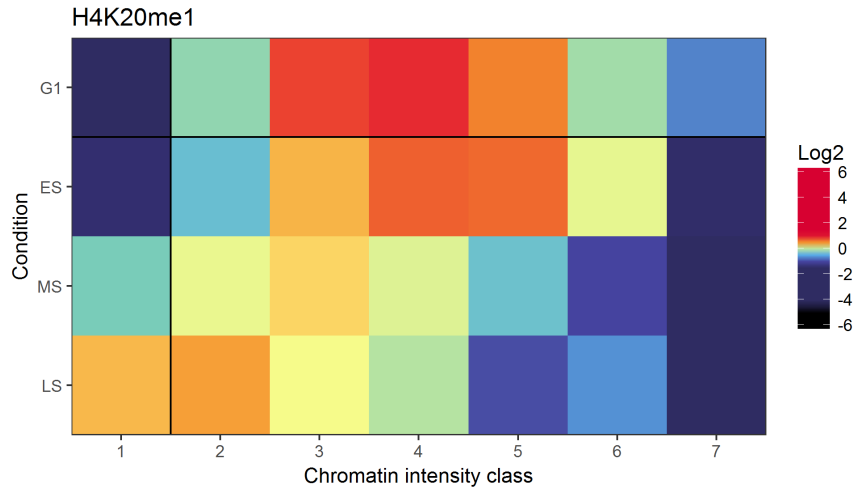


Figure 5.8: **Sphase changes to H4K20me1 distribution.** Heatmap of enrichment or depletion per chromatin class for H4K20me1 comparing G_1 to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total} = 65$ cells.

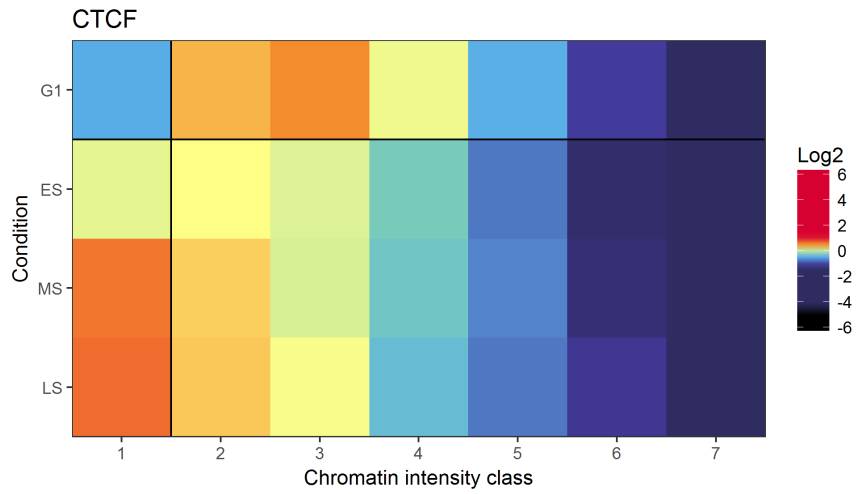


Figure 5.9: **Sphase changes to CTCF distribution.** Heatmap of enrichment or depletion per chromatin class for CTCF comparing G_1 to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total} = 65$ cells.

most noticeable for CTCF classes 2-3 in MS. However, unlike RNAPII-S2P and others, there is no increased enrichment in previously depleted classes. SAF-A is the most clear example of this with no change in the values for classes 5-7 from G_1 to ES and even to MS.

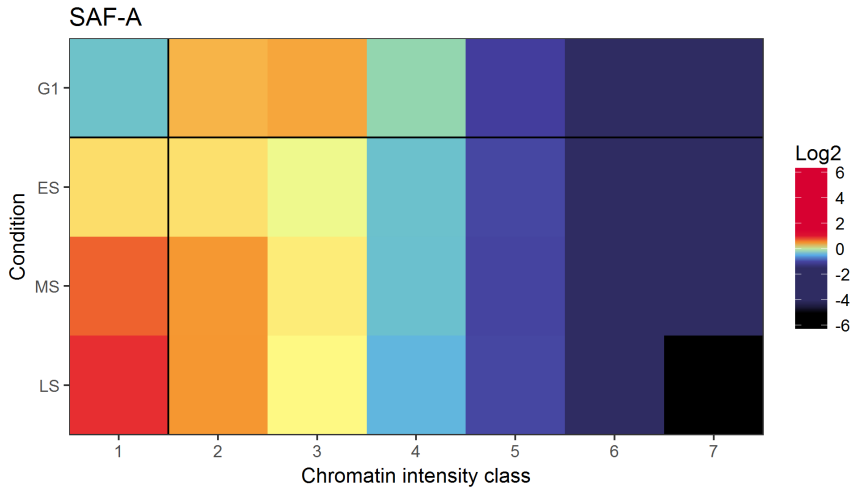


Figure 5.10: **Sphase changes to SAF-A distribution.** Heatmap of enrichment or depletion per chromatin class for SAF-A comparing G₁ to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total} = 65$ cells.

During the transition to MS and LS both CTCF and SAF-A have a clear shift of their peak enrichment class from 2 to 1. This enrichment at the IC only becomes more pronounced comparing MS to LS. There is a subtle difference between SAF-A and CTCF here, with the former having a much greater increase in class 1, reminiscent of H3K36me3 (Figure 5.5) and RNAPII-S2P (Figure 5.4). CTCF does not reach this extreme, and has also a generally less dramatic depletion in classes 5-7. SAF-A on the other hand not only keeps a strong depleted profile in classes 5-7 but shows an even deeper level of depletion in class 7 during LS, stronger still than the strongest marker for transcriptional activation, RNAPII-S2P (Figure 5.4).

Not surprisingly both markers of the same complex, the cohesin subunits, Smc3 and Scc1 behave most similarly to each other than to the other two markers for genome structure. Scc1 and Smc3 are also characterised by a visible left shift of peak enriched class, as with SAF-A and CTCF, however the former two have a stronger shift from G₁ to ES, making class 1 the peak enriched in ES (as with HP1A, Figure 5.11(a) and Figure 5.11(b)). The left shift is further established and becomes more pronounced with increased enrichments in the peak

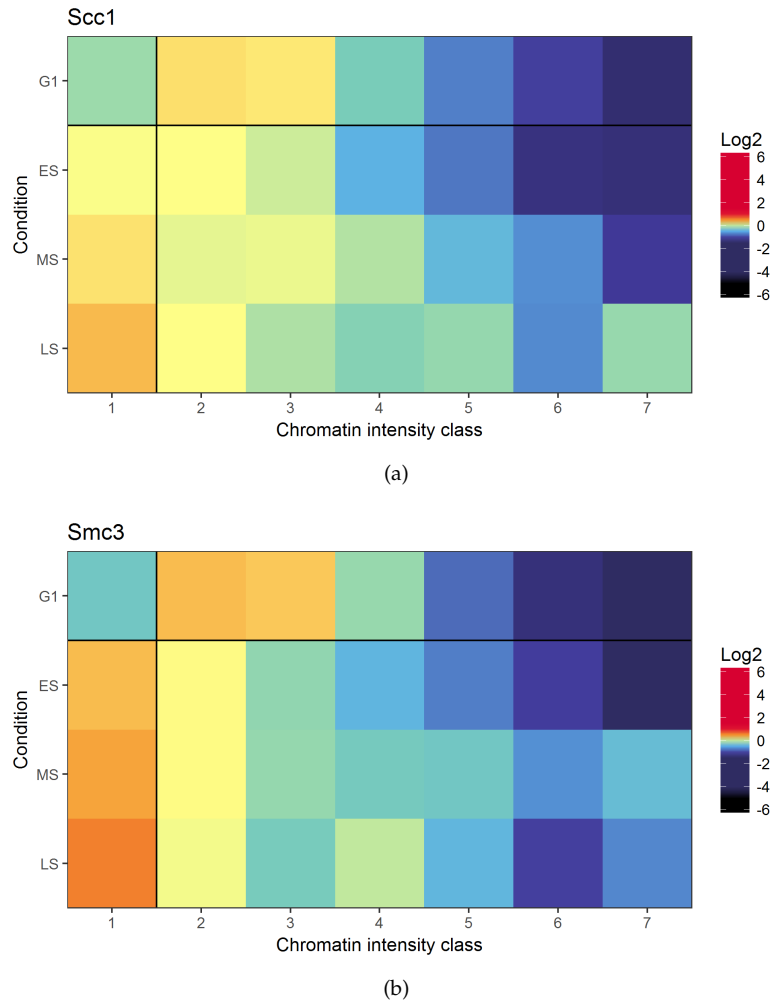


Figure 5.11: **Sphase changes to cohesin distribution.** Heatmap of enrichment or depletion per chromatin class for Scc1 (a) and Smc3 (b) comparing G_1 to S phase stages. Profiles with error bars can be found in Figure 9.6. $N_{total}=130$ cells.

class 1 from ES to MS and finally to LS, this is most strongly exemplified by Smc3. The key difference between the cohesin subunits and the other structural markers is the profile change in the higher density chromatin classes 4-7. Here both Scc1 and Smc3 become progressively less depleted in MS compared to ES and in LS compared to LS. The overall effect being a strong shift in the peak with a long tail tapering off to no great deviation in enrichment from chance from classes 3-7.

Summary of the S phase epigenome

Taking the previous analysis of marker distribution, it is clear that despite heavy depletion all markers can be detected in all classes. According to the model here, those few markers of active chromatin in higher classes would encounter the replication machinery, and after the initial fork has passed they may now be in more accessible and favourable density to interact with their corresponding machinery. Such interactions may prevent their subsequent return to the denser class. On the contrary, markers for inactive chromatin are usually found in much denser chromatin (Figure 5.3(a)). Perturbation of this chromatin does indeed lead to a decrease in enrichment. However, unlike markers of activity, markers of inactivity do not have preferential interaction with transcriptional machinery. As such they are not as highly enriched in class 1 and 2 (where the transcription machinery is, see RNAPII-S2P). However, because they have been moved from the higher density classes, inactivity markers must be in the lower density regions. These two contradicting mechanisms lead to the observed effect, of a broad enrichment profile enriched between classes 1-3, but extending far into class 5 for many. Furthermore the enrichment in class 1 and 2 is much lower than the relative enrichment observed for all the active markers.

From this interpretation it should be expected that running ChaiN again on cells fixed at longer chase times after S phase would result in the observed heatmaps for ES, MS and LS return to a pattern closely resembling the one obtained in G_1 (Figure 5.3(a)). Each individual heatmap for each of the stages will not look identical to Figure 5.3(a) as the masks used to select the coordinates for each have a bias for chromatin density. However, averaging all three heatmaps from fixation after all local rearrangements from replication have passed should yield almost the same pattern as seen in Figure 5.3(a). This might not be the case only if unknown mechanisms during G_2 are at play which may alter the 3D epigenome in comparison to the distribution analysed in G_1 , perhaps in preparation for cellular division.

Mechanisms of global chromatin perturbation

Mechanisms affecting chromatin topology

As discussed in Chapter 5, the genome can have its 3D topology remodelled by the action of cellular processes such as replication. This particular remodelling due to the progression of the denatured replication fork and DNA polymerase occurs at a local level, remodelling the region immediately around the replication foci. Nevertheless, it is also known that many other endogenous mechanism can also be conducive to an eventual rearrangement of chromatin beyond a single genomic locus. A key example that is well known and continues to be an active field of investigation is the formation of the Barr body. The action and spread of the Xist RNA across the X chromosome leads to the general compaction of chromatin in the silenced inactive X (Xi). Given that the proposed “RNA environment” of the nucleus is the IC (class 1) it is not surprising that previous findings on the 3D coverage of the Xi by Xist occurs from the regions of less dense chromatin (Smeets et al., 2014). Like a surfactant covering the Xi from the outside, not as an anchoring and internal mechanism.

Other cellular states also result in a dramatic change in the 3D topology of the genome. For this investigation, the two more pertinent conditions are: chromatin in embryonic stem cells (ESC) and chromatin in senescent (G_0) cells. According to the accessibility hypothesis, somatic adult cells have a mechanism by which they may hide genomic regions from contacting nuclear machineries. This simply is achieved by the occlusion caused from a dense packing of chromatin around these regions. This environment can be thought of as a storage compartment.

In senescent cells which are no longer replicating, the chromatin topology is far more compact (Figure 6.11(c)), potentially leading to a larger volume of the genome to be hidden away in storage. This can be an advantage when cells have been driven into senescence by erroneous or harmful events. For example, G_0 driven compaction can be observed in cells along with poor segregation of chromosomes during mitosis. They can also be screened for by the lack of expression of the Ki67 marker, as applied in this study. G_0 can be due to replicative induced senescence (RIS) or oncogene induced senescence (OIS). In the former it is caused by shortening of telomeres by erroneous telomerase repair; in the latter, as the name suggests, by the activation of a proto-oncogene to a transcriptionally active oncogene that can lead to cancer (Nelson et al., 2014; Cavalli et al., 2014). By halting replication and exaggerating the compaction of the chromatin network these cells exclude more of the genome from interacting with the transcription machinery, decreasing the danger posed from breaks or oncogenes.

Conversely one would expect a looser chromatin environment in cells which do not require, or in fact would be hindered by seclusion of any part of the genome. Such cells should have very few regions of high density storage compartments. This is the case indeed with the morphology observed for the chromatin network of ESC. By allowing a faster rate of replication, the constant perturbation of the replication machinery keeps the genome in a more decondensed state than adult somatic cells. For ESC the advantage is not only

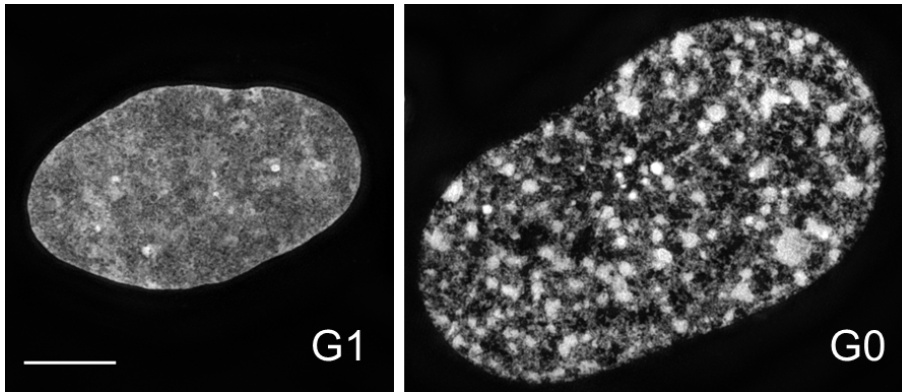


Figure 6.1: **G₀ chromatin.** Max projection of 3D-SIM z stack of G₁ and G₀ nuclei. The fine structure of chromatin networks observed in G₁ collapses to large scale chromatin macro-domains. Scale bar = 5 μ m

the higher rate of division, but also the effect of having the whole genome more easily accessible. These cells have no defined lineage yet, so they need to keep this high general accessibility until a transcriptional programme is established. Once this is established, the hypothesised model predicts that these cells hide the un-needed chromatin in their high density storage, and their pattern reverts to what can be observed in adult somatic cells.

The highest levels of chromatin compaction achievable of course belong to the nucleus of adult haploid germ line in males. During the last haploid phase of spermatogenesis almost all nucleosomes (85%-90%) are replaced by protamines. DNA winds around these in toroids of approximately 50kb each (Balhorn, 2007; Brewer et al., 1999). This extreme packing confers several advantages to sperm, predominantly the decreased volume required to store their genetic information, and so a more hydrodynamic shape for displacement towards oocytes (Woop et al., 2015). Interestingly for the accessibility model, the presence of protamines also decreases the incidence of DNA damage. This may be, as with G₀, cells that the majority of the genome is now not in reach of molecular factors which may compromise its structural integrity.

The underlying mechanisms of action that compose normal physiological chromatin topology may be elucidated by using super resolution imaging in

conjunction with ChaiN analysis on images acquired during chromatin rearrangements of both ESC and G_0 cells and on adult somatic cells treated to mimic the global perturbation in ESC and G_0 . The results obtained from these experiments are outlined below:

6.1 Global chromatin decompaction

A general state of chromatin decompaction, as observed in ESC can also be achieved in somatic cells by the introduction of exogenous drugs. One such molecule with this effect is trichostatin-A (TSA), an organic antifungal compound. Its mode of function in mammalian cells is the inhibition of class I and II HDACs, histone deacetylases. Acetyl groups which coat the N terminus of nucleosomes change their charge to neutral/negative, conferring a more loose interaction between the nucleosome and the DNA packaged around it (Vanhaecke et al., 2004). The change observed from this drug is a permanent decondensation of the chromatin network. A more transient and reversible, yet highly effective method for achieving this ESC morphology in adult cells is by changing the osmolar state of the cell, and thus nucleus (Irianto et al., 2013). Reduction in the osmolarity of the medium of incubation (hypo-osmolarity) causes the desired effect.

Both TSA and hypo-osmotic treatments were tried in this investigation after their optimum effect concentrations were qualitatively defined empirically. The experiments with these conditions resulted in the expected disruption of the wild type chromatin topology (Figure 6.2). Furthermore, antibody detection for immunofluorescence (the basis of the raw data required by the ChaiN workflow) was not adversely affected by the addition of these variables.

Nevertheless, the very nature of this disruption hindered its own subsequent analysis. As described in 1.3, super-resolution by SIM relies on the reconstruction of a structure of origin by working backwards from an observed structure and the known structured illumination pattern (here stripes). Since

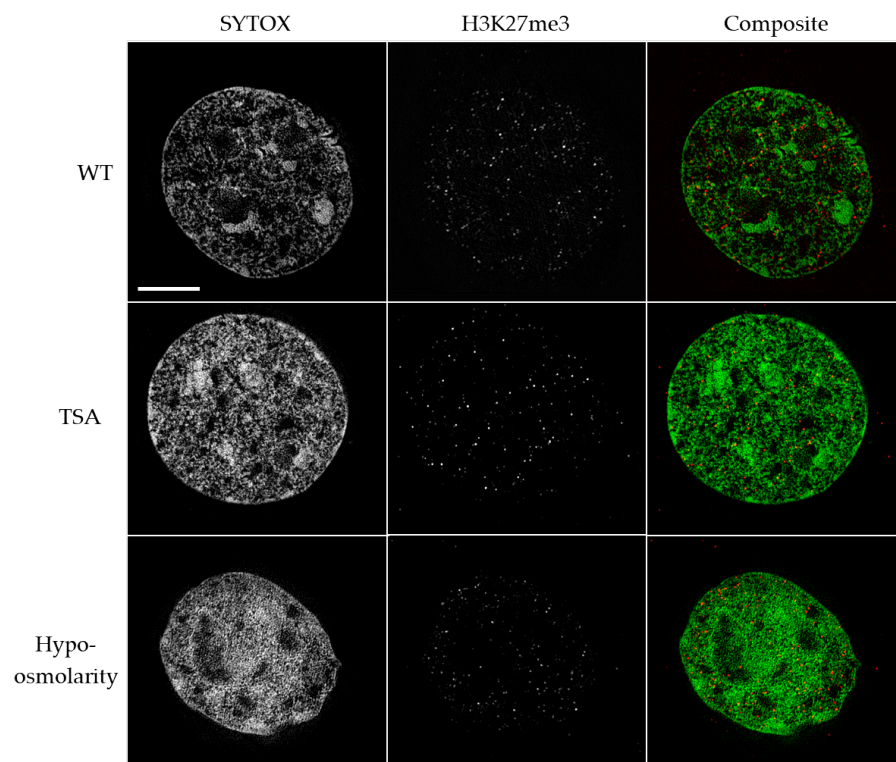


Figure 6.2: **Chromatin decompaction.** 3D-SIM images comparing exogenous methods of chromatin decompaction compatible with SYTOX and IF staining. Increased chromatin decompaction can be achieved by treating cells with TSA or hypo-osmotic conditions as described in 2.2. Scale bar = 5 μm .

in the case of TSA and hypo-osmotic treated cells the chromatin structure is made diffuse to a state of almost full homogeneity (Figure 6.2), there is very little initial structure in order to generate the raw striped acquisition for reconstruction. This means that for target cells with a general diffuse chromatin, very poor SIM reconstruction of this chromatin could be obtained. This is clearly observed using the Image J plugin: SIMcheck, from its modulation contrast function (Figure 6.3). Given that the ChaiN workflow incorporates an MCNR filter as a clean-up before processing, these stains would not yield workable data. For ESC, this inherent issue of poor structure for SIM reconstruction coupled to a very uneven and non-spherical nuclear envelope, is one of the key obstacles in obtaining useful SIM data for ESC analysis.

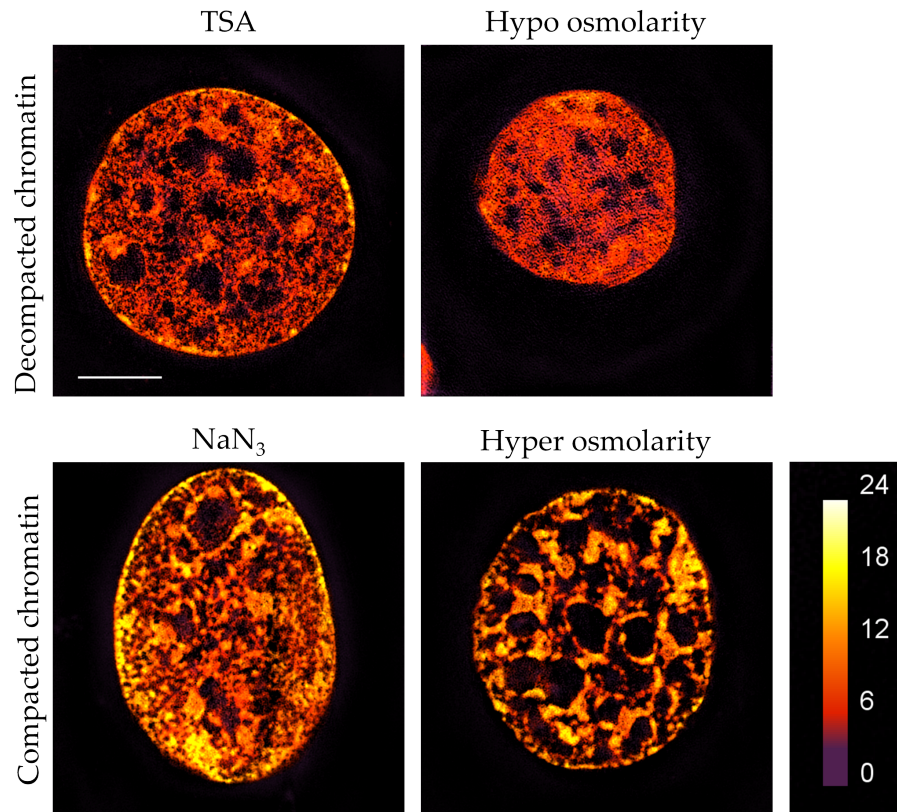


Figure 6.3: **Modulation contrast of different chromatin topologies.** Single z plane of 3D-SIM reconstructions of chromatin after different treatments. Chromatin decompaction lowers the overall fidelity of reconstruction whilst chromatin compaction increases it. Scale from SIMcheck (right). The inherent problem is that decompaction removes the structure required for sufficient contrast in 3D-SIM, this mimics the difficulty in imaging low-compaction embryonic stem cells. Scale bar = 5 μm .

6.2 Global chromatin compaction

To achieve a general state of chromatin compaction in adult somatic cells, similar to what can be observed during the onset of senescence, it is possible to use the inorganic molecule sodium azide (NaN_3). This compound forms a colourless salt and was widely used in older airplane parachutes and car airbags for its useful high-speed detonation from an electrical discharge, releasing copious nitrogen gas. It is however being phased out due to safety concerns over its other properties. NaN_3 is soluble in water and makes an ex-

tremely potent poison (Chang and Lamm, 2003). With a lethal dose of 700 mg for a human adult it has a toxicity on par with many cyanide compounds. The azide molecule irreversibly binds to the heme factor of cytochrome oxidase, a key protein in the respiratory electron transport chain, inhibiting its enzymatic activity. The result of NaN_3 poisoning is a depletion of ATP in cells. For organs such as the heart, liver and brain, with an elevated rate of respiration, such depletion is enough to cause organ failure, which is the primary cause of death by NaN_3 poisoning.

The very mechanisms of its lethality makes NaN_3 a useful tool for researchers, primarily those concerned with metabolism. Beyond the mitochondria however, ATP is still required for a variety of nuclear activities: nuclear transport, DNA replication, protein phosphorylation, RNA transcription, and the synthesis of nucleobases, nucleoside triphosphates, amino acids and loaded tRNAs. The perturbation of such a combination of elemental nuclear functions is unsurprisingly characterised by a major remodelling of chromatin architecture (Figure 6.4). Furthermore eukaryotes have a selection of ATP dependent chromatin remodellers sharing a homologous ATPase domain (Wang et al., 2007). The SWI/SNF complexes disorder nucleosomes while the ISWI complex orders them with even spaces between them and has also been implicated in higher order chromatin architecture (Lia et al., 2008). The destabilising effect of the SWI/SNF complexes in chromatin packaging is proposed to prepare chromatin for binding by transcription factors (Cristofaro et al., 2001). Nevertheless, despite the many key processes affected by sequestering ATP many known modifications to chromatin are independent of ATP hydrolysis. In this list are most of the histone markers analysed in this study, i.e: enzymatic methylation and acetylation of histones. The advantage of using NaN_3 therefore is the global rearrangement of chromatin without affecting the markers used in this investigation to locate the distribution of the 3D epigenome (see IF signal in Figure 6.4).

As alluded to previously, a more transient and reversible method for

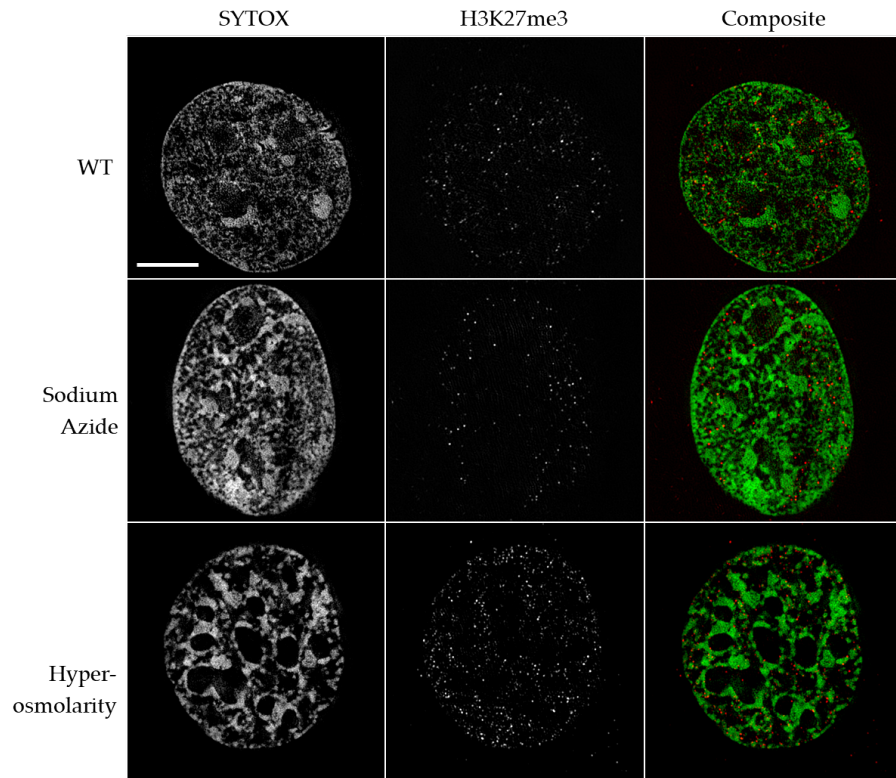


Figure 6.4: **Chromatin compaction.** 3D-SIM images comparing exogenous methods of chromatin compaction compatible with SYTOX and IF staining. Increased chromatin compaction can be achieved by treating cells with NaN_3 or hyper-osmotic conditions as described in 2.2. Scale bar = 5 μm .

achieving a perturbed chromatin topology is possible by changing the osmolar state of the cell. In the case of chromatin compaction, an increase in osmolarity (hyper-osmolarity) is enough to trigger this change with the detection of histone modification epitopes not affected by this incubation (Figure 6.4).

Quantification of network changes

In both cases, whether using the NaN_3 compound or exposure to hyper-osmolarity, the chromatin was clearly perturbed, the magnitude of said perturbation being large enough that the change is detectable by eye from confocal micrographs without requiring sophisticated image analysis. However, using super-resolution microscopy and the ChaiN workflow, the change is

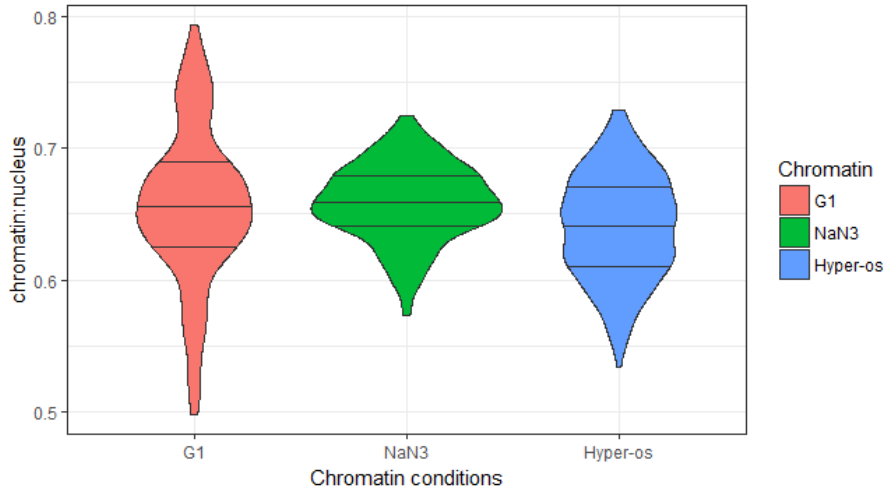


Figure 6.5: **Chromatin to nuclear volume.** Ratio of chromatin volume (binary of density classes 2-7) to whole nuclear volume plotted as a violin plot showing quartiles as horizontal lines of distribution for untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei. Despite large chromatin morphology changes observable by eye, the overall proportion of the nucleus occupied by chromatin does not seem to change. $N_{total} = 960$ cells.

chromatin network morphology can be quantitatively described for the first time.

The first metric which can easily be obtained by ChaiN allows the comparison of the overall ratio of chromatin to whole nuclear volume. This ratio in G_1 cells had a average proportion of approximately 0.65, 65%. There is no detectable change in these values when comparing cells treated with NaN_3 or hyper-osmolarity (Figure 6.5).

When comparing the fraction of the nuclear space taken by non-chromatin, that is the IC or class 1 in the ChaiN segmentation, there is also no significant difference in the average of these distributions. Although it can be noted that variance in the distribution of class 1 volumes for NaN_3 and hyper-osmotic treatments does narrow compared to G_1 , with an inter quartile range (IQR) of approximately 2% and 5% compared to 11% respectively; the proportion of nuclear volume dedicated to each compartment (chromatin and non-chromatin) remains the same.

What does change in these cells, is the proportion of chromatin which is in each of the density classes 2-7 (Figure 6.6). It is clearer when comparing G_1 to hyper-osmolarity that classes 2 and 3 show a progressive and marked decrease in proportion, from an average of approximately 26% to 20% for class 2 and from 18% to 15% for class 3. Although the absolute difference does decrease by almost half, from 6% to 3%, it is worth noting that the spread of the distribution for these classes also decreases, yielding almost no overlap between the IQR of each class.

Comparing the higher density end of the classification yields the opposite pattern. Proportionally to the nuclear volume there is a significantly elevated percentage volume occupied by classes 4-7 in NaN_3 and hyper-osmotic treated cells than in untreated cells in G_1 . Classes 5 and 6 having the highest level difference with no overlap between G_1 and hyper-osmotic distributions. As with all classes, the values for NaN_3 are usually between these two populations. Furthermore, as opposed to class 1 and 2 where the variance of the distributions was narrower for treated cells than untreated, the opposite is true for classes 4-7. In these higher density classes, the variance from hyper-osmolarity is always larger than that for the G_1 distribution.

The dramatic shift in calculated proportions, lower for classes 2-3 and higher for classes 4-7 is the quantified confirmation of the general interpretation from visual inspection. That cells which have been treated with NaN_3 or hyper-osmolar conditions tend to have a pattern of chromatin network with a more dense filament and conversely far less intricate and convoluted. Using ChaiN it is possible to quantify both such observations further.

From the modelled randomly scattered measurements fitted to a Gaussian density curve and modelling the network as tubular filaments as described in 3.3 ChaiN can calculate the average thickness of these filaments (Figure 6.7). The calculations confirm qualitative observations that the width of the average network filament is broader for treated cells. For the first time ChaiN can yield a quantitative result, measuring the average radii at 263 nm, 280 nm and

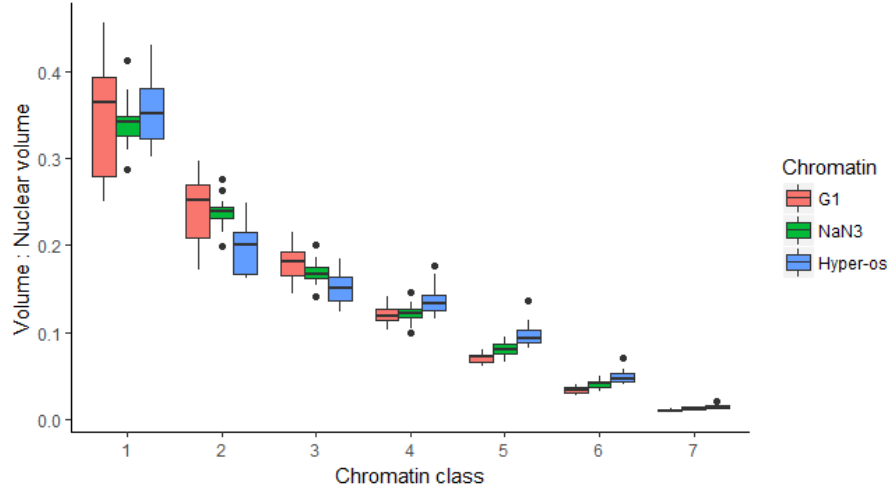


Figure 6.6: **chromatin to nuclear volume.** Distribution of 6.5 sub divided by individual box plots for each chromatin class, comparing untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei. $N_{total} = 960$ cells.

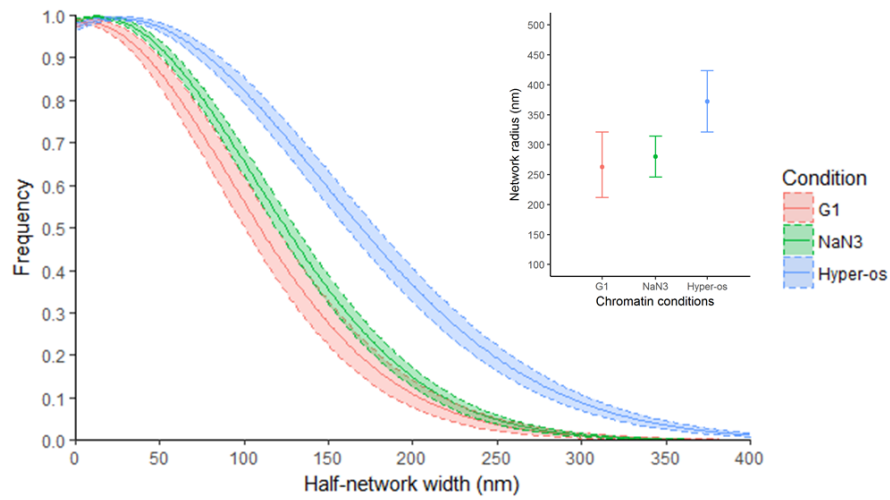


Figure 6.7: **Network filament changes.** Modelled frequencies fitted for each cell at each nm distance to the IC boundary plotted as coloured line with 95% confidence as coloured ribbon for untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei. $N_{total} = 960$ cells.

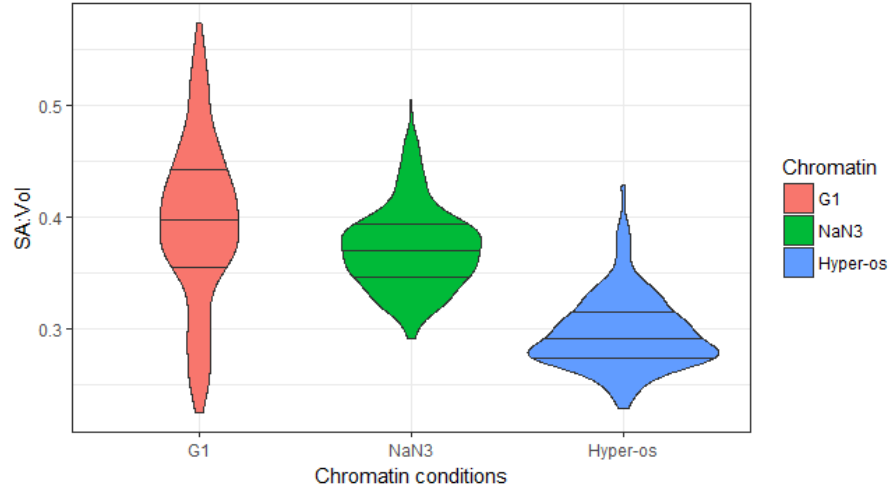


Figure 6.8: **Chromatin surface area to volume.** Surface area to volume ratio calculated from the voxel count of boundary voxels between class 1 and 2 and the chromatin voxels of classes 2-7 comparing untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei. $N_{total} = 960$ cells.

372 nm, giving a filament width of 526 nm, 560 nm and 744 nm for untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei respectively.

From the segmented IC-chromatin boundary, the consequence of increasing filament width is seen as a decrease in surface area, given the volume of chromatin segmented (Figure 6.8). The average percentage decreases from $\sim 40\%$ to $\sim 37\%$ to $\sim 29\%$, in G_1 untreated to NaN_3 to hyper-osmotic treated cells respectively. The distribution variance decreases for both treatments compared to the untreated G_1 , and while there is a high degree of overlap between NaN_3 and G_1 there is no overlap between chromatin affected by hyper-osmolar conditions and chromatin in G_1 .

Global epigenetic perturbation

Given the ability to affect chromatin without affecting the deposition and detection of the epigenetic markers analysed in the cultured G_1 cells, the corresponding immunofluorescence experiments were repeated with all markers for each of the treated samples. An equal number of 3D SIM micrographs were

acquired and passed through ChaiN under the same parameters.

Density profiles

From the hypothesised accessibility model, one predicted effect of a global chromatin compaction by ATP depletion or hyper-osmolarity would be distinct rearrangements of the 3D epigenome. This assumes, as the model suggests, that the organisation of the epigenome in 3D space is influenced by the organisation and accessibility of chromatin. It is important to note that while chromatin rearrangements may correlate with a re-distribution of the epigenome, the epigenome may also affect local chromatin such that this feedback stabilises the final configuration of both chromatin and the epigenome. Nevertheless, increased volumes of denser chromatin relative to lower density classes coupled to an larger distance from chromatin to the nearest chromatin-IC interphase surface (on average) should yield a higher level of marker depletion in the denser classes for all markers, while the enrichment at lower density classes remains unchanged.

Comparing this prediction to the observed results shows that the data obtained here backs this model (Figure 6.9). Comparing G_1 to NaN_3 then hyper-osmolarity (i.e: moving from the lower to higher levels of compaction along with decreased surface area) shows a general pattern change in the distribution of the epigenome with chromatin density. For all markers of active transcription as well as the structural markers there is an depletion (of a magnitude similar to that observed in class 7 for G_1) encroaching down to class 4. The same is also true for two of the three markers involved in multiple epigenetic functions, H4K5ac and Macro-H2A. Surprisingly though, despite hyper-osmolarity causing a significantly lower surface area to volume ratio compared to NaN_3 treatment (Figure 6.8), the strongest levels of marker depletion are observed for the active markers affected through ATP depletion by NaN_3 . Hyper-osmolarity does still have a higher degree of depletion overall in class 4 than NaN_3 for the active and structural markers.

Interestingly, the marker H4K20me1 shows almost no response in its overall enrichment/depletion profile with respect to density when the whole genome is redistributed in 3D. The only subtle change is the peak enrichment value transferring to class 3 from being previously at class 4. Nevertheless The value at class 4 remains almost unchanged. There is a small drop in enrichment in class 5, 6 and 7. A similar observation can be made for half of the inactive markers, H3K27me3 and H3K9me2. Increasing compact chromatin and decreasing surface area very subtly decreases their enrichment in classes 5-7 but does not shift the enrichment pattern for the lower density classes.

This is in stark contrast to the changes that are forced on the active and structural markers, where not only is there a very large depletion in the high density classes, but also the peak enrichment class gets progressively shifted from an average of class 2-3 in G_1 , to unequivocally centred around class 2. It is in fact very important to highlight the major contributor to class 2 being the average for both types of markers just described. The structural markers are in effect progressively pushed further and further to lower density chromatin (average at class 2-3 in G_1 , 2 in NaN_3 , and 1 in hyper-osmotic conditions). Yet, the markers of active chromatin, are depleted in class 4 and progressively enriched in class 2 and also in class 3. They are still on average still equally depleted in class 1.

This division makes sense in light of this model. With a reduction in the available surface area to act upon, nuclear machinery concentrate the deposition of active markers on the fewer remaining volumes of lower density chromatin which they are able to access. This leads to an enrichment in the normalised distributions for the active markers in class 2 and 3. However all these histone modifications are still by definition part of chromatin and so cannot be enriched in class 1 which also by definition and segmentation has non-detectable levels of chromatin. The only active histone marker to show a similar pattern to RNAPII-S2P is H3K36me3. This is mechanistically suggestive of loops of chromatin being extruded from the surface outward as tran-

scription progresses, with the 5' end of genes (marked by H3K36me3) being at the zenith of the extrusion, while promoters (even active, see H3K4me2/3) remain closer to the chromatin network. Non-chromatin proteins, such as the structural markers analysed here, do not follow the pattern of the active promoter markers, even though they are initially highly correlated to the low density chromatin classes 2-3 as well. When the volumes of these classes becomes smaller an increased proportion of these molecules will be found in the IC, similar to the effect shown by the RNAPII-S2P protein. However, given that these proteins are not intrinsic to transcription, their enrichment does not require underlying chromatin. Therefore the peak enrichment for the structural markers shifts all the way to class 1.

The final pattern of interest to discuss is the two opposing extremes to which the H3K9me3 and HP1A markers shift to. Both of these markers are functionally cooperative, with H3K9me3 binding preferentially to HP1A (Lehnertz et al., 2003). This is clearly visible being the only two markers with the same, fully-right skewed distribution in enrichment to the densest chromatin classes (Figure 4.8). Despite these clear supporting facts, both of these markers react very differently in their distribution with respect to compaction. The strongest class 7 enriched of the two, HP1A, shows a progressive depletion in classes 4-7, with the only mild enrichment detected in classes 2-3. Conversely H3K9me3, shows increased enrichment in almost all classes (3-7), with a prominent jump in enrichment from classes 4-7. This enrichment persists despite depletion of HP1A from the same densities. It is worth pointing that this enrichment is also coupled to a stronger depletion in class 1 and 2 and that the peak of the enrichment remains almost unchanged, being symmetrically centred around class 5.

It is important to note that despite making the assumption that both NaN_3 and hyper-osmotic disruption would cause a similar effect despite their different mechanisms of action, this assumption can be shown to be correct by the pattern of change for all markers. The downstream effects on the

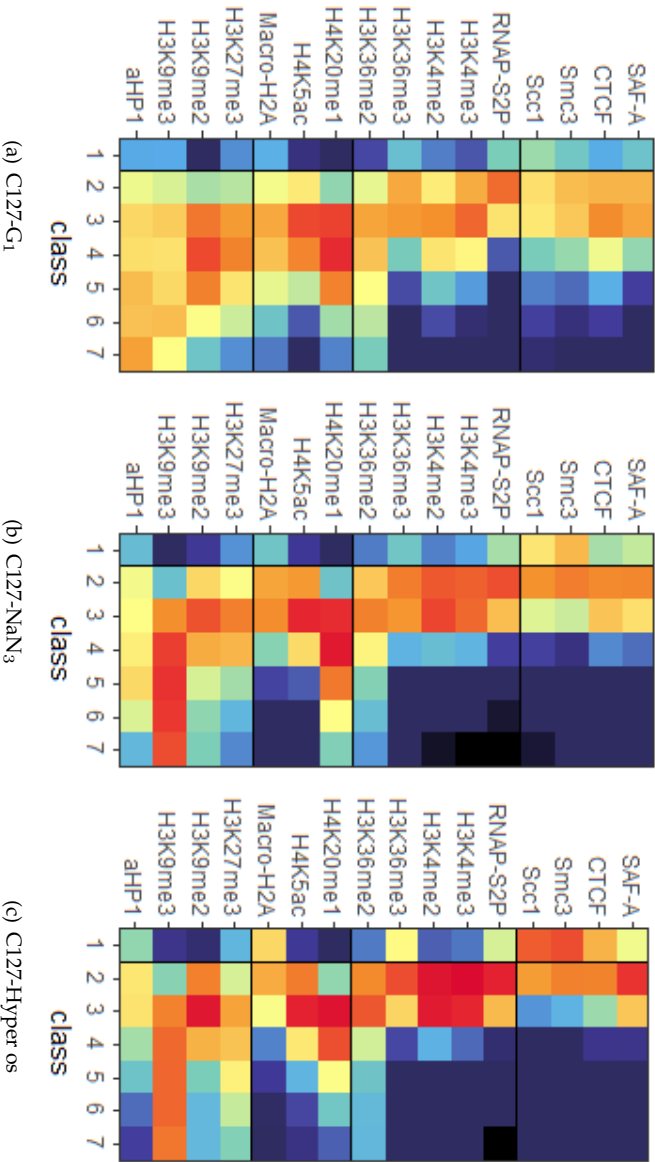


Figure 6.9: **Chromatin compaction changes to all density distributions.** Overview heatmaps of enrichment or depletion per chromatin class for all markers comparing untreated G₁ and NaN₃ or Hyper-osmotic treated C127 nuclei. Markers are grouped by function from top to bottom: Structural, Active, Mid, Inactive. For details of individual markers see Appendix 9.4, Figure 9.7. Profiles with error bars can be found in Figure 9.7. $N_{total} = 960$ cells.

epigenome can't be too dissimilar between NaN_3 and hyper-osmolarity as all markers show similar progression in a single direction. Whatever change is shown in NaN_3 (with mid-range compaction) changes in the same direction and to higher degree in hyper-osmotic conditions (which causes a higher degree of compaction). If ATP depletion only caused a particular effect on a section of the epigenome then these markers would act differently upon NaN_3 to G_1 but then either return back to the same G_1 profile for hyper-osmotic treatment, showing a regressive profile. But that is not the observed result.

Distance measurements

Using the analysis of distance ChaiN provides, between the position of the segmented marker and the boundary between the IC and chromatin, it is also possible to back up the previously alluded-to changes. Here there is no assumption between the density of chromatin and its actual three-dimensional relation to its surface area. Yet, both the distribution based on density as that based on distance depict the same broad changes based on marker functionality. Of course, as with the first analysis of G_1 distances (§4.2), the relatively broad error bars from the calculated CI95 of each distribution are unavoidable. At the level of SIM resolution, the fact that the CI95 can be brought so low as to within half the maximum diagonal of a voxel is a testament to the power of high-throughput analysis.

In spite of this limitation, it is possible to use *a priori* knowledge to show certain trends. This can be done both with the knowledge of grouping markers by their epigenetic function to get an overview of how this function is affected, and also by knowing the progressive magnitude change in chromatin compaction from NaN_3 and hyper-osmolarity. Thus, despite large overlaps in CI95, a shift in one direction from G_1 to NaN_3 followed by another progressive shift in the same direction from NaN_3 to hyper-osmotic condition is a clear indicator of an underlying mechanistic truth. More so when this pattern can be replicated and confirmed over multiple markers, as the probability of

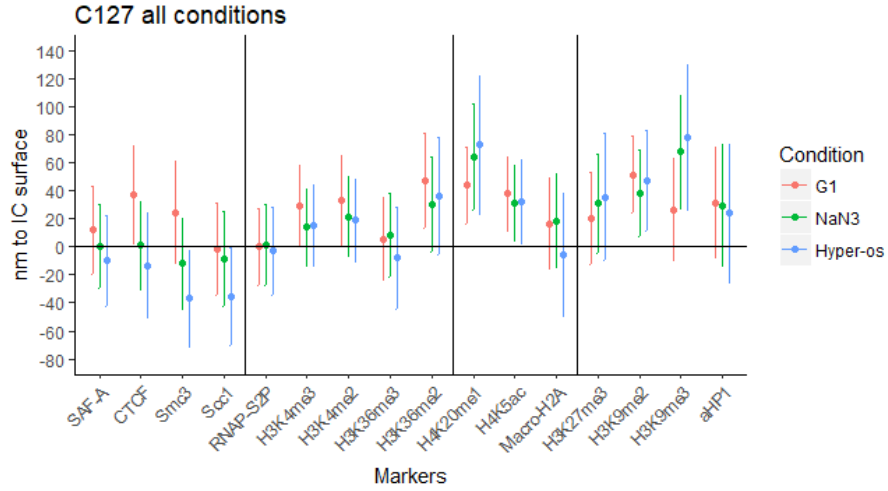


Figure 6.10: **Compaction effect on distance to IC.** Calculated mean sub-voxel distance of all IF markers to the IC comparing untreated G_1 and NaN_3 or Hyper-osmotic treated C127 nuclei. Positive distance is equivalent to marks within the chromatin network. Markers are grouped by function from left to right: Structural, Active, Mid, Inactive. Bars show 95% confidence. $N_{\text{total}} = 960$ cells.

all markers showing progressive movement in only one direction is extremely low.

A clear example of this is the distance change that occurs in the markers of genome architecture; SAF-A, CTCF and the cohesin subunits (Figure 6.10). Here the average distance moves from close to the surface but within chromatin, to close to the surface but found in the IC. There is not a large change in SAF-A, a marker which starts in G_1 close to the surface. However for SMC3, the difference between its G_1 value and its value after hyper-osmotic shock is large enough to have almost no overlap in both of the CI95 for each average. The overall trend for structural markers is evident and supports what is shown by the density distribution. Increased chromatin compaction causes these epigenetic functions to take place on average further outside of the chromatin network.

Of the markers of transcriptional activity, RNAPII-S2P shows almost no difference in distance. However markers for active promoters, H3K4me2/3

and H3K36me2 follow a very similar shift both for NaN_3 and hyper-osmolarity (compared to their G_1 averages), being still within chromatin but closer to the network surface. As was observed for the density class shifts, H3K36me3 is the only chromatin marker whose average position during extreme chromatin condensation by hyper-osmotic conditions is found just outside of the network. This marker is only enriched at the 5' end of transcribed genes and its similar behaviour to the elongating RNAPII-S2P mark is thus mechanistically understandable.

It is worth noting that neither of the H3K4 methylations follows H3K36me3, and that their CI95 tightens comparing the treated versus untreated cells. Thus supporting, prior evidence to suggest that RNAPII-S2P transcribes primarily at the same distance relative to the network, regardless of the internal composition of the network, and that it may do so by loops of active genes extending away from the network surface. Markers of active chromatin are affected by the internal composition of the chromatin network and may be pushed to a narrower band closer to the network surface by increased volumes of dense chromatin in the network interior.

A mixed distance change can be observed for markers with ambivalent functions and those of inactive chromatin. These changes follow closely those changes outlined previously for the density distribution profiles. The key affected markers are H4K20me1, macroH2A and H3K9me3. Of these H3K9me3 has the largest difference between G_1 and treated cells, moving the average distance to the boundary further away towards the inside of the network, matching the increased enrichment in the higher density classes (Figure 6.9). Surprisingly the overall distance for HP1A is not changed, though the CI95 now protrudes more outside the IC.

Other markers such as H4K5ac, H3K27me3 and H3K9me2 also show very limited shift in average calculated distances from the surface, and the overlap of their CI95 remains almost full. It is difficult to interpret from this metric only how the chromatin compaction may affect these markers. However, by

coupling this with data from the density distribution (Figure 6.9) it is possible to say that although these markers do not migrate far from their G_1 position in terms of absolute distance they all become further enriched at the same densities in which they showed enrichment in G_1 . In fact of these three markers, H3K27me3 has the largest change in density distribution profile progressively from G_1 to hyper-osmotic perturbation. This is also the case in the changes shown by its distance measurements, with its average distance moving further internally in the network where higher density chromatin is usually located.

G_0 global chromatin compaction

Of course it is also possible to induce actual senescent phenotypes in cultured cell lines as opposed to trying to mimic the morphological effect via chemical or osmolar treatments. Senescence cannot be achieved by any particular cell line, so the ChaiN analysis was re-done at G_1 for the IMR90 cell line. These human lung fibroblasts are known to undergo cellular division with a high proportion of errors. Errors during the separation of sister chromatids leads to aberrant chromosome segregation and double stranded breaks in the genome. In this cell line, this often leads not to apoptosis but to the induction of a senescent programme, thus following the previously described replicative induced senescence (RIS) pathway.

To maintain a consistent comparison between these results and those for C127 the same antibodies were used. The specificity of some antibodies was not translatable from the mouse to the human system, yielding no or little detection. The targets of these antibodies are CTCF and Scc1 for the structural markers, H4K20me1 and Macro-H2A for the ambivalent markers, and HP1A for markers of repressive chromatin. No active markers were lost from detection by changing to the human system.

The density distribution of IMR90 cells at G_1 shows a similar pattern to that of untreated C127 cells (Figure 6.11(a), Figure 6.11(b)). Meaning, that for

markers of 3D genome architecture and markers of actively transcribed chromatin, the most enriched classes are classes 2-4. C127 distribution was narrower and mostly around classes 2-3, however most of the peak enrichments for IMR90 are also in class 3. Other relative patterns here are also preserved. Of the active markers, both RNAPII-S2P and H3K36me3 the ones enriched at lower density chromatin, just as they are for C127 at G_1 . The markers of inactive chromatin are all found enriched from class 3 up to the highest density classes. H3K27me3 and H3K9me2 still have a peak enrichment distributed around class 4. H3K9me3 shows a very strong preference to classes 5-7, much more so than that for C127. Despite some differences between the two cell lines, the general pattern is maintained (Figure 6.11(b)). This suggests a conserved mechanism of 3D epigenome organisation between different cell types and species.

The largest changes occur after senescence, G_0 (Figure 6.11(c)). The main difference between the G_1 and G_0 density profiles for IMR90s is the massive enrichment of all markers (save SAF-A) by several log2 units. There is a subtle resemblance of the change from G_1 to NaN_3 treatment profiles here, particularly for the extra depletion observed in many of the structural and active markers for classes above 4. The most similar of these being RNAPII-S2P which also yields a level of depletion close to 30 times less than what is expected by chance alone in class 7.

However, one key difference between G_0 and biophysical perturbation is that most of the markers have not undergone a shift in their peak enrichment towards lower density classes. And none, not even the non-chromatin proteins such as SAF-A, cohesin or the elongating RNA polymerase are found enriched in class 1. This is in stark contrast to these very observations described for both NaN_3 and hyper-osmolar conditions in C127 cells.

The markers of inactivity also show a very marked difference in their enrichment profile compared to the chemical and osmotic chromatin compaction induced in C127 cells. Here, under RIS, both H3K27me3 and H3K9me2 be-

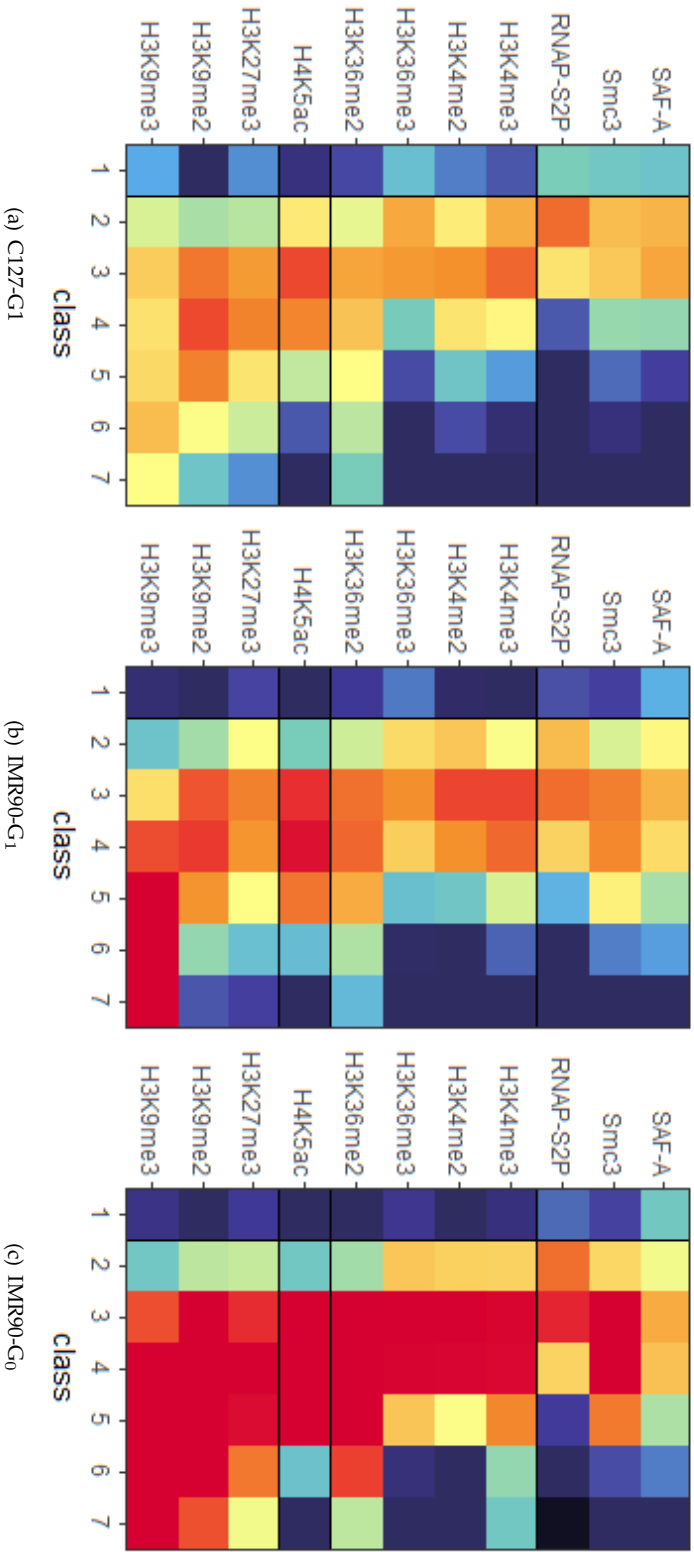


Figure 6.11: **Senescent changes to all density distributions.** Overview heatmaps of enrichment or depletion per chromatin class for most markers comparing untreated G₁ C127 and IMR90 nuclei or senescent (G₀) IMR90 nuclei. Markers are grouped by function from top to bottom: Structural, Active, Mid, Inactive. For details of individual markers see Appendix 9.4, Figure 9.8. Profiles with error bars can be found in Figure 9.8. $N_{total} = 660$ cells.

come massively enriched in classes 3-7, though still keeping the order of enrichment tail-off as with G_1 . This is to say that H3K27me3 decreases its enrichment crossing to depletion one density class lower than H3K9me2. This is both true in G_1 and G_0 .

What may be gained from comparing the downstream effects of compaction by different methods: NaN_3 , hyper-osmolarity and G_0 , is that not all of the 3D epigenome is affected to the same degree and simultaneously when differing mechanisms may be at play. One being from exogenous remodelling of a G_1 transcriptional programme in a very small time frame, while the second is established by a slower rewiring of the transcriptome to a G_0 state. While the observed chromatin topology may be morphologically similar even under super-resolution micrographs, the two rates of change may well cause the large differences observed in comparing true G_0 to chemically induced compaction in G_1 .

Single TAD labelling developments

Single TAD imaging

This investigation focused primarily on the global patterns of the 3D epigenome, both in relation to chromatin density and the surface of the chromatin network. The use of super resolution microscopy with *ChaiN*, yielded new insights into the nature of the chromatin landscape as well as the distribution of different epigenetic markers with regards to this landscape. Furthermore by exposing cells to different exogenous conditions and using *ChaiN* to measure these metrics again, this study has shown how deep-content image analysis can be used to shed light on the molecular mechanisms of this system.

Measurements at a global chromatin scale however can be subject to skew, and is therefore important to challenge the boundaries of imaging and labelling technologies in order to label individual genome loci. This would not only provide evidence that the model described here is indeed correct at the single TAD level, but may begin to untangle sub-TAD architecture. The ability to label single TADs or sub-TAD regions is therefore imperative to this

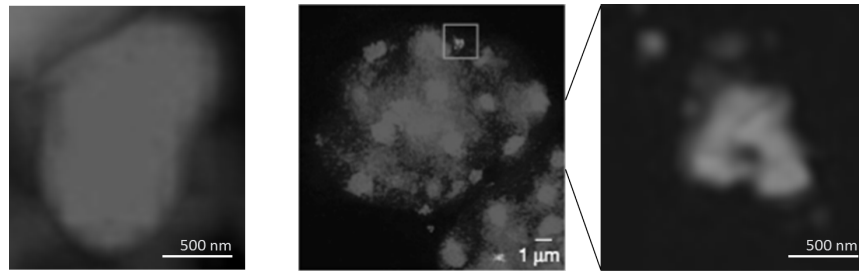


Figure 7.1: **Imaging TAD.** 1Mb TADs labelled with FISH and imaged with SIM from recent publications: Fab (Fabre et al., 2015) left, Nora (Nora et al., 2012) centre and right.

goal. Such labelling technique must be capable of labelling a single TAD with high efficiency, little-to-none cross labelling of other loci, and be amenable to multi-wavelength experiments. Furthermore, its dye must be robust and compatible with the demands of super resolution microscopy, as experiments probing these very small regions would be wasted on any other conventional microscopy technique unable to resolve structures at the 3D TAD-size scale or smaller. Recent publications have made use of FISH to label single TADs as defined by Hi-C in order to image them in super resolution microscopy (Figure 7.1). Both publications have suggested a TAD size of approximately $1\ \mu\text{m}$ to $2\ \mu\text{m}$ in diameter from SIM imaging (Fabre et al., 2015; Nora et al., 2012). Nevertheless combinatorial labelling of different sub-TAD regions in Nora et al. (2012) or the use of higher resolution STORM imaging in (Fabre et al., 2015) have also shown that the 3D structure of TADs is more complex than single spherical globules (although often depicted as such in renderings from *in silico* simulations).

In addition, this technique should be compatible with other techniques that label not chromatin but nuclear proteins or molecules of interest. This is important for analysing the location and function of these molecules in relation to the TAD architecture. Finally, for the eventual widespread use of such technique, it should be simple to perform, with reduced costs whenever possible, both economically and in the time taken for researchers to conduct the

experiment from start to finish.

Throughout the span of this investigation, several novel labelling techniques were implemented, optimised and or developed, in order to meet the aforementioned criteria. Their development along with their limitations are discussed subsequently.

7.1 LAD-ID imaging

Dam-ID is a technique developed for the identification of chromatin sequences which have undergone transient interactions with a particular target protein of interest (§2.5). Parallel studies to Hi-C using a variant Dam-LaminB fusion of the Dam-ID paradigm have shown that interactions between chromatin and the peripheral nuclear lamina also occur preferentially in domains of ~1Mb, termed lamina associated domains (LADs). Correlative bioinformatic studies on these datasets revealed a clear correlation between the domain boundaries of LADs and TADs, further supporting this modular architecture (Figure 1.6).

Imaging of LADs was performed using an adapted DamID system generated by Jop Kind who previously demonstrated LADs labelling by expressing a fusion of the bacterial m6a-binding protein domain to enhanced GFP (Figure 7.2, Kind et al. (2013), see §2.5).

Transient transfection of the LAD-ID system in C127 cells produced bright signals from high expression (Figure 7.2). However, overexpression of LaminB1 has been shown to change nuclear lamina morphology (Volkova et al., 2012). Despite starting IF staining protocol within 24 hours after transfection, the signal from GFP alone was not robust enough for high-quality SIM-reconstructions. These are limitations of the photochemistry of the fluorophore used, rather than expression efficiency. Alexa-488 conjugated antibodies with affinity to an GFP epitope were used to counter the bleaching and increase reconstruction fidelity. Phenotypes not characteristic of these cells

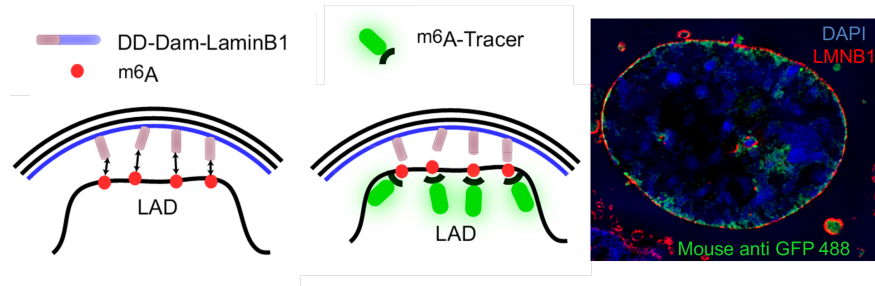


Figure 7.2: **LAD-ID**. Schematic of LAD-ID showing deposition of methyl-6 marker by Dam methyltransferase fused to lamin B on chromatin within molecular interaction to the nuclear periphery (left). A GFP-fused m6A marker tracer is then able to attach to the marked chromatin even when no longer in contact with the periphery (centre). Adapted from Kind et al. (2013). Right: 3D-SIM image of LAD-ID transfected C127 cell expressing the LAD detection system fixed and stained with DAPI, and directly conjugated A488-Mouse-anti-GFP primary antibody (to boost GFP signal) and two-antibody detection of LaminB1. Nucleus shows morphological lamina aberration such as invaginations and micronuclei.

such as: multiple micronuclei, a distorted lamina and numerous invaginations were observed, suggesting a state in which labelled chromatin may not reflect chromatin networks in physiological conditions (Figure 7.2, right).

A stable line of human LS174T colon adenocarcinoma cells (provided by Jop Kind) were used. As expected, stable integration decreased the relative expression levels of the LAD-ID system compared to transient transfections, but these cells did not show abnormal nuclear morphologies. Nevertheless this cell line presented new challenges. Firstly they were highly extended in shape away from the coverslip (10 μm to 15 μm in height), and secondly they did not form single monolayers on coverslips, but stacks of clonal colonies. A LAD-ID stable line of HT1080 cells were ultimately chosen, as they are less z-extended (7 μm to 12 μm) and form monolayers. Thus the HT1080 cell line was more suitable for 3D-SIM imaging with enhanced stripe modulation and decreased out of focus blur.

However, in spite of the optimisation for SIM imaging of these LAD-ID labelled domains, the key limitation of using such method is its stochasticity. It is known that certain genome regions have a higher propensity to interact

with the lamina, yet in single cells it is not possible to define which regions of the genome will be lamina associated. This is confounded by the calculation that the internal surface provided by the nuclear envelope is not sufficient to accommodate all LADs identified by averaging a large pool of cells from the Dam-ID methodology. As such, in single cells it is impossible just with this stain to ascertain whether a full LAD is being labelled, as opposed to sub regions of a LAD, and if the full LAD is indeed labelled which LAD/TAD it might be on the genomic linear scale. The latter is required to link 3D position to the genomic function of this chromatin domain.

Co-localisation of LAD-ID labelling with mid S phase EdU pulse replication labelling can be used to confirm such stochasticity. Replication firing has been demonstrated to be tied to TAD architecture with multiple origins of replication synchronously firing from within the single megabase range size scale of TADs (Pope et al., 2014; Chagin et al., 2016). LADs and mid S phase replication domains are enriched in facultative heterochromatin and both are found at the nuclear periphery. LADs can also be found at the nucleolar periphery where a subset of these replication domains are also found. Using 30 min EdU pulses as a measure for the extent of LAD labelling is useful as a replication pulse length of 30 minutes is enough to label an individual TAD-size foci. It is clear that although labelling with LAD-ID can closely mimic the pattern observed for replication domains, it does not do so to the same extent (Figure 7.3). Ultimately the shortfall in reproducible coverage of full genomic domains places a severe-enough limitation on this approach to be a viable method for single TAD labelling.

7.2 CRISPR imaging

The clustered regularly interspaced short palindromic repeats (CRISPR) technology is a two component system of bacterial adaptive immunity (Barrangou et al., 2007). The mechanism of the type II CRISPR system involves the

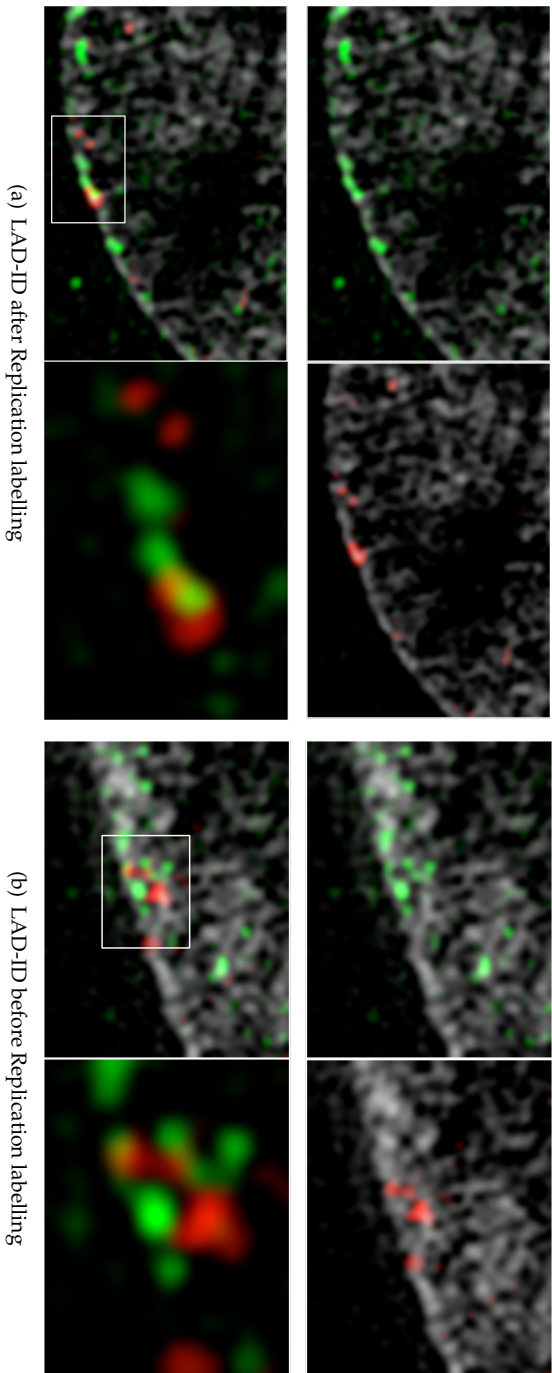


Figure 7.3: **LAD-ID vs RD.** Zoomed single z planes of 3D-SIM images showing stochasticity of labelling in both LAD-ID (green) and replication domain staining (red, RD). (a) When LAD-ID is performed after RD in mid Sphase many foci can overlap, showing direct evidence for the argument that mid Sphase RD and LADs may describe the same unit of chromatin organisation. (b) Further support comes from reversing the order of the stain. In this inverted order foci have reduced overlap. This can occur if the markers are competing for the same nuclear loci. LAD-ID is a non-covalent interaction from the m6A tracker and can be removed by the progress of DNA polymerase during replication (the process by which new traceable EdU is incorporated) Thus seeing decreased overlap in this labelling order is suggestive of labels interacting on the same region.

use of a single nuclease protein (Cas9) which is able to bind specific DNA sequences when in complex with a single guide (sg)RNA molecule. sgRNAs are a construct of the fusion of target-complementary RNA and the tracer RNA (Deltcheva et al., 2011). As such they have a 5' end complementary to the target sequence, a conserved hairpin loop secondary structure used to bind the Cas9 nuclease.

The ability to target a nuclease to loci of interest has significantly increased our ability to manipulate chromatin and perform precise genome engineering. Accurate deletions and knock out experiments have been undertaken with the use of two flanking RNAs to the genomic region of interest, in combination with *in vivo* expression of active or partially altered Cas9 for double strand cuts or single strand nicks respectively. These tuneable active Cas9 varieties along with fully inactive “dead” (d)Cas9 mutants in combinations with tolerated protein fusions have made CRISPR a multipurpose tool for investigating eukaryotic chromatin structure and function. Current applications span traditional genome engineering, transcriptional interference, recruitment of histone modifiers and transcription factors, and an imaging tool with potential for *in vivo* labelling of single loci (Sander and Joung, 2014; Barrangou and Doudna, 2016).

Proof of principle for a CRISPR-based imaging modality has been described using a GFP-fused construct on the dCas9 nuclease, targeted to telomere and pericentromeric satellite repeat regions (Chen et al., 2013; Anton et al., 2014). The current CRISPR imaging methodology is non-denaturing but the technique suffers from the inherent photo-physical limitations of GFP, as well as a low signal-to-noise ratio with only few binding events decreasing its capacity to discriminate single loci. Furthermore, multiplexing CRISPR labelling with different colour wavelengths is not possible without transfection of multiple fusion constructs of Cas9 orthologues (Chen et al., 2016; Ma et al., 2015). This is because each Cas9 orthologue has a specific recognition to a guide RNA. Thus using multiple guides for different targets generated from

the same guide RNA type will only lead the same Cas9 to both loci making them indistinguishable microscopically. Other Cas9-based labels have been generated by replacing the fluorescent protein fusion with a HALO tag (Knight et al., 2015). This is useful in order to detect single binding events with super resolution as the HALO fluorophore is more stable than fluorescent proteins.

In this study, an alternative CRISPR based imaging modality was tested. This relies on the *in vitro* incorporation of fluorescent moieties in the guide RNA (Figure 7.4). Multiple labels per binding would increase the signal-to-noise ratio and allow low-copy and non-repetitive loci to be imaged. The purpose for developing such technique would be the ability to specifically label individual TADs in a manner non-denaturing to all chromatin, thus preserving both the chromatin network's structure as well as the proteome for multiplexing IF signals of the epigenome around single TADs. A similar adaptation of CRISPR using the RNA as the labelling molecule has been published utilising an extended sgRNA (Qin et al., 2017). The extension does not interfere with Cas9 recognition or binding and is comprised of multiple MS2 RNA loops. A fluorescent protein fusion construct with the MCP protein able to recognise MS2 renders a high density and specificity of labelling. Although this method still relies on fluorescent proteins, the density of labelling allows imaging of non-repetitive loci in confocal microscopy (Qin et al., 2017).

To be able to label single TADs this investigation required four major milestones: sgRNA design, synthesis of labelled constructs (Figure 7.4), *in vitro* control of sgRNA:protein function and *in situ* delivery to live cells (§2.5). Initial developments were made using a sgRNA against mouse major satellite repeats (MSsgRNA), as used to publish the previous CRISPR labelling system with FISH validation (Deng et al., 2015). This allowed multiple iterations of tests on label incorporation, labelling densities and delivery protocols with comparable data.

During *in vitro* synthesis of the MSsgRNA (as described by Bassett et al.

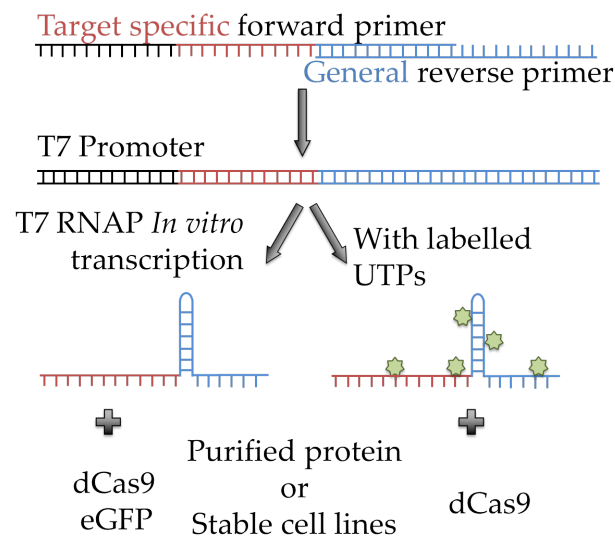


Figure 7.4: **CRISPR methodology.** Schematic of protocol used for *in vitro* synthesis of labelled and unlabelled sgRNA for combination with GFP-fused or non-fused Cas9 protein respectively. Adapted from Bassett et al. (2013).

(2013)) the sgRNA was labelled with a selection of NTP homologues (§2.5). Afterwards, an *in vitro* cleavage assay was used to control the viability of the labelled sgRNA molecule population (Figure 7.5). The labelled RNA by T7 RNAP transcription should not affect secondary structure of the sgRNA or its 5' end as these are key for Cas9 recruitment or hybridisation respectively.

Running an agarose gel of the *in vitro* cleavage assay results shows which of the incorporated labels (and at which concentration) are still compatible with Cas9 function (Figure 7.6). The expected result is that the sgRNA should be able to bind and guide the Cas9 to two sites on a plasmid, flanking a fragment of DNA of known size, here the open reading frame (ORF) of GFP. Thus when a band at 700bp is detected (the size of the GFP ORF) the sgRNA is successfully compatible with Cas9 function despite its label.

The concentration of sgRNA is sufficient to be detected as a smear at the bottom of the gel. If this smear is not present, the T7 polymerase was not able to conduct sgRNA synthesis, meaning the size of the dye/label or its concentration was detrimental to RNA synthesis. A good control showing this

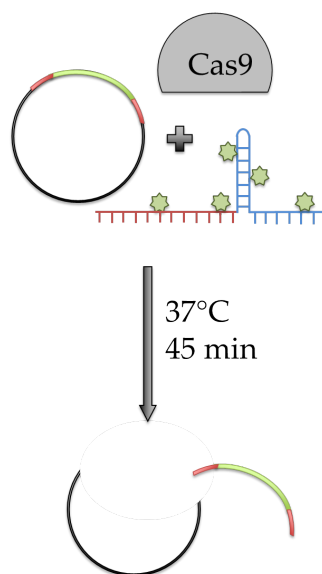


Figure 7.5: ***in vitro* cleavage methodology.** Protocol for controlling the viability of labelled sgRNA. Using catalytically active Cas9 nuclease in conjunction with sgRNA and a plasmid with two sgRNA target sites allows a fragment of known size to be excised if the viability of sgRNA is not affected by *in vitro* label incorporation.

can be performed by incubating the Cas9 and target plasmid with a sgRNA that does not base pair to the flanking target sequences. In such lane, the smear is present but no fragment can be detected (Figure 7.6, negative). From this assay, a 1:6 dilution ratio of labelled NTP was found to be the limit for sgRNA synthesis and Cas9 function while still yielding detectable sgRNAs. Not all labels were incorporated equally, possibly limited by their size and by their moiety's chemistry. Thus out of all the tested labels, EU, Biotin and Alexa 488 labels could produce sgRNA (Figure 7.6).

After the synthesis of labelled MSsgRNA, these were transfected into cells after pre-complex formation *in vitro* with purified dCas9. However, none of the large RNA-protein complexes could be successfully transfected with high-enough efficiency for fluorescence detection (after fluorescent Click reaction of the EU labelled MSsgRNA or fluorescent streptavidin detection of the Biotin MSsgRNA). The issues with delivery of dCas9-RNA complex with cationic

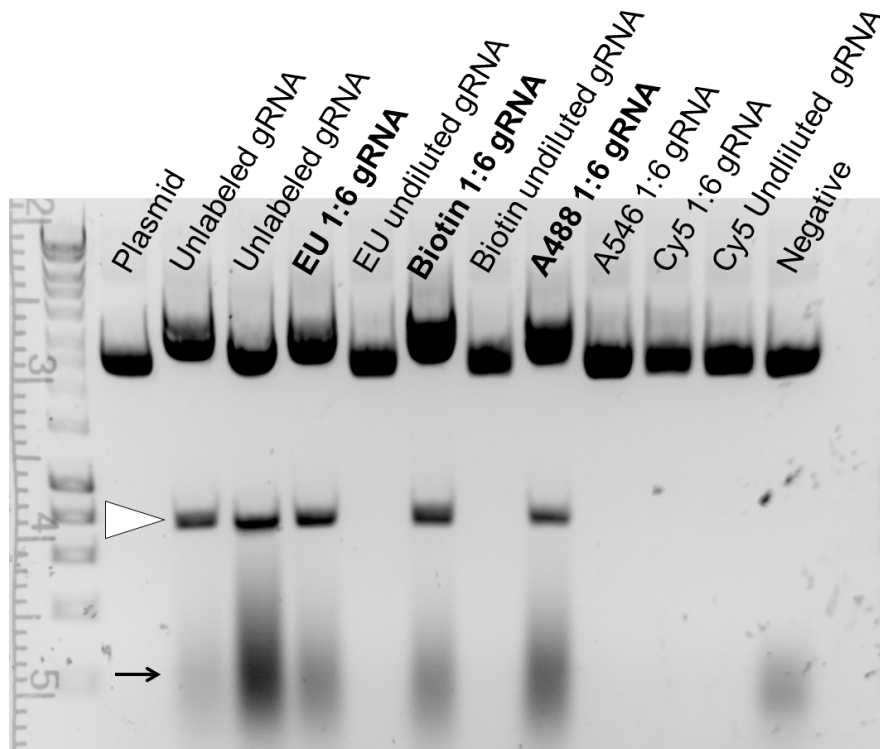


Figure 7.6: **Results of cleavage assay.** Electrophoresis gel results of a range of tested labelled sgRNA. First and last lanes show controls of plasmid + Cas9 only or with a sgRNA not homologous to the flanking target sites in the plasmid, black arrow shows excess sgRNA bands. White arrowhead shows size of excised band in wells whose sgRNA is still viable, of these, the labelled sgRNA are highlighted in bold font.

transfection reagents was surprising given previous studies showing the efficiency of this method when evaluated by the downstream genome editing of active Cas9 (Zuris et al., 2015). However, to observe the effects outlined in such studies, a single Cas9 molecule and a single sgRNA molecule are technically required, given days of incubation. However for labelling, many multiplexed protein and RNA molecules are needed.

Large quantities of the labelled complexes were found in the cytoplasm, by this method but not in the nuclear volume. To ascertain whether this was due to the chemistry and size of the label or that of the dCas9 protein, two stable cell lines were created expressing variants of dCas9. One cell line ex-

pressed a dCas9-GFP used as a control for the transfection of unlabelled MSsgRNA and the second line expressed unlabelled dCas9 to allow detection of the tagged MSsgRNAs. Furthermore, simple cationic lipid based transfections were replaced by an optimised commercial electroporation protocol “Nucleofection”, from Lonza. A range of cell line specific electroporation reagents can be browsed and purchased, each accompanying a dedicated electrical pulse protocol for achieving an empirically defined highest electroporation efficiency. This however is also dependent on the proprietary hardware for electroporation a “Nucleofector”.

Nucleofection can be achieved for cells suspended in solution, as well as for adhered cells on coverslips. Both approaches were tested and found to work for the electroporation of fluorescently labelled dUTP, with nucleofection for cells in solution yielding a higher nucleofection efficiency. Adherent cell nucleofection should be used preferentially over nucleofection in solution when subsequent live cell imaging is required immediately after electroporation as there is a minimum amount of time taken for cells to re-adhere to the coverslip for imaging. However in this investigation requiring fixed cell imaging, the higher efficiency of incorporation, reduced cost and higher practicality were sufficient reasons to proceed with nucleofection in solution.

The results from these experiments were as follows. Unlabelled MSsgRNA was successfully transfected to the nucleus of dCas9-GFP cells, causing their GFP enriched nucleoli pattern (Figure 7.7, Fix top) to migrate to a GFP enriched chromocentre pattern (Figure 7.7, Fix middle). This control showed that a sufficiently efficient transfection of sgRNA was possible via this method in order for enough labelled molecules to localise at the target sites (major satellites at chromocentres) and be detected. For the A488-MSsgRNA labelled sgRNA the same successful chromocentre labelling pattern was detected after transfection into unlabelled-dCas9 expressing cell line (Figure 7.7, Fix bottom). Neither Biotin or EU labelled MssgRNA could be successfully transfected to the nuclear interior in order to replicate the positive control.

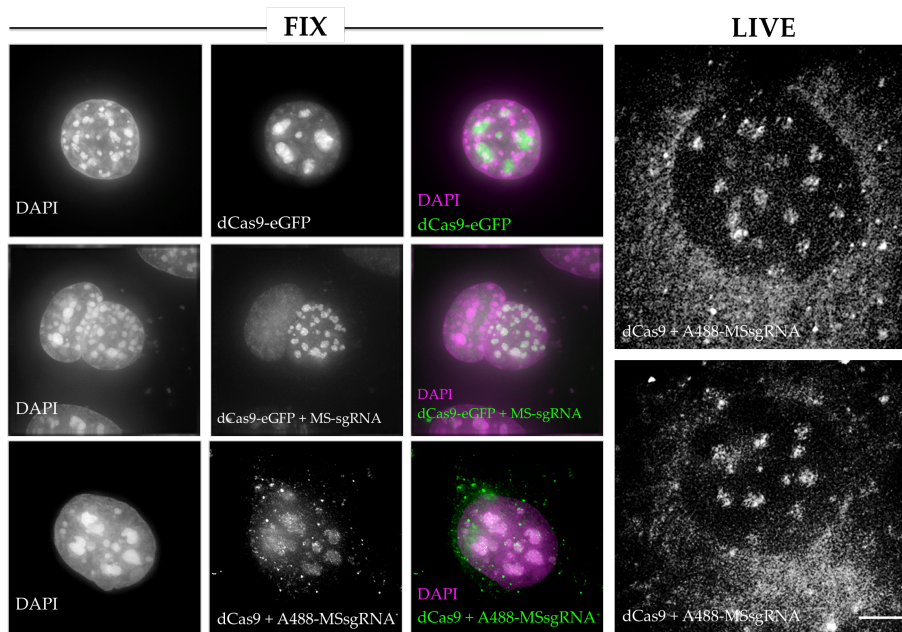


Figure 7.7: **CRISPR-delivery. Fixed:** Top panels: Mouse C127 somatic cell stably overexpressing dCas9-eGFP localised at nucleoli (control). Middle panels: Positive control of same stable line transfected with plasmid encoding MS-sgRNA with signal colocalised to chromocentres. Bottom panels: C127 stably expressing dCas9 nucleofected with A488-MS-sgRNA showing positive results of pericentromeric colocalisation. **Live:** live cell imaging of same experiment as fixed bottom panels 4 hours after nucleofection also reveal pericentromeric pattern. All images are Z-projected from widefield data, deconvolved. Scale bar 5 μm .

As mentioned previously, adherent cell nucleofection can be performed if subsequent live cell imaging is required immediately after electroporation. This nucleofection protocol was used to check the viability of a CRISPR labelled live cell imaging experiment. In principle, labelled sgRNA could be imaged in live cells if directly labelled with a fluorophore, not requiring fixation and antibody based epitope detection. Indeed, when A488-MssgRNA was nucleofected into adherent dCas9 expressing cell line the labelled chromocentre pattern was replicated, characterised by more labelled foci of smaller size than the few and larger nucleoli (Figure 7.7, Live). However, the efficiency of labelling by this approach was far lower than required for live cell imaging. A small signal to noise ratio could be detected, even with confocal microscopy.

Attempts at super resolution SIM imaging of these datasets proved futile in yielding useful images.

Major satellites are large tandem repeats of the same target sequence, over several tens and hundreds of megabases, making for an ideal target to test the development of this labelling technology. Nevertheless, labelling of individual TADs would require multiplexing many thousands of sequence specific sgRNAs in order to label only a 1 Mb region. Given the results obtained from MSsgRNA the likely outcome of labelling a region at least one or two orders of magnitude smaller would not produce samples of high-enough label contrast for SIM.

The development of CRISPR based tools for genome editing, epigenetic control and *in situ* labelling are all still at their infancy. Although the continued optimisation of this particular sgRNA-labelled approach was beyond the scope of this investigation it is possible that this may be the way to achieve non-denaturing single TAD labelling for super resolution imaging in the near future. A way to bypass some of the major delivery issues could be by the chemical synthesis of sgRNA with a wider fluorophore palette selection. Here, the limitations of the T7 RNA polymerase could be lifted so that larger dyes may be used. Chemical synthesis is also independent of template plasmids thus the cost and effort of generating thousands of sequence-specific sgRNA for multiplexing would be significantly lower. Another advantage of the chemical synthesis route is the ability to know with precision the location of the incorporated tag as well as the number of tags incorporated per sgRNA. With this information it would be easier to control the viability of labelled sgRNA but more importantly, the known number of molecules allows their titration and to work out, given relative brightness, how many sgRNAs are found at target regions.

Another possibility is the use of this method in bacterial systems. Bacteria lack nuclear envelopes and therefore would not suffer the problems encountered for delivery of the sgRNA-protein complex to cells via electroporation.

7.3 Raser FISH imaging

After the limitations of both LAD-ID and CRISPR based labelling for imaging single TADs in SIM, another method was required. FISH labelling of individual TADs has already been shown to be a viable approach (Nora et al., 2012). However, FISH, even at its mildest, involves the large scale denaturation of DNA. DNA denaturation is of course unavoidable for selective labelling of the genome as the double stranded molecule must be separated in order for probes to bind to their complementary sequences by base pairing. However, for FISH this DNA denaturation step is often done in a way which also irreversibly denatures the proteome of cells, therefore affecting the target epitopes for antibody detection. Thus the use of all such disruptive FISH protocols was not compatible with mapping the 3D epigenome by immunofluorescence as has been shown in this investigation.

What would be required in order to use the known labelling capabilities of FISH at the TAD level in combination with the ChaiN protocol would be a FISH DNA denaturation that is non-disruptive to the proteome. Fortunately, one such “non-denaturing” FISH variant has recently been developed by Dr Jill Brown in the lab of Professor Veronica J Buckle as a combination of the interests of the Molecular Medicine and Wolfson Imaging groups of the Weatherall Institute of Molecular Medicine, Oxford. Originally developed with the aim of understanding the relevance of nuclear organisation to gene expression during erythropoiesis, this FISH variant uses light at the ultraviolet end of the visible spectrum to induce nicks in double stranded DNA, which is then made single stranded by a specific exonuclease. This technique is also compatible with super resolution imaging and as such has been termed RASER-FISH, for “resolution after single-stranded exonuclease resection” FISH (§2.5).

The FISH probes used to test RASER-FISH for full TADs were the same probes previously published by the research group of Professor Edith Heard

for covering neighbouring TADs in the X chromosome (Nora et al., 2012). By using these, the specific sequence required for full coverage of TADs has already been verified. The probes were generated with incorporated digoxigenin for subsequent anti-digoxigenin detection IF as described in §2.5. Although denaturing FISH has been used for full TADs before, the combination of this labelling method with super resolution is still in its early days. A recent publication using both SIM and STORM microscopy was able to reconstruct individual TADs of 1Mb to approximately 2 μm in diameter (Figure 7.1, left) (Fabre et al., 2015).

The initial result from the combination of Raser-FISH and SIM on single TADs was very promising. Comparing the same TADs from prior publications as well as other TADs of the same size it is clear that the labelling, SIM acquisition and reconstruction in this investigation give a much higher resolution image of these TADs (Figure 7.8). The overall size of TADs is reduced from what has been reported in other studies, with TADs being approximately 500 nm in length and down to 200 nm at the narrowest. This is in agreement with the reported average network filament width of 526 nm from previous *ChaiN* parameters (Figure 4.4). This resolution critically reveals a sub-TAD organisation (eluded to in Fabre et al. (2015)). TADs are not just amorphous round or oval signals, but display intricate 3D patterns, as well as a high level of diversity in their 3D structure between TADs, even for those in the same cell (Figure 7.8). This is very likely due to the different compaction of the active and inactive X chromosome where these different TADs reside.

The original Raser-FISH protocol was developed for detection using the FITC dye. However, more robust dyes are now available (eg Alexa 488) and these are generally chosen over FITC. Here, Raser-FISH was trialed with optimised A-488 detection. The results show a much higher signal to noise ratio can be achieved as theorised (Figure 7.8, bottom left). This extended dynamic range translates to a cleaner background, of lower intensity, for the same signal brightness of the TAD in question.

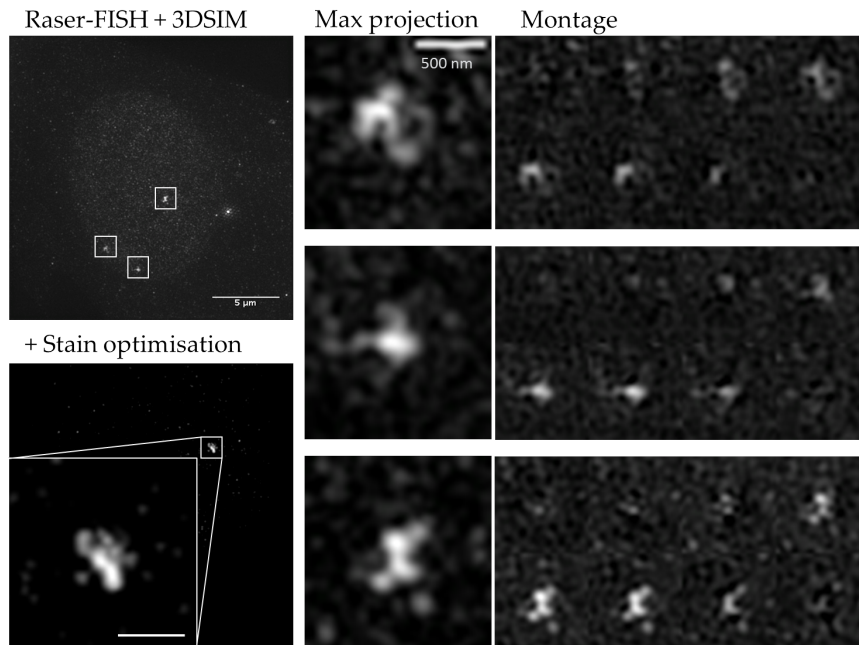


Figure 7.8: **Raser FISH**. Top left: A higher resolution can be achieved from 3D SIM in combination with Raser-FISH. Here the same FISH probes for the TAD labelled on Figure 7.1 (right) have been used to have a direct comparison. Zooming in and displaying max projections and montages expanded from on the white boxed regions from top image (and displayed to the same scale as TADs in Figure 7.1) show much better structural resolution for image analysis. The TADs take many intricate 3D shapes but maintain a radial diameter no larger than 500 nm. The Raser-FISH can be optimised with organic dye labelling to achieve a higher contrast benefiting 3D SIM (bottom left). This is clearly observed by comparing the bottom left inset of max projection with any of the three non-optimised max projections (centre column) magnified to the same scale.

Finally Raser-FISH was attempted in combination with IF (Figure 7.9). This is a requisite in order to make this labelling method compatible with classification of the epigenome around TADs. The results confirmed the hypothesised preservation of the proteome and epitopes by the Raser-FISH approach. a successful IF of transcription factor Islet1 (ISL1) shows a clean IF pattern for this transcription factor in combination with Raser-FISH TAD labelling of the X chromosome TAD (Figure 7.9). It is worth noting that the ISL1 transcription factor was chosen for its relatively high abundance without being too highly concentrated that individual spots could be resolved. ISL1 is not known to

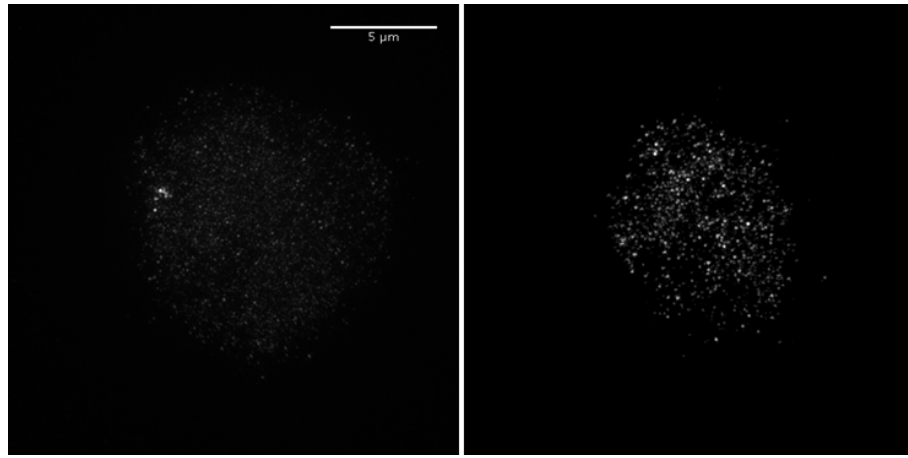


Figure 7.9: **Raser FISH vs IF.** Raser-FISH (left) does not disrupt the proteome (right, IF) when denaturing DNA for FISH probe incorporation. This means epitopes are preserved for detection by IF. Here Raser-FISH is demonstrated in combination with an IF for the ISL1 TF in a single cell to demonstrate proof of principle.

have any interactions with the genomic region labelled in this TAD as it is a factor for the expression of the *Mnx1* gene in motor neuron differentiation from embryonic stem cells (Lee and Pfaff, 2003). Therefore it is not expected that the signals from the IF correlate strongly, if at all, to the interior or surface of this TAD.

Although the implementation of this developed technique is beyond the scope of this thesis, its future application with super resolution microscopy has potential to impact the field of chromatin biology and epigenomics. The structure of TADs, which are still shown in diagrams as an amorphous blobs of disorganised DNA or chromatin, will begin to be understood. Large-scale quantitative experiments like the ones performed in this investigation, but at the single TAD level should reveal the differences in position between constitutively active or inducible promoters within the same TADs, as well as how the proportion of active versus inactive chromatin within TADs affects their overall shape, size and density. Dual labelling of TADs and their sub-regions, such as promoters, gene bodies and intergenic regions would give a clearer understanding of how these regions of diverging function interact within the

3D constraints of a single TAD. Multiplexed labelling of TADs and nuclear machineries, such as transcription factors and RNA polymerase, but also the epigenetic writers of the histone code (eg: methyltransferases) as well as the progression the replication machinery during S phase should yield an even deeper understanding of the accessibility model proposed from this global study at the TAD scale.

The answers to these questions can be asked at the single cell level and not as an average ensemble of contact probabilities from a pool of cells. With the development of single-cell Hi-C variant approaches, this deep image analysis method would be ideally suited for complementary investigations of individual loci of interest at a scale and resolution never before achieved.

Discussion

8.1 Aims and ChaiN developments

Three principal goals were attempted in this thesis. First and most crucial, the development of a high-throughput semi-automated image analysis pipeline for the quantitative analysis of super resolution micrographs. Second, the implementation of this pipeline to analyse the mammalian nuclear volume. Specifically, using SIM to resolve the chromatin landscape in conjunction with immunofluorescence detection of histone modifications and key epigenetic markers. From hundreds of such datasets, to obtain the average attributes of both chromatin and the 3D epigenome in single cells. Comparisons of these metrics in the G_1 state versus a variety of factors for the perturbation of chromatin (both exogenous and endogenous) could in principle lead to a better understanding of the architecture of chromatin at the TAD size-scale with relation to its genomic function. Finally, methods for the labelling of individual TADs for super resolution studies were developed and optimised.

Many of the aims set forth for this thesis were achieved, being successfully completed in the given timeframe. A semi-automated image analysis pipeline termed ChaiN (for ChaiN analysis of the *in situ* nucleome) was de-

veloped. ChaiN is able to capitalise on the increased spatial information from super resolution micrographs. By employing automation, it also boasts high-throughput rates for the rapid analysis of hundreds-thousands of 3D datasets with overall millions of datapoints. The capacity to process large numbers of samples allows for the first time a robust statistical analysis of the spatial environment acquired. One major choke-point of this investigation was the individual acquisition of thousands of micrographs. Despite the speed of SIM acquisition, the acquisition and reconstruction of thousands of nuclei still amounts to hundreds of man hours at the microscope. As automated acquisition for super resolution becomes optimised and more established, this rate-limiting step will have a decreased impact on the overall workflow.

The ultimate purpose for the development of ChaiN was not solely for the results obtained in this investigation. The measurements and metrics which can now be obtained as shown in this study demonstrate that ChaiN and ChaiN -modified pipelines can be effective to explore super resolution three-dimensional datasets for characterising nano-scale environments, both inside and outside of the nucleus.

8.2 Chromatin landscape

Attempts at describing chromatin depend on the resolution and labelling methods employed. Studies using electron microscopes, at a much higher resolution than their fluorescence counterparts, have shown that chromatin may be bundled as fibres of 30 nm in diameter (Figure 1.1)(Robinson et al., 2006). These “30 nm” fibres served to highlight that at the nano scale chromatin may be organised in higher order 3D architecture. This posed the question of how would such higher order organisation accommodate all the known (and yet unknown) genome functions described by traditional linear-genome biochemical assays, such as Chromatin immuno precipitation (ChIP). The presence of these “30 nm” fibres *in situ* under physiological conditions is

now debated (Maeshima et al., 2010) by many studies which support the view that these fibres are a crystalline product of the fixation methodology used for EM. The field is still not in consensus as to what the true model for higher order chromatin folding is, but chromatin is mostly portrayed as disordered strands which may exhibit fractal patterns (fractal globule) of local cohesion (Mirny, 2011). Some in the field have also begun to question how such globule domains of chromatin interact not just with other regions of the genome but with other non-chromatin molecules of the nuclear volume.

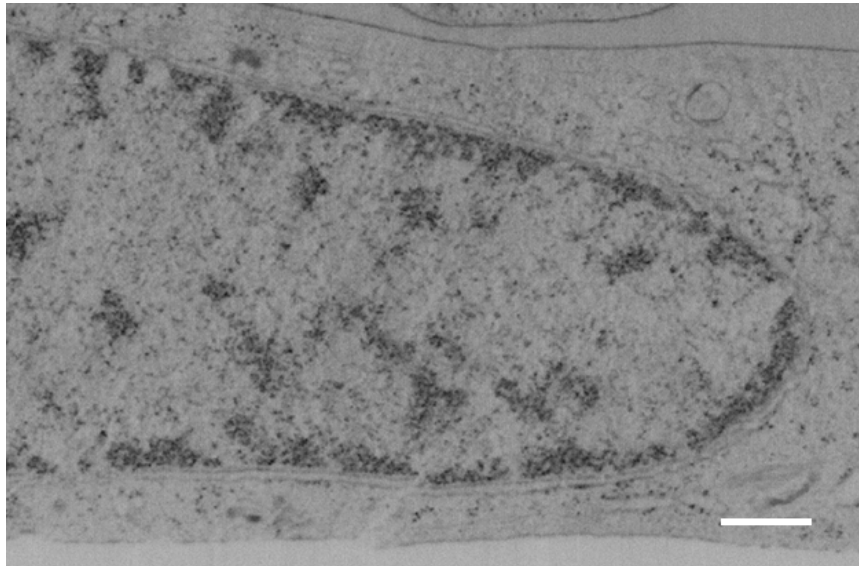
When considering these interactions, the general accessibility of chromatin is an important factor. A droplet model of chromatin with a gradient of densities has been proposed from *in silico* modelling of chromatin as a melted polymer (Maeshima et al., 2016). It is important to remember that although chromatin may be remodelled locally to small individual “10 nm to 11 nm, beads-on-string” fibres, DNA is an extremely large molecule overall. As such, any proposed higher order organisation not only requires a high-enough efficiency of packing (for the whole genome to fit inside the nucleus of each cell) but also balance the constraints on flexibility. Much like rope and cordage, multiple iterations of winding individual small fibres yields a structure which has decreased flexibility and higher tensile strength.

Recent work combining electron microscopy tomography (EMT) and a novel chromatin dye has also failed to visualise “30 nm” fibres *in situ* (Ou et al., 2017). Instead it has shown DNA organised as disordered chains of 5 nm to 24 nm in diameter. Notably, they have found that these chains, putatively composed on nucleosomes (as beads-on-string), are flexible at various lengths and can achieve many levels of chromatin compaction by their packing density. Such EM studies are not without their limitations as they must have strong fixation in glutaraldehyde followed by dehydration and embedding at room temperature. Lack of cryo conditions in EM can lead to alteration of the chromatin structure. Furthermore, it is clear that in this technique nucleoli are densely stained despite containing no DNA. This calls into ques-

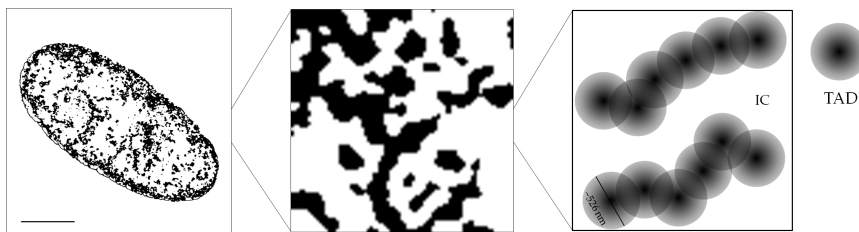
tion whether much of the non-nucleolar signal is indeed purely chromatin or a mix of chromatin with coagulated ribonuclear proteins. CryoEM work by the lab of Herald Hess have also produced imaging of chromatin *in situ* (Figure 8.1(a)). Through cryogenic preservation the results show chromatin in a network approximately 200 nm to 500 nm wide. The discrepancy between the two EM techniques may be due to loss of ultrastructure preservation in non-cryogenic conditions, leading to an unravelling of this large-scale network into a more homogenous nuclear environment where the highest chromatin folding structure is beads-on-strings.

A model that reconciles the single TAD view, the holistic “droplet of chromatin” model (Maeshima et al., 2016) and the observations of disorganised 5 nm to 24 nm chains would be one in which the chromatin compartment is made up of strung-up TAD domains. Each TAD organised within itself as the fractal globule predicts made from loops of beads-on-string 5 nm to 24 nm chains. The areas in which TADs contact each other would maintain a high chromatin density and many TADs together would start to form long filamentous structures with a high surface area (Figure 8.1(a)). This surface between chromatin and non-chromatin environments would be composed of decreasing chromatin density, with increased accessibility, as hypothesised by the melted polymer.

From the results obtained by Cha/N in this thesis, chromatin in single cells can be thought of as a convoluted network taking up approximately 65% of the nuclear volume and with filaments ~526 nm in diameter on average (Figure 8.1(b)). From whole-chromatin labelling experiments, chromatin shows a gradient of intensities with the majority being less compacted at the interphase region between chromatin and non-chromatin regions. This is in agreement with a droplet model of varying density. Results from the single TAD labelling developments produced evidence in support of single TAD based architecture. TADs maintained a singular domain no larger than 500 nm in maximum diameter. However single TADs showed not a simple spherical



(a) CryoEM image of chromatin



(b) Proposed arrangements of TADs

Figure 8.1: TAD filament network model. (a) Cryo electron microscopy slice of a nucleus of a U2OS cell showing chromatin *in situ*. A network of ~ 200 nm to 500 nm wide compact chromatin regions can be seen, predominantly at the periphery of the nucleus next to the double nuclear envelope (also visible). Interestingly, chromatin domains of the approximately similar size can be seen in the nuclear interior surrounded by large volumes of non-chromatin (IC) space. This image is reproduced courtesy of Herald Hess. Scale bar = $500 \mu\text{m}$. (b) Binary image of a C127 cell with all non-chromatin voxels as white. From the metrics obtained by the ChaiN analysis of hundreds of super-resolution 3D SIM micrographs the G_1 chromatin network can be described as a composed of filaments which are on average 526 nm in diameter. Single TAD labelling with the most optimised FISH staining (Raser-FISH) and super resolution also shows TADs have a maximum diameter of approximately 500 nm. A feasible combination of both observations would be the hypothesis that chromatin in single cells is arranged in a filamentous network of TADs strung one after the other. The predicted partial overlap between neighbouring TADs in single cells yields a continuous core of dense chromatin as the backbone for the chromatin filament. Scale bar = $5 \mu\text{m}$.

globular shape, but mostly a distorted globule stretched into sub domains. The width of these being approximately 200 nm. Such architecture of TADs fits with the previous filamentous network description. TADs are still entities of high 3D spatial contact predominantly within a radial volume of 250 nm, yet their shape often shows a finer filament-like arrangement.

How this network of adjacent TADs with decondensed-surface to the IC reacts to global change also provides insight on chromatin architecture. From compaction perturbations followed by *ChaiN* analysis carried out in this thesis, a marked and biased change was quantified. Increased levels of chromatin compaction collapsed the finer filament network into broader filaments with average widths of 744 nm. These changes did not affect the proportion of nuclear volume occupied by chromatin. However it is worth pointing out that the whole nuclear volume is inferred from the chromatin stain §9.4. This is a limitation of the technical set up as all light-paths were used in a quadruple stain, see §9.3. For a truer representation of the entire nuclear volume a lamina-stain should also be used. Nevertheless, as expected, broadening filaments lead to a marked decrease in the surface area to volume ratio of the chromatin network. Because lower density chromatin (classes 2 and 3) was primarily found at this interphase boundary, classes 2-3 made up a decreased proportion of chromatin during compaction compared to G_1 . As mentioned, the overall proportion of chromatin remained constant and as no DNA is lost during compaction, the higher density classes increased in volume. Interestingly the increase was most notable for the lower end of these classes (4 and 5), supporting the notion that the addition % for these is due to a direct result of the loss of the lower density classes in the vicinity. In other words, chromatin in classes 2-3 becomes compacted to the point of becoming classes 4-5 but there are no mechanisms pushing more of classes 4-5 to classes 6-7

From the current model it is possible to interpret these observations as TAD filaments that had previously been kept apart, now forced together, decreasing the available surface area of each filament (Figure 8.2). Regions in

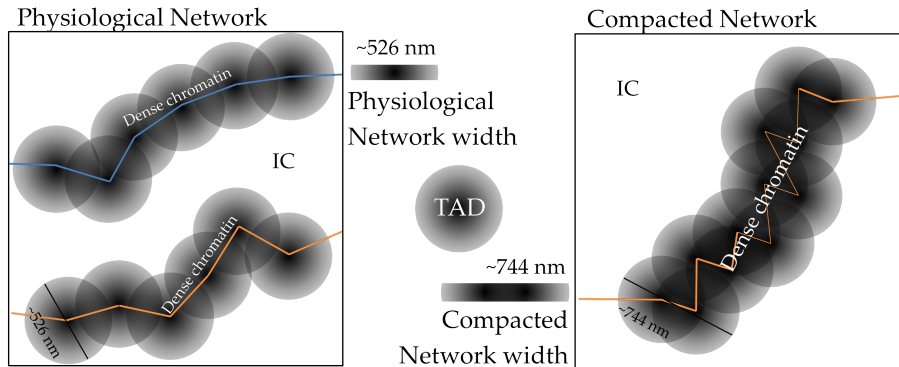


Figure 8.2: Perturbations of the TAD filament network model. Schematic of chromatin network as described in Figure 8.1 with added chromosome backbone highlighted in blue and orange for chromatin in trans. Exogenously induced compaction of the chromatin network is possible from the proposed TAD filament network by creating overlaps of TADs in orthogonal directions. This could be from tighter packing of genomically neighbouring TADs, highlighted by the change in the orange backbone. This creates a new, wider, filament network of 744 nm (right, to scale) at this degree of overlap. The exogenous methods of chromatin compaction used in this investigation do not directly affect cohesin or CTCF so presumably individual TAD architecture is preserved. Therefore single TADs in the network filament fibre are not themselves compacted but multiple TADs are brought closer together. Other chromatin perturbations such as overloading cohesin should be expected to cause similar change and affect the whole network as a consequence, since TADs are the base unit of the filament. Indeed this is the observed result from the knock-out of the cohesin removing factor Wapl whereby interphase chromatin forms thick network bundles which look like small worms “*vermicelli*” (Tedeschi et al., 2013). It is worth noting in the condensed filament the surface area to volume ratio is decreased as observed and the volume of denser chromatin hidden inaccessibly inside the network is increased which would lead to the depletion of markers in more chromatin. This is exactly what the ChaiN analysis has shown in Figure 6.9.

the periphery of these TADs which would have held low density are now occluded from access to action by the nuclear machinery due to being flanked on either side by whole filaments.

Decreasing chromatin compaction by the action of TSA, hypo-osmolarity or even in ESC is not likely to function as a direct reversal of the mechanisms induced by compaction. In the G_1 state, individual TADs are already shaped to the width of the chromatin network filaments. When imaging cells with decondensed chromatin, a network of finer filaments is not observed despite SIM being able to resolve down to 120 nm in x and y. As compaction is hypothesised to occur from bringing single TAD filaments together decompaction cannot pull single TAD filaments apart (unless denaturing DNA to two single strands). TADs represent a unit of contiguous chromatin in *cis*. Thus, decompaction from G_1 state most likely occurs from a breakdown of the TAD substructure into a more homogenous and accessible state. This would decrease the level of high contrast structure which would cause poorer SIM reconstruction, a fact that is observed when attempting to image these datasets.

8.3 The 3D epigenome

One of the key questions that immediately emerges from any proposed model of chromatin higher order organisation is the way in which this model affects, or is affected by, the functions of the genome. Overall this study has shown that a striking distribution bias can be observed for multiple histone markers at the global scale. Primarily this is relative to the underlying chromatin density where the marker is found, but to some extent the absolute distance from the network surface. Although most histone modifications demonstrated a statistical significant preference to a chromatin density class, the magnitude of this bias changed dramatically between epigenetic marks and chromatin states. Yet regardless of the individual differences in magnitude, grouping several classes by *a priori* known function yields a striking organisation. In-

terestingly a general correlation was found between the molecular size of the complexes involved in the deposition of the epigenetic markers analysed and their enrichment with respect to chromatin density (Figure 4.6). The relationship being that with increased size of complex, lower enrichment (or greater depletion) was found with higher density chromatin.

During G_1 the chromatin network seems to organise as follows. Markers of active transcription have high affinity for lower density chromatin and are strongly biased against deposition at density classes beyond classes 2-3 (Figure 8.3, left, green). Their density distribution profile is mirrored closely by all markers of chromatin architecture, adding further evidence to suggest that lower third of chromatin density (two classes out of the 6 total chromatin classes) makes up the bulk of the active genome region. Given that these markers are also predominantly at the network periphery it makes logical sense in 3D space. The non-chromatin IC makes up the highways of transport to and from nuclear pores. RNA and nuclear enzymes may use these channels to manoeuvre unimpeded, and Occam's razor dictates that the chromatin with preferential interaction to these processes would also be the closest to where they are found.

This however does not imply that just being a surface boundary is enough to cause local decompaction of chromatin. Although some Brownian motion is expected to affect chromatin once encountering a chromatin void region it is worth remembering that the IC is not devoid of all steric hindrances and other molecular moieties. In fact, given the hypothesised challenge that chromatin poses to molecular accessibility, the IC may very well have a high local concentration of non-chromatin entities, forced there by the 3D barrier of the chromatin network.

A likely explanation for the concurrence of low density chromatin at the network periphery is also the enrichment of markers signalling active transcription. Transcription requires the local remodelling of chromatin where histones immediately surrounding the transcription fork are displaced by the

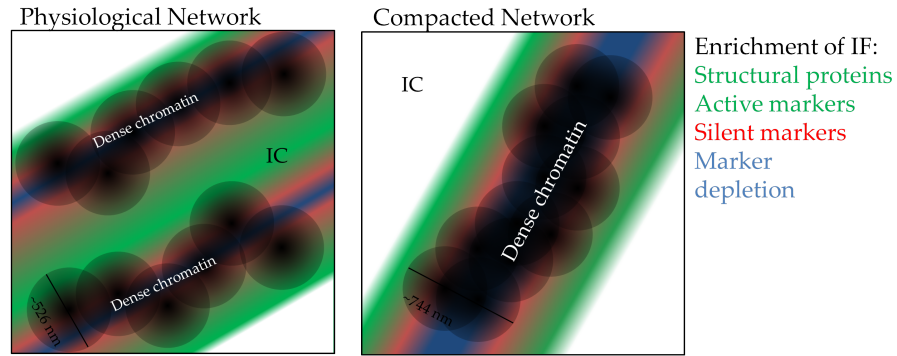


Figure 8.3: **Epigenetic functions on the TAD filament network model.** Colour coded schematic of the regions of distinct epigenetic functions with relation to chromatin density as shown by ChaiN analysis. The exogenous compaction of chromatin as suggested in Figure 8.2 is compatible with observation of the changes in epigenetic marker enrichments and depletions Figure 6.9.

action of chromatin remodellers. Immediately behind the fork, chromatin is also modified to reflect the transcriptional event, with certain markers being more prominent in the bodies of active genes (eg H3K36me2/3).

The fact that markers of chromatin architecture regulation are in such high correlation to markers of transcriptional activity in lower density chromatin is very interesting. A previous investigation on the correlation between TAD boundary and architectural elements found also a very strong correlation in these genomic regions to many transcription factors (Dixon et al., 2012). Some previous interpretation of these results suggested that the fact that architectural proteins bound to these regions, created non-TAD decondensed areas which were more susceptible and accessible to binding by transcription factors. However, together with our results, an alternative interpretation gains backing. In a network of large filaments of TAD-size, chromatin at the interphase of the network to the IC can be read and remodelled by chromatin readers/ remodellers. This leads to active transcription at these sites, favouring the density equilibrium to less dense chromatin. Proteins involved with TAD architecture are also found in the IC and bind to sequences at the lower density surface, acting as stabilising surfactants of the network surface. Their

action may controls the lower density chromatin from becoming too unwound to the point of revealing or making accessible chromatin that should not be accessed by such processes (Figure 8.4).

Some epigenetic marks that don't have a clear cut distribution bias were also observed. These are linked to multiple diverse genomic functions which may take place at different chromatin densities. Furthermore it has been shown that some of these histone modifications may play a role in epigenetic memory and the establishment of poised environments for activation after major chromatin remodelling in a cell cycle dependent manner. However without the ability to label these markers in a way that also allows the segmentation of their diverse functions it is not possible to say with certainty what fraction of these markers is involved with what function at each chromatin density. From the two extremes of unambiguous active and inactive markers it can be assumed that when these markers of multiple functions are found in less dense regions their functions may be due to genome activity, whilst their population in higher density chromatin classes could correlate to epigenetic bookmarking of repressed chromatin poised for activation.

Markers of repressed genome functions have a right-shifted density profile compared to active markers. That is to say that, for the most part, these markers are found in the denser two-thirds of the chromatin density classes (classes 4-7)(Figure 8.3, red). Unlike the architectural markers and markers of active transcription, there is far more variability in the profiles of inactivity markers during G₁. The heterochromatin marker, aHP1, and its associated histone modification H3K9me₃ were found to have a strong linear relationship between enrichment and chromatin density. These markers are strong for genome inactivity because they also correlate strongly to genome regions poor in genes. On the other hand, H3K27me₃, while also being a strong and well studied marker of heterochromatin, has strong correlation to gene promoter (silent genes). Therefore, its mid-range density profile is indicative that although transcriptional activity occurs in the lower third of chromatin

density, inactive genes are mostly present in the densities immediately and slightly higher, and not buried deep in the higher density core of the network (Figure 8.4(a), top middle red, right). Non-genic regions are enriched in the highest density chromatin, which suggest an architectural function for many “junk” DNA and pseudo genes which make up a significant portion of the genome with no known function. Under this model, these regions are kept despite their replicative burden on cells as chromatin that can be densely packed to make up a stable core of the network. As these regions do not have the potential for transcription, markers of transcription silencing (such as H3K27me3) are not required (Figure 8.4(a), top middle blue, right). Having the lowest frequency of interaction with transcriptional machinery also makes these chromatin regions recurrently packed as dense chromatin, and most likely include what is known as constitutive heterochromatin.

Work using STORM imaging on active, inactive or repressed (tissue specific) domains has also shown that chromatin of different function and epigenetic marker composition can be packed at different densities (Boettiger et al., 2016). They showed active domains to be the least densely packed, but with high locus-specific variation in packing which may be linked to the specific histone modifications and interactions at these loci. The least dense of all being the active domains with highest CTCF interaction. An observation which matches the distribution of CTCF by *ChaiN* in this investigation. In inactive domains (characterised by H3K27 methylation) higher levels of compaction were found, as much as would be expected from the fractal globule model of TAD formation (Mirny, 2011; Fudenberg et al., 2016). The repressed tissue specific domains showed the highest level of compaction, more so than expected from the fractal globule model. In *Drosophila* cells polycomb proteins were found to be enriched in these domains and shown to be required for their increased compaction. Ultimately they suggest that the epigenome helps delineate 3D environment which may work in concert with biochemical processes in the nucleus to regulate genomic function. In the latter, repressed

tissue specific regions, the addition of polycomb to compact chromatin even further may impose very low accessibility to chromatin to ensure aberrant transcription of developmental genes is not feasible. They suggest this might also explain the clustering observed for many developmental loci (eg: Hox genes), ie: these may all be packaged efficiently into a single repressed domain once cells are differentiated. Furthermore repressed chromatin domains strongly exclude active chromatin with very little mixing between them despite some being relatively close in the genome. This may avoid bystander activation from the histone modifications of active chromatin domains in trans.

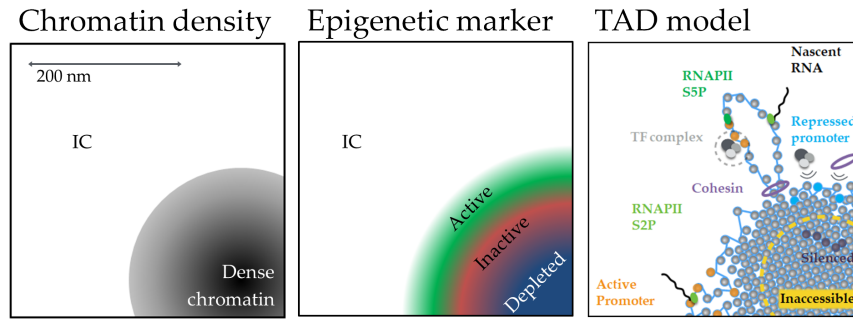
Epigenome rearrangements

Support for the network model with decondensed surface area also comes from the analysed changes in the epigenome from perturbation of chromatin. As discussed in each relevant chapter, chromatin can be perturbed by both endogenous and exogenous mechanisms. The most dramatic examples of endogenous chromatin remodelling occur at two different stages of the cellular cycle: S phase and mitosis, in order of increasing cellular rearrangement. Whilst the mitotic breakdown of the nuclear envelope and the condensation of chromosomes and sister chromatids looks dramatic under the microscope, genomic processes must be halted making this stage less interesting for the study of genome function. Thus in this thesis, the actions of the replication machinery during the 3 main stages of S phase (early, mid and late) were analysed with ChaiN . As S phase limits its remodelling around the active replication forks, exogenous global chromatin remodelling was also employed. This was achieved by different chemical or osmolar changes in the growth medium. For decompaction hypo-osmolar conditions or TSA were used, however this effect was found to yield poor reconstructions from SIM imaging. Thus the focus shifted to chromatin compaction. In the C127 cell line this was achieved by hyper-osmolarity or the addition of sodium azide (NaN_3).

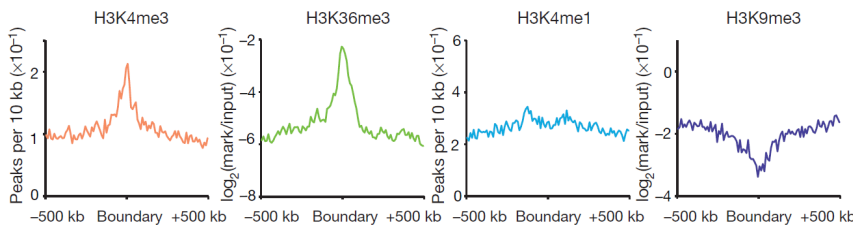
During ES, lower chromatin density classes (1-4) were disrupted. This was

manifest as a higher impact on the markers and functions which in G_1 had their enrichment peaks there, i.e: the active markers. The markers of inactivity have for the most part an unperturbed right skewed distribution. In MS the pattern of change has a more uniform effect on all markers; many of them exhibiting a prominent enrichment in class 1 and 2. During LS the highest density regions of the genome were perturbed and so mostly the markers of inactive chromatin had a decrease in their density profile enrichment at high density classes. These changes to the epigenome are what would be expected from the model where low density chromatin harbours active transcription markers but it is also the most accessible to the replication machinery and so it becomes affected by replication first, in ES. This rationale means that the last part of chromatin to be affected by replication, LS, should be the dense core of the chromatin network, and thus the markers here should only be affected at this stage.

For chromatin compaction, the network proposed collapses so that the average filament width is no longer that of single TADs. For this model, increased volumes of denser chromatin relative to lower density chromatin means a higher level of marker depletion in the denser classes for all markers and larger marker to surface distances for markers of inactive chromatin (Figure 8.3, right), the latter caused by larger distances on average to the nearest chromatin-IC interphase surface. As the levels of compaction increase, this becomes more pronounced. While one might expect that the enrichment of markers at lower density classes might remain unchanged, it is worth noting that with a reduction in the available surface area, nuclear machinery concentrate the deposition of active markers on the smaller remaining volume of lower density chromatin which they are able to access. This leads to an enrichment in the normalised distributions for the active markers in class 2 and 3. Unlike the structural markers where this mechanism leads to an increase in class 1 almost as prominent as classes 2 and 3, the active markers are mostly histone modifications and so are by definition still part of chromatin and so



(a) Density-function characteristics of TAD model



(b) TAD alignment of Histone PTMS

Figure 8.4: TAD functional model. (a) The relationship between chromatin density (left) and epigenetic function (centre) can be explained in single cells by focusing on the basic subunit of the proposed filament model, the TAD (right). Whereby the action of ATP dependent processes at the lower density TAD-IC boundary keep this region decompacted. Genomic loci (dark purple) beyond a certain threshold of density (dashed yellow circle) are excluded from interaction with nuclear machinery by the steric hindrance imposed from nucleosome crowding. In the event that, through Brownian motion, one such locus becomes exposed at the TAD-IC boundary its histones are modified with gene specific repressive markers followed by the non-genic markers at slightly denser chromatin. This explains the similar density distribution between active and repressive promoter markers (eg: H3K4me3 and H3K27me3 respectively). The same accessibility issue also predicts the observed depletion of all markers at very dense chromatin. Both due to the inaccessibility of the histone remodeller protein (proportional to the size of each remodeller) and because no marker (even for repression) is required where there is no chance of interaction with nuclear machinery. The model proposed here is also backed by recent alignments from genomics techniques of histone PTMs and TAD boundaries (b). Adapted from Dixon et al. (2012).

cannot be enriched in class 1. This is what can be observed for NaN_3 and hyper-osmotic conditions for both structural and active markers respectively.

These conclusions are not drawn from a single metric of density. Although distances measurements even with SIM resolution were inconclusive from

marker to marker in G_1 , it was clear that a pattern was visible when grouping markers by function. The problem lies in the fact that the average filament of the chromatin network during G_1 is at the limit of SIM resolution (§4.2). However, by artificially enlarging the filament widths from chromatin compaction experiments, these hypothesised functional zones become even more clear. The difference between these markers grouped by function become larger, and though not fully without overlap of the CI95 the trends are hard to ignore (Figure 6.10).

The subtle differences between the enrichment of structural markers and active markers between the IC and class 2 become clearer when in the context of distances. As described, only the structural markers being by definition not an integral part of chromatin should respond to a decreased surface area by accumulation in the IC and low density chromatin. Histone markers could only really be found enriched more in low density chromatin as a response to the same compaction. Distance measurements of the group of each of these functions backs this up (Figure 6.10). The active histone markers do show a small shift in their distance distribution to chromatin closer to the surface, but it is the structural markers that all display a large shift in the same direction, with the mean distance of all the structural markers now lying in the IC.

The RNAPII-S2P epitope shows the strongest dense-chromatin depletion from compacted chromatin (followed by the strongest markers for active promoters, H3K4me2/3). Its enrichment in the IC and density classes 2 and 3 increases with increased compaction and filament widening. Yet it is clear from distance measurements that its distance distribution shows the least amount of variation of any of the markers analysed, both in in/out shift or widening/narrowing of the CI95 (Figure 6.10). This is a strong indication that, as predicted by this model, the interphase between the IC and chromatin (regardless of the compaction of the chromatin network beyond the surface) is the main site of transcriptional elongation (Figure 8.4, top right).

One final surprising result, which may have mechanistic implications is

the reciprocal interplay between aHP1 and H3K9me3. Analysing the position of these markers in G₁ cells leads to an unsurprising strong correlation of increased enrichment with increases in chromatin density. This is true both because these are strong markers of heterochromatin, and also because the presence of H3K9me3 has been linked to the prior occupancy of aHP1 in chromatin (Lehnertz et al., 2003). Thus, it is strange to see a major decoupling between these two markers after the onset of compaction. The distance measurements reveals for aHP1 hardly any change in the maximum depth achieved by this marker into the chromatin network (a very subtle increase in the distance away into the IC progresses with increased compaction which pulls the average distance closer to the surface). Given that there is a decrease in enrichment of aHP1 in the dense chromatin classes, but the distance measurements shows almost no change it can be assumed that the same depths for compacted chromatin do not reach the same densities.

All these changes are still small compared to the large shift in H3K9me3 distance deeper to the network interior for both types and magnitudes of chromatin compaction. Concurrently, the H3K9me3 density profile becomes significantly enriched in classes 3-7 and depleted in classes 1 and 2.

With increased width of the chromatin network it appears that aHP1 cannot penetrate further than its usual range, which keeps it away from denser chromatin at the core of the network. H3K9me3 remains in dense chromatin which is now further away from the network surface, leading to a deeper distance measurement. The increased enrichment of H3K9me3 in the denser classes implies little hindrance of marker deposition, pointing to the smallest of the tri-methyl methyltransferase complexes SUV39H1/2 (KMT1A/KMT1B; 48 kDa, 50 kDa) as the most likely complex. With a molecular weight smaller than the antibodies used for detection, the deposition of H3K9me3 at the densest and deepest chromatin can still be expected from the proposed accessibility model. What genome regions and what function is undertaken by H3K9me3 present in an aHP1-independent manner are still to be investigated, although

from the accessibility model the predicted nature of chromatin at this depth would most likely be non-genic repetitive tandem sequences.

8.4 Concluding remarks

The development of the ChaiN workflow and the results obtained from it in this thesis provide strong evidence for an intricate organisation of the epigenome in single cells with regards to a filamentous network of chromatin composed of TADs (Figure 8.1). The TAD network interior is composed of denser chromatin and its exterior is comprised of a surface of less dense chromatin (Figure 8.2) that acts as the interphase between the large chromatin molecule and the enzymes and nuclear machineries that mostly inhabit the channels of the IC. This lower density chromatin is therefore the functional genome region (Figure 8.3) with a variety of genome activities, such as transcription, relying on protein-chromatin non covalent interactions and ATP driven chromatin remodelling (Figure 8.4). Taken together, similar studies of chromatin architecture using electron microscopy (see §8.2, Ou et al. (2017)) and of the epigenetic relationship of chromatin packing by STORM (see §8.3, Boettiger et al. (2016)) are in line with results from the ChaiN analysis. The advances on these and the major implication of the results from this work is the concept of a TAD-based filament network (Figure 8.1(b)) which also provides a framework for understanding the distribution of epigenetic markers at this scale (Figure 8.3) while still keeping the known TAD-based distribution of the same markers (Figure 8.4).

Although this study focused primarily on wild type cells under physiological and non-physiological conditions, the development of ChaiN also has far-reaching applications. The chromatin of retina cells of nocturnal mammals has been shown to have a macro-morphology inversed from other somatic cells, with high-density chromatin at the nuclear interior and less dense chromatin at the periphery (Solovei et al., 2009). It is hypothesised that this acts

as a lens to focus light at photoreceptors in low-light conditions. How the 3D epigenome is arranged in these cells would shed light on the robustness of the system able to tolerate such structural inversions and maintain genome regulation. Furthermore, many diseases are characterised by abnormal nuclear or nuclear lamina morphology, eg: progeria. These laminopathies are hypothesised to directly interfere with genome architecture and thus deregulate downstream function in somatic and embryonic cells (Perovanovic et al., 2016). By employing ChaiN to characterise the 3D epigenome rearrangements in these pathologies we may elucidate the epigenetic link between physical perturbations of the nuclear structure and genetic output. As noted previously, novel studies employing ChaiN with large nuclear morphology alterations should incorporate a dedicated nuclear lamina stain to bypass current limitations whereby the nuclear volume is estimated solely from chromatin staining. This is particularly true for laminopathies where the quantification of interactions from chromatin at the periphery is of highest importance.

Lastly, as the scientific community moves towards higher automation, throughput and machine learning, the core analysis performed by ChaiN could in future form one of many characterisation readouts in a much larger epigenetic fingerprinting workflow. Even now, efforts are being made to collate data at the single cell level from super resolution STED imaging to quantify the total levels of certain epigenetic marks and correlate these to the transition stages of stem cells before arriving at a particular lineage (Kim et al., 2017). The unique combination of markers deposited at different quantities acts as an identifier for that particular cell stage. A higher level of fingerprinting precision could be gained by incorporating measurements from ChaiN which can also show the subtle changes in marker distribution to chromatin density and not just the total marker deposition.

Future studies will also need to employ equally powerful automatic image analysis pipelines in conjunction with super resolution imaging plus staining of single TADs, multiple neighbouring TADs, and sub-TAD regions along

with IF stainings of epigenetic markers. As shown in this study, such combination of state-of-the-art imaging, chemical staining and high-throughput processing are the minimum requirements to begin to elucidate the structure-function duality of the genome in single cells.

Appendix

9.1 Controls

The primary focus of this investigation was the screening of spatial relationships between IF labelled histone modifications, phosphorylated RNA polymerase II (RNAPII) variants and key TAD architectural proteins, (all linked to diverse genomic functions) to chromatin density in the entire nuclear volume as characterised by an unbiased global SYTOX counterstain. When passed through the ChaiN workflow, these 3D SIM datasets allowed the spatial organisation of the functional epigenome to be overlaid onto the 3D global chromatin landscape. In order to make such comparisons several assumptions must be made, the validity of which were tested.

3D SIM is capable of resolving sufficient structure in both the genome and immunofluorescence for segmentation

The use of super resolution microscopy in this investigation is predicated on the shortcomings of diffusion-limited widefield microscopy for the resolution of chromatin. Under optimal conditions, widefield images are able to resolve chromatin beyond a single amorphous stain, hinting at a more intricate distribution clearly revealed by SIM (Figure 9.1(a)). It should then be asked whether the mapping of epigenetic markers may be achieved under non-super-resolved conditions.

To assay whether super resolution is indeed required for the characterisation of the 3D epigenome, the ChaiN output for the distribution of RNAP-S2P and the proportion of chromatin under each density class were compared in either super resolution or pseudo-widefield (PWF, Figure 9.1). PWF can be achieved by the ImageJ plugin SIMcheck (Ball et al., 2015). Furthermore as both the chromatin and IF channels are affected by resolution limits, I have generated a hybrid image with super resolved IF spots but with widefield chromatin (PWF-SIR). The results from this analysis shows that for the RNAP-S2P (the most strongly correlated marker for transcriptional activity) only

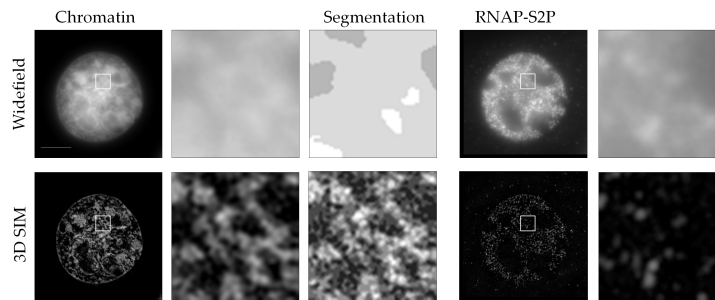
fully super-resolved imaging yields a distribution by which it is enriched in the low density chromatin and consistently under-represented in higher density chromatin in agreement with ChIP-Seq data of RNAP-S2P. PWF images have the distribution of RNAP-S2P enrich as the class density increases with a massive depletion in class 1, and strong depletion in class 2 (Figure 9.1(c)). It could be argued that as distribution is measured as a fold change normalised to class size, then a high depletion in class 1 may be due to an increase in the volume of class 1 segmented in both PWF and PWF-SIR. However, the relative class volume for each of these is in fact lower in both class 1 and 2 and higher in classes 4-7 than the values observed for SIM (Figure 9.1(b)). This shows that without super-resolution microscopy, the fine detail of the IC channels and low-density chromatin throughout the nucleus is lost to the overwhelming out-of-focus blur from diffusion-limited widefield imaging (Figure 9.1(a)). Thus, super-resolution microscopy is essential for the accurate characterisation of the *in situ* 3D epigenome.

Of the super resolution techniques SIM is nominally the one with the least resolvable power at a 2 fold increase in resolution in axial and lateral dimensions, constituting an 8 fold increase in volumetric resolution compared to diffusion-limited microscopy. This amounts to a resolution of 120 nm in xy and 300 nm in z. In comparison, other techniques such as 3D STED or PALM/STORM(in TIRF) boast of 60 nm to 30 nm in xy and 200 nm to 100 nm in z respectively (Scheremelleh et al., 2010).

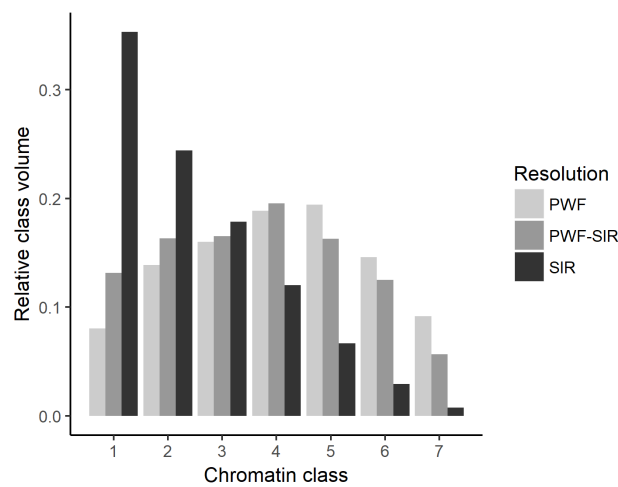
However the advantages of SIM are: its speed, as well as the ability to use a palette of all standard dyes (as long as they are robust enough to photobleaching). This makes SIM amenable to experiments requiring the acquisition of many wavelengths, such as 3 or 4 channel images required in this investigation. Furthermore for immunofluorescence of epigenetic targets, where the image is comprised of distinct foci, the lower resolved elliptical point spread function can be mathematically modelled and its centre pin-pointed with a high degree of accuracy (§3.2). For stainings such as chromatin, where this method cannot be used SIM only has its nominal resolution. Yet, due to the high density and therefore out of focus blur of chromatin labels, neither of the more powerful techniques is able to resolve chromatin to a significantly higher degree than SIM (Figure 1.12). A case has been made for the detection of single nucleosome clutches via single molecule localisation (Ricci et al., 2015). However when taking into account various chromatin epitopes these individual clutches merge to give the same nebulous network pattern observed with both SIM and STED (Figure 1.12, a). SIM is clearly not at a significant disadvantage for fulfilling the technical requirements of this investigation and boast of greater practicality when considering experimental design of multiple wavelengths, and acquisition of thousands of samples needed for statistical power.

Chromatin stains used are linearly proportional to the density of the underlying chromatin

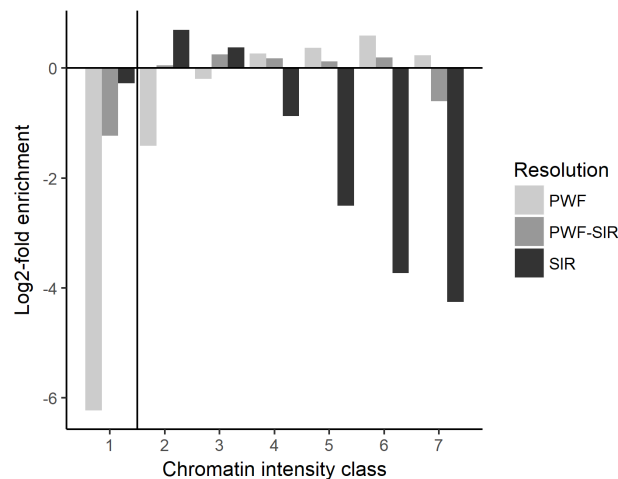
One potential assessment of the environment where diverse genome functions are found would be chromatin density. In the model proposed in this study, chromatin density plays a role in the accessibility of chromatin to machinery found in non-chromatin regions (the interchromatin compartment, IC). How-



(a) Resolution of widefield vs 3D-SIM



(b) Resolution in chromatin segmentation



(c) Resolution in RNAP-S2P distribution

Figure 9.1: Resolution controls. Control for the need of super-resolution microscopy for characterising the 3D epigenome. (a) Gallery comparing the same cell imaged as widefield or 3D-SIM micrographs for both the chromatin or IF detection of RNAP-S2P. Position of insets are outlined by a white box. For chromatin the loss of segmentation quality is highlighted. Scale bar = 5 μ m. (b) Proportion of chromatin density represented for each chromatin class pseudo widefield (PWF), structured illumination reconstruction (SIR) or a hybrid of PWF chromatin and SIR IF channels (PWF-SIR). (c) Log2 fold enrichment and depletion of IF marker abundance for RNAP-S2P relative to chance in each chromatin class in the different resolution micrographs.

ever, as SIM is based on fluorescence microscopy, direct values for chromatin density cannot be measured. Instead, the intensity of fluorescence from chromatin labelling dyes can be measured and a scale of relative density inferred from such values. This requires that the label is specific only to chromatin and is distributed evenly across chromatin without bias, such that the relationship between intensity scale and density scale be linear. Furthermore chromatin is DNA wound around nucleosomes, so it is important to link chromatin to the epigenome and not loose DNA.

Thankfully many dyes for DNA labelling exist with both strong specificity and almost 100% genome cover (given excess stoichiometry of label to nucleotide when incubating). DAPI (4'-Diamidino-2-phenylindole) is one of the most regularly used. Using GFP fused histones in a dual stain experiment with DAPI DNA stain there is a strong correlation to chromatin *in situ*. With an average Manders' Correlation Coefficient of $M1=0.8655$ and $M2=0.8700$ for the fraction of voxels of H2B overlapping DAPI and *vice versa* respectively. However DAPI labelling is not entirely unbiased, with a preference for AT rich sequences (Smeets et al., 2014). This would make the relationship non-linear. To avoid this confounding effect, SYTOX green dye was used. This dye has no base bias but does have a weak affinity to RNA, requiring a prior RNase incubation.

Density of chromatin does not hinder antibody detection

Since the final aim of this study is to compare the distribution of epigenetic markers as identified by immunofluorescence, it is important to make sure that the density of chromatin does not affect the penetration of IgG based antibodies. DNA density does not hinder introduction of fluorophores on the scale of 300 Da. Nevertheless IgG antibodies have a molecular weight several orders of magnitude larger at (150 kDa). Using ChaiN to map the distribution of an immunofluorescence stain against the ubiquitous H2B protein (which correlates to chromatin density) it is possible to control for antibody accessibility. Results from this analysis (Figure 9.2) show no significant accessibility hindrance across the chromatin density classes with a generally flat profile centered at zero log2 fold change.

Local chromatin density is reversibly disrupted by replication during S phase

One of the key assumptions in Chapter 5 is the temporal local perturbation of chromatin by the replication machinery. In the aforementioned chapter, the IF experiments were performed immediately after staining for replication as this would theoretically represents maximum disruption to the local chromatin environment. To check whether this is indeed the case, a 10 minute EdU incorporation was followed by a chase period before fixation and staining.

The expected results were observed at each stage of S phase, with the strongest decompaction of chromatin seen at 0 minutes after EdU incorporation, then reverting to more dense chromatin the longer the chase (Figure 9.3). That is, as the replication machinery moves away from these loci. For ES the peak shifts from class 2 to class 3 as the chase length continues (Figure 9.3(a)).

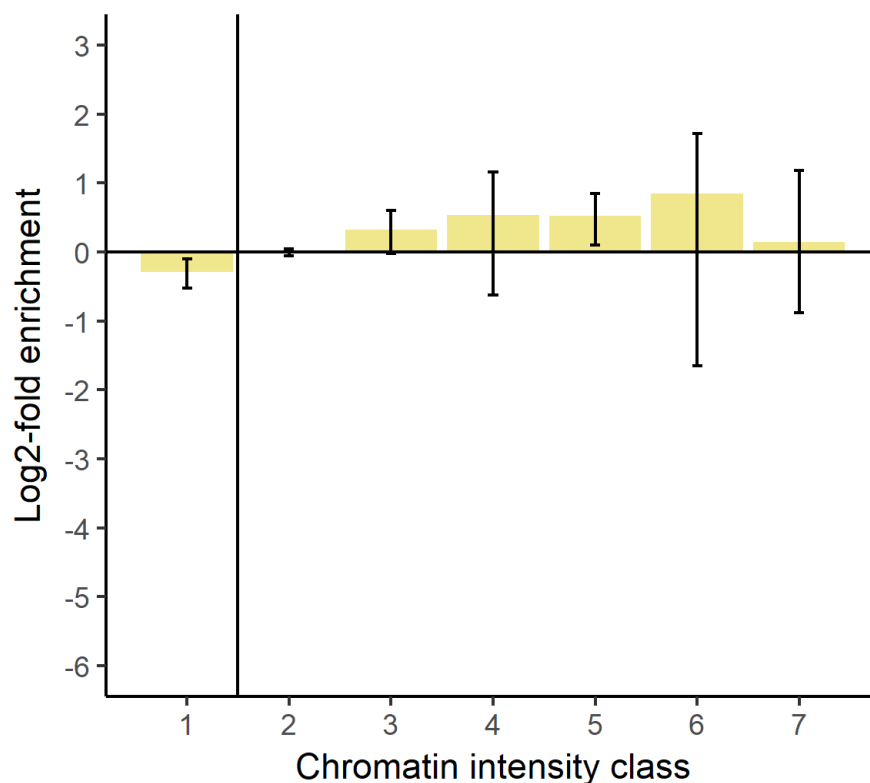


Figure 9.2: **Antibody penetration.** log2 fold enrichment and depletion of IF marker abundance for H2B relative to chance in each chromatin class. Error bars = 95% Confidence interval.

For MS the peak shifts from class 3 to class 4 as the chase length continues (Figure 9.3(b)). And for LS the peak shifts from class 3-4 to class 4-5 as the chase length continues (Figure 9.3(c)). This work was conducted by Miss P. Oldach as part of her post-graduate rotation research position in the laboratory of Dr L. Schermelleh, under my supervision.

Immunofluorescence is not compromised by exogenous chromatin rearrangements

One of the core assumptions in comparing the distribution of epigenetic markers as chromatin landscapes change in Chapter 6, is that the deposition of said markers or their IF detection should not be adversely affected by these changes. Taking the average spot number per cell for each marker type under each chromatin condition compared to G_1 reveals that for the most part this is the case (Table 9.1). CTCF shows an increase with from G_1 and hyper osmotic treatment, as does Smc3 between G_1 and both NaN_3 and Hyper osmolarity. Although the numbers of final coordinates assayed per cell is low compared to some ChIP approaches, it is worth noting that stringent cut-offs are in place to discard low quality data from poor reconstruction or labelling.

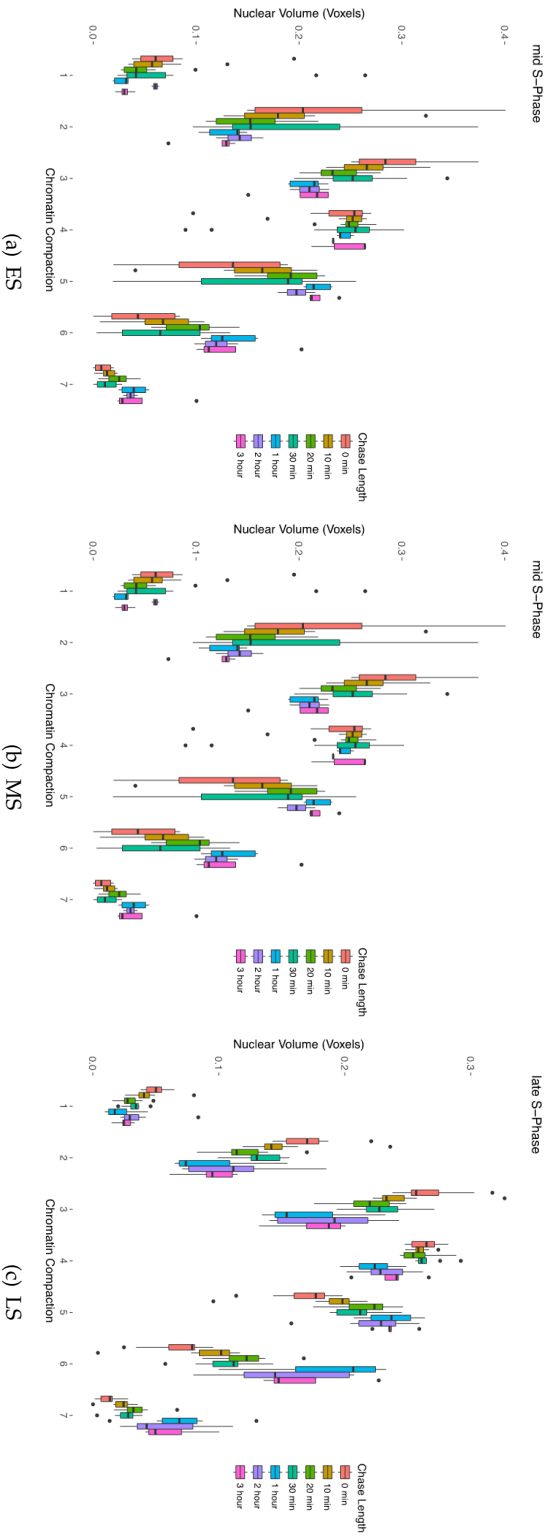


Figure 9.3: **Reversible S phase disruption to chromatin.** Volume of each chromatin class detected in the ROIs of ES (a), MS (b) and LS (c) EdU stain after a chase time from the EdU incorporation. $N_{total} = 57, 57$ and 61 cells respectively.

Marker	G ₁	NaN3	Hyper-os
aHP1	2462.4	3663.8	3706.55
CTCF	5396.1	4096.45	8032.15
H3K27me3	2426.4	2076.45	3692.65
H3K36me2	3259.75	2951.8	3391.55
H3K36me3	969.35	1334.2	825.9
H3K4me2	2707.7	4105.4	3030.75
H3K4me3	4900.65	5814.05	5591.7
H3K9me2	4949.9	4124.4	4658.25
H3K9me3	811.4	2725.6	1504.6
H4K20me1	2850.1	3699.9	2604.5
H4K5ac	5177.45	5228.1	4999.55
Macro-H2A	4406.75	6115.55	4143.05
RNAP-S2P	4360.5	6263	6160.6
SAF-A	4654.95	4031	3418.6
Scc1	1271.5	1271.5	1046.3
Smc3	614.7	3016.25	3138.25

Table 9.1: **Spot numbers.** Average spot numbers after MCNR cleanup for each marker under each exogenous chromatin perturbations used in this thesis compared to endogenous, G₁ numbers.

9.2 Statistical analysis

Sampling, replicates

To achieve statistical power for detecting significant differences in datasets it is often desirable to acquire as large a sample as possible. Measuring the *true* value the epigenetic distribution is not only highly impractical technically, but downright impossible, as the value for all cells not just on the coverslip, but in existence must be measured. The correct method is to calculate the minimum number of samples that must be gathered to have a large enough statistical power to detect significance. This can be achieved through estimation and the following relationship: $N = (\frac{\sigma Z}{\alpha \bar{x}})^2$ where N is the desired number of cells, α represents the statistical power (the probability of committing type II errors subtracted from unity, $1 - \beta$), Z is the z score from the normal distribution at the confidence required, ie the multiple of standard deviations that encompass your confidence (eg: 1.96 for 95%) and finally σ and $\bar{x}$ are taken as the best approximations of the true standard deviation and mean (μ) of the population. When using this formula the values of σ and $\bar{x}$ are those that must be estimated empirically from preliminary results and also dependent on the N of these first results. The more precise they are preliminary, the better the approximation. Here a preliminary N of 10 cells were used yielding a preferable sample size between 10-25 cells on average (dependent on the marker). A value of 20 cells was used, informed from this estimation, split into 2 technical replicates to avoid a technical bias.

Error calculations

In this investigation several parameters were extracted (see §3.3). For the density distribution analysis each cell produces an absolute value of spots found in each of the 7 segmented classes. So no sample error is introduced by averaging for each cell. However when pooling the results of measurements on multiple cells then the mean of these cells is generated. This incurs a standard error $SE = \pm \frac{\sigma}{\sqrt{N}}$, where σ is the standard deviation from the mean, and N is the number of cells. From these values the error bars used in the graphical representations are derived. Only 95% confidence intervals or standard errors are plotted and not just the standard deviation. The rationale behind this is that both confidence interval and standard error take into consideration N . Plotting σ as error bars cannot account for N and therefore does not give enough information to judge statistical significance from overlaps. It would only show the spread, ie: the variation of effects within that sample, and so was not used as a value for plotting error bars.

Confidence intervals and standard error estimate how far the mean $\bar{x}$ of this sample might be from the true mean of the population, μ . When using standard error as error bars, the bars may not overlap, yet that does not always mean the difference between the two datasets is significant. Alternatively, when plotting confidence intervals as error bars, the bars may overlap and the difference between the two datasets still be statistically significant. In this study 95% confidence intervals were used in preference to standard error, as this means that when two bars are not overlapped there must be significant difference. This decision is a conservative one, risking possibly missing statistically significant values in order not to call significance when none is there. Essentially, potentially increasing type I errors (rejecting true positives), to avoid type II errors (retaining false negatives).

When performing the distance measurement analysis, each cell gives a distribution of values, with its own mean, standard deviation and standard error. Here also a measurement error must be taken into account. In SIM with a 60x magnification used in this study, the pixel size accounts for 41 nm in both x and y and 125 nm in the z dimension. Using simple Pythagorean algebra the maximum distance d between the opposing vertices can be calculated as $d = \sqrt{x^2 + y^2 + z^2}$, 137.8 nm. Given that measurements from sub-voxel points to the nearest voxel surface are a half max voxel distance, $d/2$, this makes the maximum error E expected from each measurement at least 68.9 nm. The error of the mean in each cell would then be inversely proportional to n the number of spots segmented in that cell, $E_c = \sqrt{\frac{E^2}{n}}$, or 0.81 nm. Thus the error from the limitations imposed on the pixel size by SIM can become negligibly small due to the thousands of spots measured per cell.

Once the mean from each cell are pooled together, the errors must be propagated. Pooling each individual distance measurement for all spots regardless of cell and having one single mean for all is a statistical fallacy known as *pseudoreplication*. The general formula for propagation of errors used is:

$$E_f = \sqrt{\left(\left(\frac{\delta f}{\delta x_i}\right)\Delta x_i\right)^2}$$

Given that our function f is the mean of means:

$$\mu = \frac{\bar{x}_1 + \bar{x}_2 + \dots \bar{x}_N}{N}$$

Therefore:

$$\frac{\delta_f}{\delta_\mu} = \frac{1}{N}$$

So:

$$E_f = \sqrt{((\frac{1}{N} \Delta x_i) \Delta x_i)^2 + \dots}$$

Given Δx_i is SE_i

$$E_f = \sqrt{((\frac{1}{N} SE_i) \Delta x_i)^2 + \dots}$$

Where the best estimate of the true distribution will have a mean $\mu \pm E_f$.

For distance calculations after pooling this means that n drops from a few thousand spots per cell to N , just 20 cells sampled per marker. The total standard error for each marker SE_m becomes proportional to both the calculated error per cell E_c and the standard error of the mean SE from the cells samples:

$SE_m = \sqrt{E_c + \frac{SE^2}{N}}$. As shown for the density distributions, the measurement error per cell is often negligible given such a large n of spots, so a condensed formula $SE_m = \sqrt{\frac{SE^2}{N}}$ can be used. Thus to decrease the SE_m , the total number of cells sampled per marker must increase. Of all super resolution techniques SIM has the highest acquisition speed. However, this still has a limit at approximately 10 minutes per acquisition (with reconstruction) in this experimental set up. Just doubling N from 20 to 40 would increase the time required for human-based acquisition from 3 Hours 20 minutes to 6 hours 40 minutes per marker, or approximately 4 days, 5 hours for all markers. Increasing N by just one whole order of magnitude to 200 cells per marker takes the acquisition time to 2.5 days per marker, and by extension 40.5 days for all marker. Without automated acquisition, continuous acquisition beyond a few hours is the limit for human-based high-throughput quantitation.

9.3 Optical set up

The 3D SIM microscope used in this study was the OMX V3 from Oxford Micron advanced bioimaging unit. The optics drawer configuration for the acquisition of data from four wavelengths was kept as outlined in Figure 9.4. The objective used was a 60x magnification, 1.42 NA Oil immersion objective from Olympus.

9.4 Supplementary data

Nuclear volumes

Using the ChaiN algorithms the volume of the entire nucleus can be estimated from the chromatin stain. These volumes can be compared in cells exposed to different conditions (Figure 9.5).

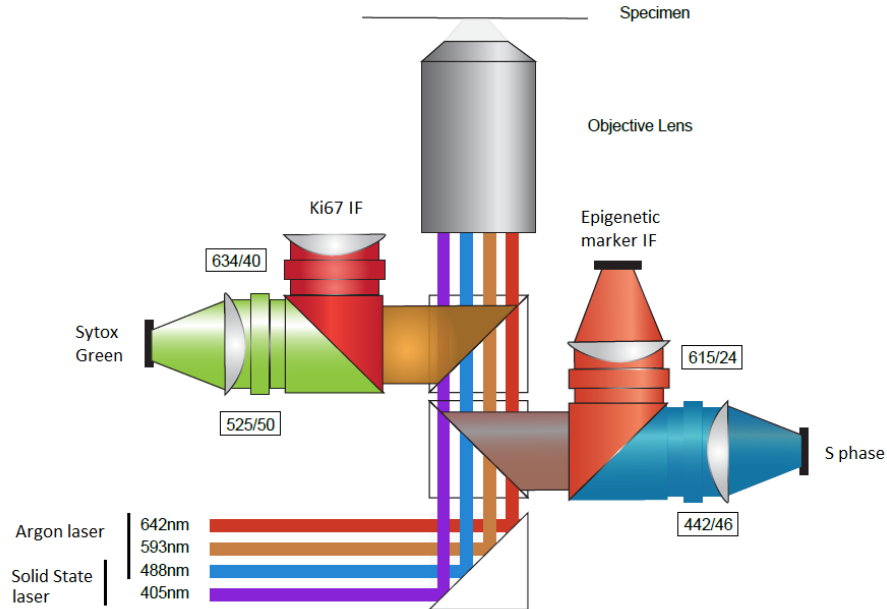


Figure 9.4: **Lightpath.** Schematic of lightpath from both laser sources for the acquisition of SIM data in four wavelengths. Figure adapted from Oxford Micron <https://micron.ox.ac.uk/microngroup/facilities/OMXV3.php>.

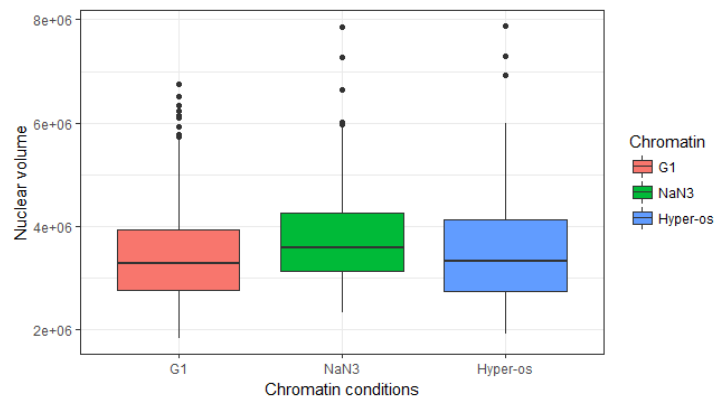


Figure 9.5: **Nuclear volumes.** Volumes of C127 nuclei in voxels comparing untreated G_1 and NaN_3 or Hyper-osmotic treated cells. One voxel represents $\sim 210 \text{ nm}^3$. $N_{\text{total}} = 960$ cells.

Single marker profiles

In this section are the individual density distribution profiles for each epigenetic marker in all conditions. These have been condensed into heatmaps for ease of comparison in the main text of this work however by doing so data is lost (Figure 5.3, Figure 6.9 and Figure 6.11). Particularly it is important to have a visual representation of the statistical comparisons by showing error bars of 95% confidence interval as displayed in this section.

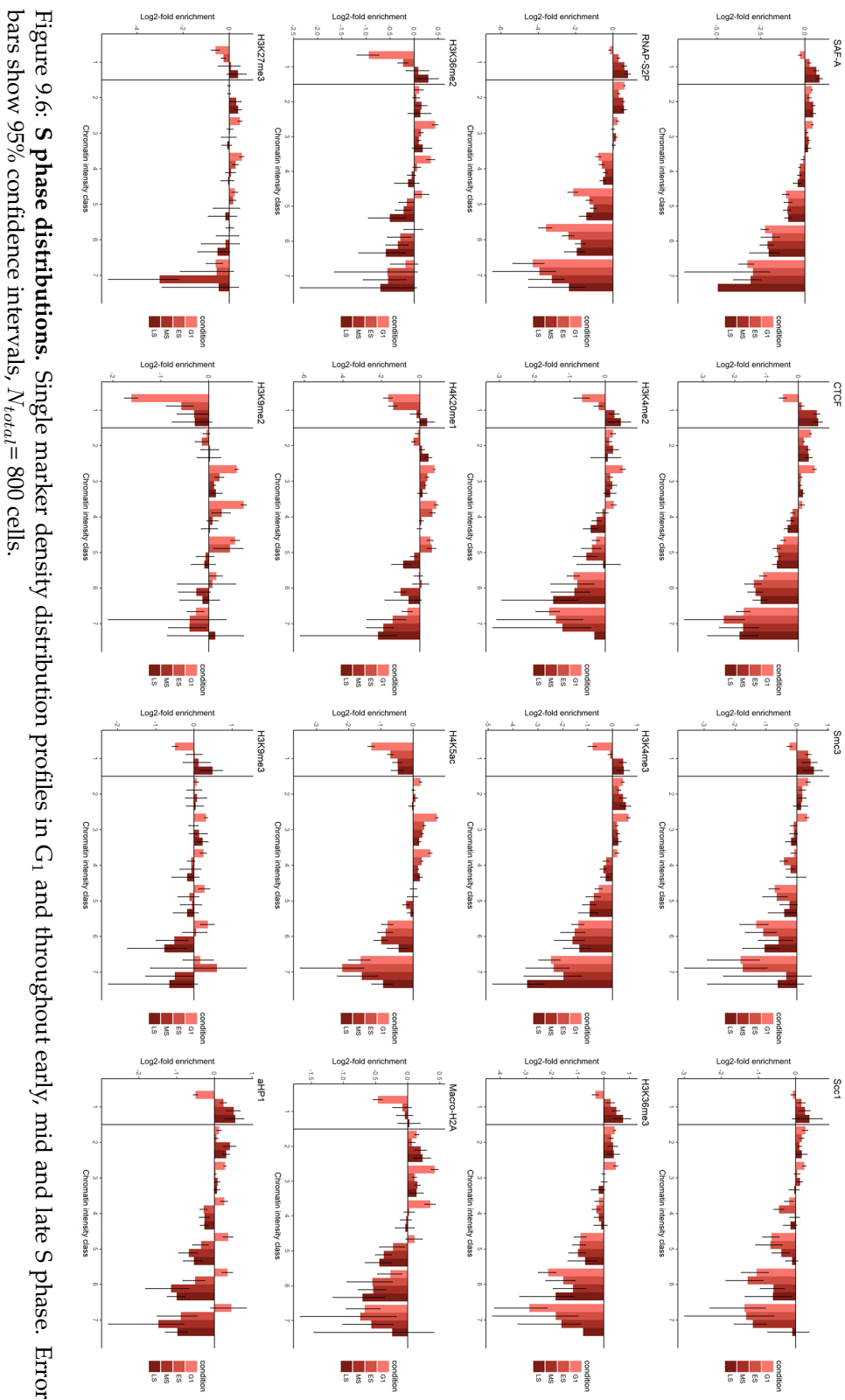


Figure 9.6: **S phase distributions.** Single marker density distribution profiles in G_1 and throughout early, mid and late S phase. Error bars show 95% confidence intervals, $N_{total} = 800$ cells.

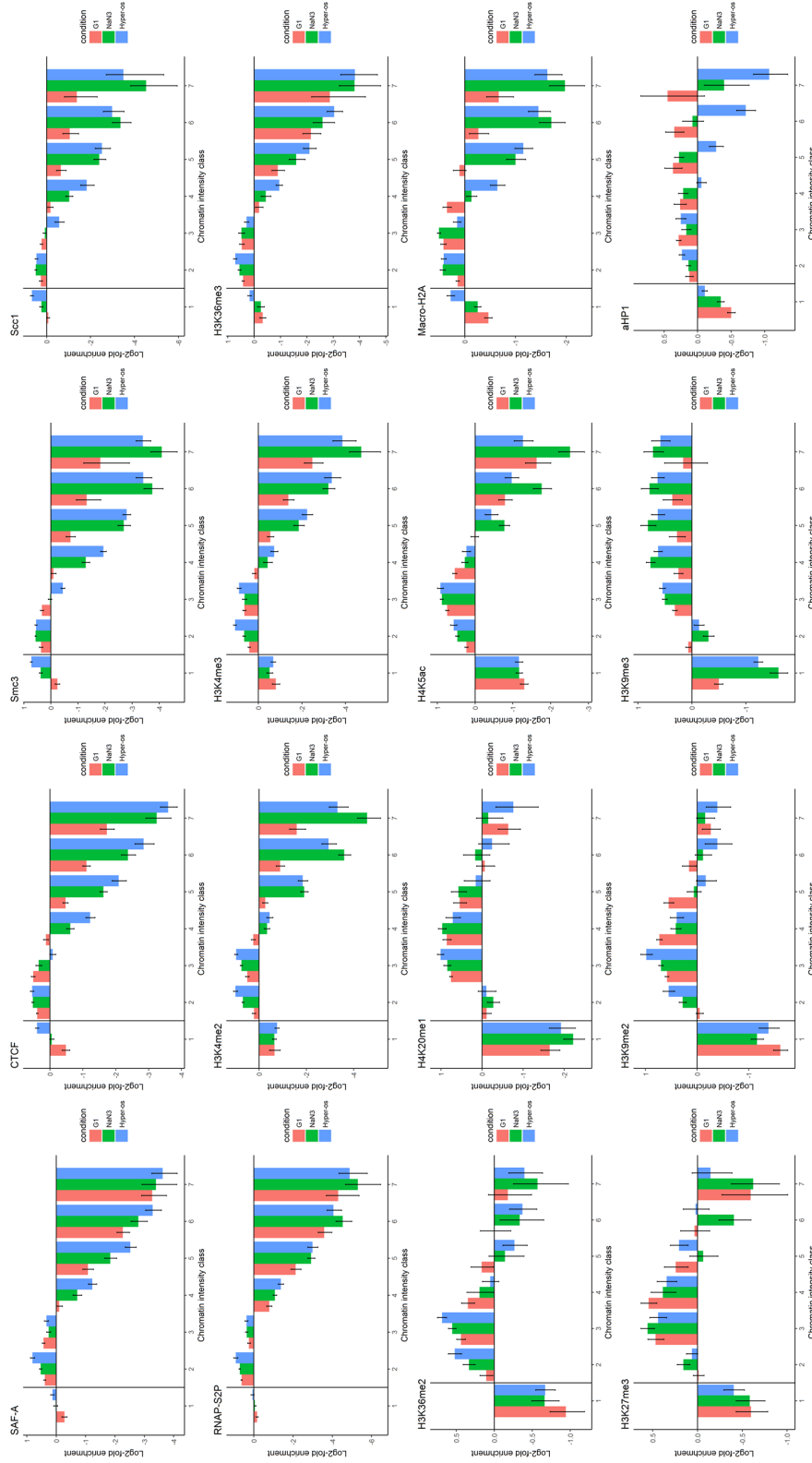


Figure 9.7: **Exogenous compaction distributions.** Single marker density distribution profiles in wild type C127 cells in G₁ and compaction by NaN₃ and hyper-osmolarity. Error bars show 95% confidence intervals, $N_{total} = 960$ cells.

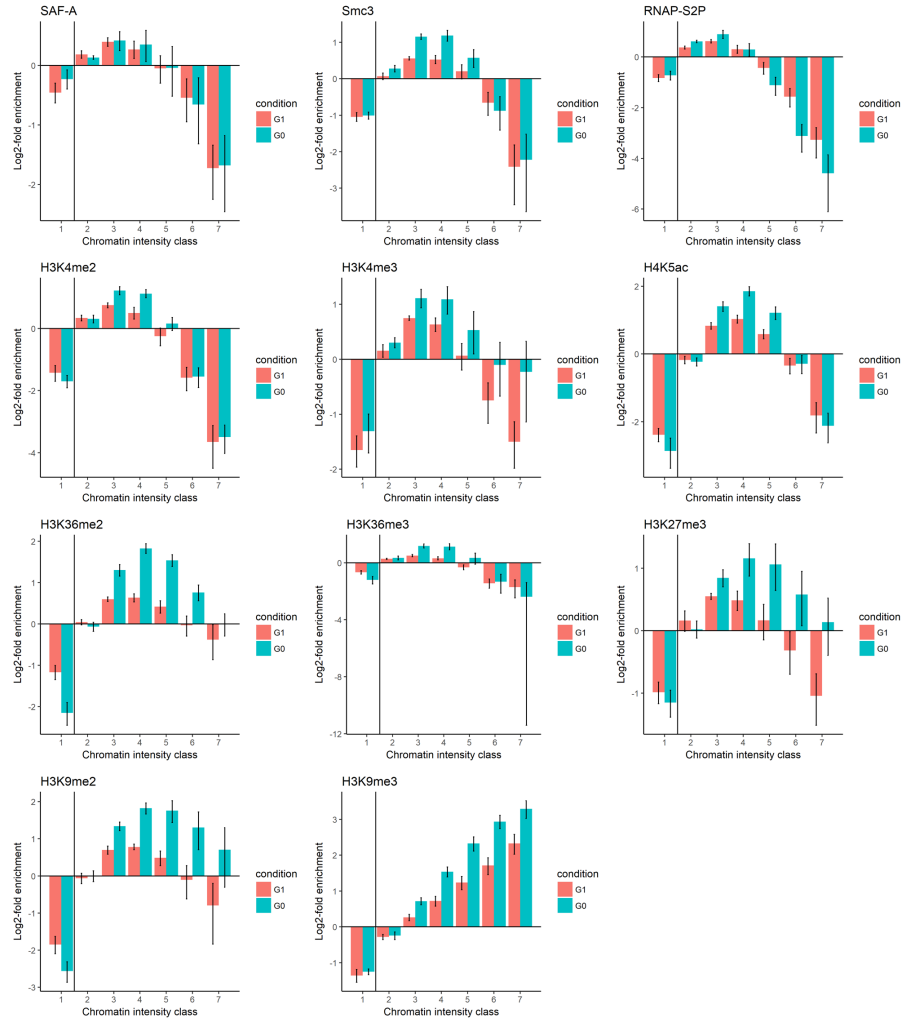


Figure 9.8: **Senescence distributions.** Single marker density distribution profiles in wild type IMR90 G₁ and senescence (G₀). Error bars show 95% confidence intervals, $N_{total} = 660$ cells.

9.5 Materials

This section outlines the reagents, solutions, equipment and antibodies (both primary and secondary) used in this investigation and their respective manufacturers and suppliers when known. Whenever possible, item/stock codes are also given for ease of identification and experimental replication. Where this is not noted, it is demarcated by “N/A”. Furthermore, where common laboratory equipment was used from various manufacturers this is demarcated by “multiple” as these are not brand/stock dependent.

Reagents

Table 9.2: **Reagents.** Table of reagents used throughout this thesis.

Item	Source
Competent <i>E. coli</i> bacteria DH5 α	Invitrogen, 18258-012
SOC solution	Invitrogen, 18258-012
Kanamycin	Sigma-Aldrich Co Ltd
Chloramphenicol	Sigma-Aldrich Co Ltd
Neomycin	C0378-5G
Puromycin	Sigma-Aldrich Co Ltd
DNA purification kit	Sigma-Aldrich Co Ltd P8833
Taq DNA Polymerase	Insight Biotechnology Limited
Reaction buffer for Taq polymerase	NP100024
dATP, dTTP, GTP, dCTP, 0.5 mM	Sigma-Aldrich Co Ltd
Nuclease-free H ₂ O	11647679001
Agarose, analytical grade	Sigma-Aldrich Co Ltd
Ethidium bromide (10 mg/ml)	Sigma-Aldrich Co Ltd
DNA loading dye	New England Biolabs
LB media	N0446S
Culture media (DMEM)	Thermo Fisher Scientific
Culture media (DMEM F-12)	750023
Culture media (Glutamax)	Sigma-Aldrich Co Ltd
Trypsin TrypLE Express Enzyme (1X)	Sigma-Aldrich Co Ltd
DMSO	Sigma-Aldrich Co Ltd
Hydrochloric acid	Sigma-Aldrich Co Ltd
Thymidine	Sigma-Aldrich Co Ltd
Lipofectamine, 2000	Invitrogen
Lipofectamine, RNAiMAX	Invitrogen
Fuse-It	Ibidi
TransIT-X2	Mirus
Triton X-100	Sigma-Aldrich Co Ltd
MEGashortscript T7 Kit	Ambion, Life Technologies
	AM1354

Maxblock	Active Motif, 15252
DAPI	Sigma-Aldrich Co Ltd D9542-5MG
SYTOX (green)	Life Technologies, S7020
Glycerol	Sigma-Aldrich Co Ltd
Vectashield	Vector Laboratories, H-1000
Prolong-gold/diamond	Thermo Fisher Scientific
RNAseI	Life Technologies, EN0601
Chloroform	Sigma-Aldrich Co Ltd
EdU	Life Technologies, A10044
F-ara-EdU	Sigma-Aldrich Co Ltd T511293-5MG
EU	Life Technologies, E-10345
Sodium ascorbate	Sigma-Aldrich Co Ltd
Copper sulfate	Sigma-Aldrich Co Ltd
CF405M azide	Cambridge Bioscience Ltd BT92092
Alexa Fluor 594 azide	Life Technologies, A10270
Alexa Fluor 488 azide	Life Technologies, A10266
4D Nucleofection kit	Lonza
Doxycycline	Sigma-Aldrich Co Ltd PHR1145-1G
Shield 1	Takara Clontech, 632189
β MerCaptoethanol 100 mM	Life Technologies, 21985023
DNAse (2000 U/ml)	New England Biolabs, M0303S
DNA polymerase I (10 000 U/ml)	New England Biolabs, M0209S
MgCl ₂ , 50 mM	Sigma-Aldrich Co Ltd
Cot1 DNA (m/hCot)	Thermo Fisher Scientific 18440-016/15279-011
Salmon sperm	Sigma-Aldrich Co Ltd, D-7656
Sodium acetate	Sigma-Aldrich Co Ltd
BrdU	Sigma-Aldrich Co Ltd B7166-5MG
BrdC	Sigma-Aldrich Co Ltd
Ammonium chloride	Sigma-Aldrich Co Ltd
Hoechst 33258	Sigma-Aldrich Co Ltd, 94403
Exonuclease III	New England Biolabs MO206L
Exonuclease III buffer 10x	New England Biolabs MO206L

Solutions

Table 9.3: **Solutions.** Table of solutions used throughout this thesis.

Solution	Constituents
----------	--------------

PBS (phosphate buffered saline)	10 mM sodium hydrogen phosphate (Na_2HPO_4), 137 mM sodium chloride (NaCl), 1.8 mM potassium hydrogen phosphate (KH_2PO_4), 2.7 mM potassium chloride (KCl) adjusted to pH 7.4 with hydrogen chloride (HCl) in double distilled water (ddH ₂ O)
PBST	0.02 % Tween-20 in PBS
Permeabilization solution	0.2 % Triton X-100 in PBS
Fixation solution	2% or 4% formaldehyde/PBS freshly made either from electron microscopy grade 16 % formaldehyde ampules (Thermo Scientific) or from molecular biology grade 37 % solution stabilized with 10 % methanol.
Click reaction solution	(per 100 μl) 55 μl ddH ₂ O, 10 μl Tris- HCl buffer (1 M, pH 7), 10 μl sodium ascorbate (500 mM), 5 μl copper sulfate (100 mM), 20 μl fluorescent azide dye (0.1 mM)

Equipment

Table 9.4: **Equipment.** Table of equipment used throughout this thesis.

Item	Source
Pipette	Gibson
Pipette tips	StarLab, TipOne
Serological pipettes (1/5/10/25ml)	Falcon
Reaction tubes (0.5/1.5/2ml)	Eppendorf
Ice bath	Multiple, N/A
Heatblock	Techne DB-2P
Heated incubator, dry	Thermo Fisher Scientific, X5
Inoculation loop	Multiple, N/A
Ethanol burner	Multiple, N/A
Heated shaker incubator	New Brunswick Scientific
	Innova 44
Centrifuge	Eppendorf
	5430R/5804R
Spectrophotometer	Nanodrop 2000
PCR tubes	Qiagen, 981105

PCR thermocycler	Techne 3Prime
Weighing boat	Multiple, N/A
Balance	Multiple, N/A
Bottle (100 ml to 250 ml)	Schott-Duran
Beaker (50 ml to 100 ml)	Schott-Duran
Microwave	Multiple, N/A
Gel boat mold	BioRad
Gel well comb mold	BioRad
Electrophoresis tank	BioRad
UV illuminator	AlphaInnotech
Gel extraction kit	Qiagen, 28704
Plasmid midi kit	Qiagen, 12943
Large construct kit	Qiagen, 12462
Heated water bath	Grant
Tissue culture dishes (10 ml)	Nunc
Tissue culture dishes (6 ml)	Nunc
Tissue culture dishes (6 well, 2 ml)	Nunc
Centrifuge tube (15 ml to 50 ml)	Falcon
Microcentrifuge	Eppendorf
Tissue culture incubator	Thermo electron corporation HeraCell 150
Microscope slides	Thermo Fisher Scientific Superfrost
High precision cover slips	Marienfeld Superior 18mm ² or 22mm ² No 1.5H 170±5 µm
Stainless steel forceps	Multiple, N/A
Sonicator	Multiple, N/A
Desktop pump	Thermo Fisher Scientific
Paraffin based film	Parafilm
Nail varnish (clear)	Multiple, N/A
Humidified dark chamber	Multiple, N/A
Hypodermic needles	BD Microlance
4D Nucleofector	Lonza
Illustra G-50 columns	VWR International Ltd GE healthcare 27-5330-01
UV crosslinker	Spectrolinker XL-1500
Rubber cement	Multiple, N/A

Antibodies

Table 9.5: **Antibodies.** Table of antibodies used throughout this thesis.

Primary Antibody	Source, working dilution
Rabbit mAb anti-CTCF	Cell Signalling Technologies 3418, D31H2
Mouse mAb anti-SCC1	Merck Millipore, 05-908

Rabbit pAb anti-SMC3	Bethyl Laboratories Inc. A300-060A
Rabbit pAb anti-SAF-A	Abcam, ab20666
Rabbit pAb anti-RNAPII Ser2P	Abcam ab5095
Mouse mAb anti-H3K27me3	Active Motif, 61017
Mouse mAb anti-H3K4me2	Active Motif, 39679
Rabbit pAb anti-H3K4me3	Active Motif, 39159
Rabbit pAb anti-H3K36me2	Active Motif, 39255
Rabbit pAb anti-H3K36me3	Active Motif, 61021
Mouse mAb anti-H3K9me2	Abcam, ab1220
Mouse mAb anti-H3K9me3	Active Motif, 61013
Mouse mAb anti-H4K20me1	Abcam, ab78513
Rabbit mAb anti-H4K5ac	Merck Millipore, 04118
Rabbit pAb anti-macroH2A1	Abcam, 39593
Rabbit pAb anti-HP1a	Sigma-Aldrich Co Ltd H2164-200UL
Mouse mAb anti-Ki67	Abcam, ab8191
Rabbit pAb anti-Ki67	Abcam, ab15580
Mouse mAb anti-Cas9	Merck Millipore, MAC133
Rabbit mAb anti-GFP	Life Technologies, G10362
Rabbit pAb anti-LMNB1	Abcam, ab16048
Secondary Antibody	Source, working dilution
Alexa-488 Goat pAb anti-Mouse-IgG	Life Technologies, A11029
Alexa-488 Goat pAb anti-Rabbit IgG	Life Technologies, A11034
Alexa-594 Donkey pAb anti-Mouse-IgG	Thermo Fisher Scientific A21203
Alexa-594 Donkey pAb anti-Rabbit-IgG	Thermo Fisher Scientific A21207
Alexa-647 Goat pAb anti-Mouse-IgG	Abcam, ab150115
Alexa-647 Donkey pAb anti-Rabbit-IgG	Abcam, ab150075

Cell lines

Table 9.6: **Cell lines.** Table of cell lines used throughout this thesis.

Cell line	Description
C127	Mouse female. Mammary gland tumour, adherent epithelial.
IMR90	Human female. Lung, adherent fibroblast.

HT1080 Human male. Connective sarcoma tissue, adherent epithelial.

9.6 Code

Subsequently are the outline of the major scripts employed throughout this thesis, both during the ChaiN workflow but also for pre-processing and clean-up before ChaiN as well as data mining post-ChaiN. Filenames, input arguments and output arguments throughout ChaiN have been controlled automatically through recipes in the Makefile for the Make build system. A schematic representation of the relationship between the input and output of each script is given in Figure 3.4. All code for the running of ChaiN as well as for compiling this thesis has been version controlled through Git and can be accessed via public repositories in Github:

ChaiN

<https://github.com/ezemiron/Chain>

Thesis

<https://github.com/ezemiron/thesis>

Scripts for Pre-processing clean-up

16thrMCNR.imj

ImageJ script - not Makefile managed.

- Requirements:
 - Bioformats importer plugin (silent).
 - SIMcheck plugin.
 - Directory of reconstructed and aligned SIM image “_SIR_EAL.dv”.
 - Directory of aligned RAW SIM image “_EAL.dv”.
- User Inputs:
 1. Choose target directory (from GUI).
 2. Choose destination directory (from GUI).
- For every “SIR_EAL.dv”:
 - Opens and get dimensions.
 - Finds its corresponding RAW.
 - Runs SIMcheck 16 bit threshold and saves “_SIR_EAL_THR.tif”.
 - Opens RAW, and removes negative pixel values if any.
 - Runs SIMcheck modulation contrast map, then closes RAW.

- Thresholds MCNR output (5-100, HC) and makes binary mask.
- Removes overlays if present and scale output to the size of SIR_EAL.dv.
- Saves as “_EAL_MCNR-mask.tif” and closes all images.

Scripts for ChaiN pipeline

Nucleus_mask.m

Octave script - Makefile managed.

- Requirements:
 - Image pkg for octave.
 - Assumes 2 channels.
- Makefile inputs:
 1. Channel of chromatin (HC by makefile).
 2. Path to “_SIR_EAL_THR.tif”.
 3. Path to “_SIR_EAL_THR_mask.tif” (to save it).
- For every argument 2:
 - Opens argument 1 channel.
 - Applies Gaussian blurring.
 - Creates 3D structural element (HC: shape -disk, diameter, z).
 - Thresholds image with “Otsu” algorithm.
 - Dilates it with structural element.
 - Fills holes.
 - Closes the dilation to better match chromatin.
 - Filters out particles lower than 10000 voxels (HC).
 - Writes mask to argument 3.

Foci_centroids.m

Octave script - Makefile managed

- Requirements:
 - Image package.
- Makefile inputs:
 1. Channel of foci (HC by makefile).
 2. Path to “_SIR_EAL_THR_mask.tif”.
 3. Path to “_SIR_EAL_THR.tif”.
 4. Total number of channels in arg 3 (HC by makefile).
- For every argument 3:

- Opens argument 1 channel.
- Mask it with argument 2.
- Threshold with "Otsu".
- Watershed.
- Generate weighted centroids.
- Converts sub-pixel centroids to nm using voxel size (HC 41 41 125).
- Save to standard output: results/.. ../*_SIR_EAL_THR_centroids.csv.

Chromseg.R

R script - Makefile managed

- Requirements:
 - EImage package.
 - Bioimagerools package.
- Makefile inputs:
 1. Channel of chromatin (HC by makefile).
 2. Path to "_SIR_EAL_THR_mask.tif".
 3. Path to "_SIR_EAL_THR.tif".
 4. Path to "_SIR_EAL_THR_SEG.tif" (to save it).
- For every argument 3:
 - Opens argument 1 channel.
 - Use "segment" HMM from bioimagerools (7classes, 30 iterations; HC).
 - Scale to 8bit so imageJ can display output properly.
 - Save to argument 4.

distribution.R

R script - Makefile managed

- Requirements:
 - EImage package.
 - Bioimagerools package.
- Makefile inputs:
 1. Path to "_SIR_EAL_THR_centroids.csv".
 2. Path to "_SIR_EAL_THR_SEG.tif".
 3. Path to "_EAL_MCNr-mask.tif".
- Reads arguments 1, 2 and 3 and checks 2 and 3 have same dimensions.
- For every argument 2:

- Makes mask logical.
- Rounds centroids using voxel (HC 41 41 125).
- Generates 3D linear index using voxel dimensions and dimensions of mask (x,y,z -> a).
- Index mask to cut out centroids that indexed NULL (outside mask).
- Use trimmed list to index SEG.
- Tabulate SEG to find number of voxels in each class (required for normalisation).
- Count how many spots were index in each class (absolute distribution).
- Remove class 0 if any.
- Normalise distribution, make log2 and compile in single dataframe.
- Save to standard output: results/.. ../*_SIR_EAL_THR_distn.csv.

bound.m

Octave script - Makefile managed

- Requirements:
 - Image package.
 - Statistics package.
- Makefile inputs:
 1. Number of reference class, IC.
 2. Path to "_SIR_EAL_THR_SEG.tif".
 3. Path to "_SIR_EAL_THR_bound.tif".
- For every argument 2:
 - Turn SEG to binary, either voxels of the IC or non IC.
 - Use "bwperim" function to find the boundary voxels between the previous binary volumes.
 - Save binary image of boundary voxels to argument 3.

d2bound.m

Octave script - Makefile managed

- Requirements:
 - Image package.
 - Statistics package.
- Makefile inputs:
 1. Number of reference class, IC.

2. Path to “_SIR_EAL_THR_SEG.tif”.
 3. Path to “_SIR_EAL_THR_centroids.csv”.
 4. Path to “_SIR_EAL_THR_bound.tif”.
 5. Path to “_SIR_EAL_THR_mask.tif”.
 6. Path to “_SIR_EAL_THR_md2bound.tif”.
 7. Path to “_SIR_EAL_THR_rd2bound.tif”.
- Reads arguments 1, 2 and 3 and checks 1 and 3 have same dimensions.
 - For every argument 2:
 - Turn SEG to binary, either voxels of the IC or non IC.
 - Rounds centroids with voxel dimensions.
 - Turn all boundary voxels to xyz coordinates.
 - Find all distances between all marker centroid and boundary coordinates.
 - Find minimum distance for each marker centroid.
 - Index SEG and turn all distances from centroids in the IC to negative.
 - Save distances to standard output CSV in argument 5.
 - Generate random spots (used for normalisation controls).
 - repeat distance finding and IC indexing.
 - Save distances from random to standard output CSV in argument 6.

Scripts for post-analysis data mining

summary.R

R script - **not Makefile managed**

- Requirements:
 - Correct data architecture. A directory for each marker (M), under a directory for each chromatin condition (X).
- For every condition directory X:
 - Makes list of marker directories (M) under X.
 - Makes first M directory the working directory.
 - Gets list of all its files.
 - Filters only files ending in “_distn.csv”.
 - For each of these reads 7th column (HC) and adds this column to a new dataframe.
 - Gets average of all columns.
 - Gets log 2 of the average.
 - Saves all columns in one dataframe in the M directory as: M_distn_summary.csv.

Bibliography

- S. H. Ahn, M. Kim, and S. Buratowski. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell*, 2004.
- B. Alberts, A. Johnson, J. Lewis, D Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. 1983.
- V. Ambros. The functions of animal microRNAs. *Nature*, 2004.
- T. Anton, S. Bultmann, H. Leonhardt, and Y. Markaki. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*, 2014.
- R. Balhorn. The protamine family of sperm nuclear proteins. *Genome Biology*, 2007.
- G. Ball, J. Demmerle, R. Kaufmann, I. Davis, I. M. Dobbie, and Schermelleh L. SIMcheck: a toolbox for successful super-resolution structured illumination microscopy. *Scientific Reports*, 2015.
- R. Barrangou and J. A. Doudna. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 2016.
- R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007.

- A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007.
- A. R. Bassett, C. Tibbit, C. P. Ponting, and J. L. Liu. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell reports*, 2013.
- D. B. Beck, H. Oda, S. S. Shen, and D. Reinberg. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes & Development*, 2012.
- A. S. Belmont and K. Bruce. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *The Journal of Cell Biology*, 1994.
- E. Benizri, A. Ginouvès, and E. Berra. The magic of the hypoxia-signaling cascade. *Cellular and Molecular Life Sciences*, 2008.
- E. A. Bennett, H. Keller, R. E. Mills, S. Schmidt, J. V. Moran, O. Weichenrieder, and S. E. Devine. Active Alu retrotransposons in the human genome. *Genome Research*, 2008.
- B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 2006.
- E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacio, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 2006.
- N. Bobroff. Position measurement with a resolution and noise-limited instrument. *Review of Scientific Instruments*, 1986.
- A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, M. Fudenberg, G. Imakaev, L. A. Mirny, C. T. Wu, and X. Zhuang. Super-resolution imaging

- reveals distinct chromatin folding for different epigenetic states. *Nature*, 2016.
- M. Bohn and D. W. Heermann. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS ONE*, 2010.
- A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, R. Eils, C. Cremer, M. R. Speicher, and T. Cremer. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS biology*, 2005.
- B. P. Bratton and J. W. Shaevitz. Simple experimental methods for determining the apparent focal shift in a microscope system. *PLoS ONE*, 2015.
- L. R. Brewer, M. Corzett, and R. Balhorn. Protamine-induced condensation and decondensation of the same DNA molecule. *Science*, 1999.
- B. Burgess-Beusse, C. Farrell, M. Gaszner, M. Litt, V. Mutskov, F. Recillas-Targa, M. Simpson, A. West, and G. Felsenfeld. The insulation of genes from external enhancers and silencing chromatin. *Proceedings of the National Academy of Sciences*, 2002.
- M. Buschbeck and L. Di Croce. Approaching the molecular and physiological function of macroH2A variants. *Epigenetics*, 2010.
- M. J. Carrozza, B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 2005.
- D. Carter, L. Chakalova, C. S. Osborne, Y. F. Dai, and P. Fraser. Long-range chromatin regulatory interactions in vivo. *Nature Genetics*, 2002.
- G. Cavalli, R. Biavasco, B. Borgiani, and L. Dagna. Oncogene-induced senescence as a new mechanism of disease: the paradigm of Erdheim–Chester disease. *Frontiers in Immunology*, 2014.

- V. O. Chagin, C. S. Casas-Delucchi, M. Reinhart, L. Schermelleh, Y. Markaki, A. Maiser, J. J. Bolius, A. Bensimon, M. Fillies, P. Domaing, Y. M. Rozanov, H. Leonhardt, and M. C. Cardoso. 4D Visualization of replication foci in mammalian cells corresponding to individual replicons. *Nature Communications*, 2016.
- S. Chang and S. H. Lamm. Human health effects of sodium azide exposure: a literature review and analysis. *International Journal of Toxicology*, 2003.
- B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas System. *Cell*, 2013.
- B. Chen, J. Hu, R. Almeida, H. Liu, S. Balakrishnan, C. Covill-Cooke, W. A. Lim, and B. Huang. Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Research*, 2016.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 2001.
- P. R. Cook. Predicting three-dimensional genome structure from transcriptional activity. *Nature Genetics*, 2002.
- C. Costanzi and J. R. Pehrson. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, 1998.
- M. Cremer, K. Küpper, B. Wagler, L. Wizelman, J. von Hase, Y. Weiland, L. Kreja, J. Diebold, M. R. Speicher, and T. Cremer. Inheritance of gene density? related higher order chromatin arrangements in normal and tumor cell nuclei. *The Journal of cell biology*, 2003.

- T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2001.
- M. F. Cristofaro, B. L. Betz, C. J. Rorie, D. N. Reisman, W. Wang, and B. E. Weissman. Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. *Journal of cellular physiology*, 2001.
- J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. A. Bickmore. Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of Cell Biology*, 1999.
- L. Cui and J. Miao. Chromatin-mediated epigenetic regulation in the malaria parasite *Plasmodium falciparum*. *Eukaryotic Cell*, 2010.
- C. Davidovich, K. J. Goodrich, A. R. Gooding, and T. R. Cech. A dimeric state for PRC2. *Nucleic Acids Research*, 2014.
- E. de Wit and W. de Laat. A decade of 3C technologies: insights into nuclear organization. *Genes Dev*, 2012.
- A. Dean. On a chromosome far, far away: LCRs and gene expression. *Trends in genetics*, 2005.
- K. M. Dean and A. E. Palmer. Advances in fluorescence labeling strategies for dynamic cellular imaging. *Nature Chemical Biology*, 2014.
- J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 2002.
- E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, and E. Charpentier. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011.
- J. Demmerle, C. Innocent, A. J. North, G. Ball, M. Müller, E. Miron, A. Matsuda, I. M. Dobbie, Y. Markaki, and L. Schermelleh. Strategic and practical

- guidelines for successful structured illumination microscopy. *Nature Protocols*, 2017.
- W. Deng, X. Shi, R. Tjian, T. Lionnet, and Singer R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proceedings of the National Academy of Sciences*, 2015.
- N. Dillon and R. Festenstein. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends in Genetics*, 2002.
- J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 2012.
- J. Dostie, T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R. D. Green, and J. Dekker. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 2006.
- S. R. Eddy. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2001.
- W. E. Evenson, L. M. Boden, K. A. Muzikar, and D. J. O'Leary. ^1H and ^{13}C NMR assignments for the cyanine dyes SYBR Safe and thiazole orange. *The Journal of Organic Chemistry*, 2012.
- I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, and M. L. Tress. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Human Molecular Genetics*, 2014.

- P. J. Fabre, A. Benke, E. Joye, T. H. Nguyen Huynh, S. Manley, and D. Duboule. Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. *Proceedings of the National Academy of Sciences*, 2015.
- E. F. Fornasiero and S. O. Rizzoli. *Super-resolution microscopy techniques in the neurosciences*. 2014.
- G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, 2016.
- M. J. Gamble and W. L. Kraus. Multiple facets of the unique histone variant macroH2A: from genomics to cell biology. *Cell Cycle*, 2010.
- A. A. Gavrilov, I. S. Zukher, E. S. Philonenko, S. V. Razin, and O. V. Iarovaia. Mapping of the nuclear matrix-bound chromatin hubs by a new M3C experimental procedure. *Nucleic Acids Research*, 2010.
- J. H. Gibcus and J. Dekker. The Hierarchy of the 3D Genome. *Mol Cell*, 2013.
- J. Goedhart, D. Von Stetten, M. Noirclerc-Savoye, M. Lelimousin, L. Joosen, M. A. Hink, L. van Weeren, T. W. Jr. Gadella, and A. Royant. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Communications*, 2012.
- T. R. Gregory. Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics*, 2005.
- F. Greil, C. Moorman, and B. van Steensel. DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. *Methods in enzymology*, 2006.
- F. Grosveld, N. Dillon, and D. Higgs. The regulation of human globin gene expression. *Bailliere's clinical haematology*, 1993.
- S. Gruber, C. H. Haering, and K. Nasmyth. Chromosomal cohesin forms a ring. *Cell*, 2003.

- B. Gu and M. G. Lee. Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. *Cell & bioscience*, 2013.
- L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 2008.
- M. G. Gustafsson, L. Shao, P. M. Carlton, C. J. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical journal*, 2008.
- O. Hanscombe, D. Whyatt, P. Fraser, N. Yannoutsos, D. Greaves, N. Dillon, and F. Grosveld. Importance of globin gene order for correct developmental expression. *Genes & Development*, 1991.
- J. C. Hansen. Human mitotic chromosome structure: what happened to the 30-nm fibre? *The EMBO Journal*, 2012.
- B. Harke, J. Keller, C. K. Ullal, V. Westphal, A. Schönle, and S. W. Hell. Resolution scaling in STED microscopy. *Conference on Lasers and Electro-Optics*, 2008.
- S. Hell and J. Wichmann. *Breaking the Diffraction Resolution Limit by Stimulated Emission Depletion Fluorescence Microscopy*. 1994.
- A. D. Hershey and M. Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology*, 1952.
- S. T. Hess, T. P. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 2006.
- N. Hoze, D. Nair, E. Hosy, C. Sieben, S. Manley, A. Herrmann, J. B. Sibarita, D. Choquet, and D. Holcman. Heterogeneity of AMPA receptor trafficking

- and molecular interactions revealed by superresolution analysis of live cell imaging. *Proceedings of the National Academy of Sciences*, 2012.
- F. J. Iborra, A. Pombo, D. A. Jackson, and P. R. Cook. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *Journal of cell science*, 1996.
- J. Irianto, J. Swift, R. P. Martins, G. D. McPhail, M. M. Knight, D. E. Discher, and D. A. Lee. Osmotic challenge drives rapid and reversible chromatin condensation in chondrocytes. *Biophysical journal*, 2013.
- M. K. Iyer, Y. S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T. R. Barrette, J. R. Prensner, J. R. Evans, S. Zhao, A. Poliakov, X. Cao, S. M. Dhanasekaran, Y. M. Wu, D. R. Robinson, D. G. Beer, F. Y. Feng, H. K. Iyer, and A. M. Chinnaiyan. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, 2015.
- D. A. Jackson and A. Pombo. Replicon Clusters Are Stable Units of Chromosome Structure: Evidence That Nuclear Organization Contributes to the Efficient Activation and Propagation of S Phase in Human Cells. *The Journal of cell biology*, 1998.
- X. Ji, D. B. Dadon, B. E. Powell, Z. P. Fan, D. Borges-Rivera, S. Shachar, A. S. Weintraub, D. Hnisz, G. Pegoraro, T. I. Lee, T. Misteli, R. Jaenisch, and R. A. Young. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell*, 2016.
- A. A. Joshi and K. Struhl. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell*, 2005.
- D. Jost, C. Vaillant, and P. Meister. Coupling 1D modifications and 3D nuclear organization: data, models and function. *Current opinion in cell biology*, 2016.

- T. Juven-Gershon, J. Y. Hsu, J. W. Theisen, and J. T. Kadonaga. The RNA polymerase II core promoter - the gateway to transcription. *Current opinion in cell biology*, 2008.
- J. T. Kadonaga. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2011.
- J. J. Kim, N. K. Bennett, M. S. Devita, S. Chahar, S. Viswanath, E. A. Lee, G. Jung, P. P. Shao, E. P. Childers, S. Liu, A. Kulesa, B. A. Garcia, M. L. Becker, N. S. Hwang, A. Madabhushi, M. P. Verzi, and P. V. Moghe. Optical High Content Nanoscopy of Epigenetic Marks Decodes Phenotypic Divergence in Stem Cells. *Scientific reports*, 2017.
- A. Kimura and M. Horikoshi. Tip60 acetylates six lysines of a specific class in core histones in vitro. *Genes to Cells*, 1998.
- J. Kind and B. van Steensel. Genome–nuclear lamina interactions and gene regulation. *Current opinion in cell biology*, 2010.
- J. Kind, L. Pagie, H. Ortabozkoyun, S. Boyle, S. S. de Vries, H. Janssen, M. Amendola, L. D. Nolen, W. A. Bickmore, and B. van Steensel. Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 2013.
- D. Kioussis, E. Vanin, T. deLange, R. A. Flavell, and F. G. Grosveld. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, 1983.
- S. C. Knight, L. Xie, W. Deng, B. Guglielmi, L. B. Witkowsky, L. Bosanac, E. T. Zhang, M. El Beheiry, J. B. Masson, M. Dahan, Z. Liu, J. A. Doudna, and R. Tjian. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, 2015.
- H. C. Kolb, M. G. Finn, and K. B. Sharpless. Click chemistry: diverse chemical function from a few good reactions. *Angewandte Chemie International Edition*, 2001.

- M. Kozak. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, 1987.
- F. Kraus, E. Miron, J. Demmerle, T. Chitiashvili, A. Budco, Q. Alle, A. Matsuda, H. Leonhardt, L. Schermelleh, and Y. Markaki. Quantitative 3D structured illumination microscopy of nuclear structures. *Nature Protocols*, 2017.
- P. Křížek, T. Lukeš, M. Ovesný, K. Fliegel, and G. M. Hagen. SIMToolbox: a MATLAB toolbox for structured illumination fluorescence microscopy. *Bioinformatics*, 2016.
- O. I. Kulaeva, E. V. Nizovtseva, Y. S. Polikanov, S. V. Ulianov, and V. M. Studitsky. Distant activation of transcription: mechanisms of enhancer action. *Molecular and Cellular Biology*, 2012.
- A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Development*, 2002.
- J. R. Lackowicz. *Principles of fluorescence spectroscopy*. 1983.
- S. K. Lee and S. L. Pfaff. Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. *Neuron*, 2003.
- B. Lehnertz, Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. Peters. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biology*, 2003.
- G. Lia, M. Indrieri, T. Owen-Hughes, L. Finzi, A. Podesta, P. Milani, and Dunlap D. ATP-dependent looping of DNA by ISWI. *Journal of Biophotonics*, 2008.

- E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragozy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009.
- C. G. Lobe. Transcription factors and mammalian development. *Current topics in developmental biology*, 1992.
- K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 1997.
- D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 2015.
- H. Ma, A. Naseri, P. Reyes-Gutierrez, S. A. Wolfe, S. Zhang, and T. Pederson. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences*, 2015.
- K. Maeshima, S. Hihara, and M. Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Current opinion in cell biology*, 2010.
- K. Maeshima, K. Kaizu, S. Tamura, T. Nozaki, T. Kokubo, and K. Takahashi. The physical size of transcription factors is key to transcriptional regulation in chromatin domains. *Journal of Physics: Condensed Matter*, 2015.
- K. Maeshima, S. Ide, K. Hibino, and M. Sasai. Liquid-like behavior of chromatin. *Current Opinion in Genetics and Development*, 2016.

- N. L. Mahy, P. E. Perry, S. Gilchrist, R. A. Baldock, and W. A. Bickmore. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of Cell Biology*, 2002.
- D. Managadze, A. E. Lobkovsky, Y. I. Wolf, S. A. Shabalina, I. B. Rogozin, and E. V. Koonin. The vast, conserved mammalian lincRNome. *PLoS Computational Biology*, 2013.
- R. Margueron and D. Reinberg. The Polycomb complex PRC2 and its mark in life. *Nature*, 2011.
- Y. Markaki, M. Gunkel, S. Schermelleh, L. Beichmanis, J. Neumann, M. Heide-
mann, H. Leonhardt, D. Eick, C. Cremer, and T. Cremer. Functional Nuclear
Organization of Transcription and DNA Replication A Topographical Mar-
riage between Chromatin Domains and the Interchromatin Compartment.
Cold Spring Harbor Symposia on Quantitative Biology, 2010.
- Y. Markaki, D. Smeets, S. Fiedler, V. J. Schmid, L. Schermelleh, T. Cremer, and
M. Cremer. The potential of 3D-FISH and super-resolution structured illu-
mination microscopy for studies of 3D nuclear architecture. *BioEssays*, 2012.
- M. R. Matarazzo, S. Boyle, M. D'Esposito, and W. A. Bickmore. Chromosome
territory reorganization in a human disease with altered DNA methylation.
Proceedings of the National Academy of Sciences, 2007.
- L. A. Mirny. The fractal globule as a model of chromatin architecture in the
cell. *Chromosome Research*, 2011.
- E. Miron, C. Innocent, S. Heyde, and L. Schermelleh. In Vivo and In Situ
Replication Labeling Methods for Super-resolution Structured Illumination
Microscopy of Chromosome Territories and Chromatin Domains. *Chromo-
some Architecture*, 2016.
- C. B. Moens and L. Selleri. Hox cofactors in vertebrate development. *Develop-
mental biology*, 2006.

- C. Mondello and A. I. Scovassi. Telomeres, telomerase, and apoptosis. *Biochemistry and Cell Biology*, 2004.
- U. V. Nägerl, K. I. Willig, B. Hein, S. W. Hell, and T. Bonhoeffer. Live-cell imaging of dendritic spines by STED microscopy. *Proceedings of the National Academy of Sciences*, 2008.
- H. Nakayasu and R Berezney. Mapping replicational sites in the eukaryotic cell nucleus. *The Journal of cell biology*, 1989.
- D. M. Nelson, T. McBryan, J. C. Jeyapalan, J. M. Sedivy, and P. D. Adams. A comparison of oncogene-induced senescence and replicative senescence: implications for tumor suppression and aging. *Age*, 2014.
- M. Nicodemi, B. Panning, and A. Prisco. A thermodynamic switch for chromosome colocalization. *Genetics*, 2008.
- D. A. Nix and M. B. Eisen. GATA: a graphic alignment tool for comparative sequence analysis. *BMC bioinformatics*, 2005.
- E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, and Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 2012.
- R. T. O’Keefe, S. C. Henderson, and D. L. Spector. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *The Journal of cell biology*, 1992.
- A. L. Olins and D. E. Olins. Spheroid chromatin units (v bodies). *Science*, 1974.
- F. P. M. Oliveira and J. M. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 2012.
- C. T. Ong and V. G. Corces. CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics*, 2014.

- H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 2017.
- A. F. Palazzo and A. Akef. Nuclear export as a key arbiter of "mRNA identity" in eukaryotes. *Biochimica et biophysica acta*, 2012.
- R. J. Palstra, B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, 2003.
- E. Passarge. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *American journal of human genetics*, 1979.
- T. Pawson. Signal transduction—a conserved pathway from the membrane to the nucleus. *Developmental genetics*, 1993.
- T. Pederson. Half a century of "the nuclear matrix". *Molecular Biology of the Cell*, 2000.
- B. Pei, C. Sisui, A. Frankish, C. Howald, L. Habegger, X. J. Mu, R. Harte, S. Balasubramanian, A. Tanzer, M. Diekhans, A. Reymond, T. J. Hubbard, J. Harrow, and M. B. Gerstein. The GENCODE pseudogene resource. *Genome biology*, 2012.
- J. Perkel. Mapping chromosome neighborhoods. *BioTechniques*, 2015.
- J. Perovanovic, S. Dell'Orso, V. F. Gnoch, J. K. Jaiswal, V. Sartorelli, C. Vigouroux, K. Mamchaoui, V. Mouly, G. Bonne, and E. P. Hoffman. Laminopathies disrupt epigenomic developmental programs and cell fate. *Science Translational Medicine*, 2016.
- J. E. Phillips and V. G. Corces. CTCF: master weaver of the genome. *Cell*, 2009.
- J. E. Phillips-Cremins. Unraveling architecture of the pluripotent genome. *Current opinion in cell biology*, 2014.

- T. A. Planchon, L. Gao, D. E. Milkie, M. W. Davidson, J. A. Galbraith, C. G. Galbraith, and E. Betzig. Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination. *Nature Methods*, 2011.
- A. F. Pluta, A. M. Mackay, A. M. Ainsztein, I. G. Goldberg, and W. C. Earnshaw. The centromere: hub of chromosomal activities. *Science*, 1995.
- B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gülsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren, and D. M. Gilbert. Topologically associating domains are stable units of replication-timing regulation. *Nature*, 2014.
- F. G. Prendergast and K. G. Mann. Chemical and physical properties of aequorin and the green fluorescent protein isolated from *Aequorea forskalea*. *Biochemistry*, 1978.
- D. Pribnow. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 1975.
- P. K. Puvvula, R. D. Desetty, P. Pineau, A. Marchio, A. Moon, A. Dejean, and O. Bischof. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nature Communications*, 2014.
- P. Qin, M. Parlak, C. Kuscü, J. Bandaria, M. Mir, K. Szlachta, R. Singh, X. Darzacq, A. Yildiz, and M. Adli. Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nature communications*, 2017.
- V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 1993.

- T. P. Rasmussen, T. Huang, M. A. Mastrangelo, J. Loring, B. Panning, and Jaenisch R. Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. *Nucleic Acids Research*, 1999.
- N. Rhind and D. M. Gilbert. DNA replication timing. *Cold Spring Harbor Perspectives in Biology*, 2013.
- M. A Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, and M. P. Cosma. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 2015.
- P. W. Rigby, M. Dieckmann, C. Rhodes, and P. Berg. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *Journal of Molecular Biology*, 1977.
- P. J. Robinson, L. Fairall, V. A. Huynh, and D. Rhodes. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences*, 2006.
- H. Romig, F. O. Fackelmayer, A. Renz, U. Ramsperger, and A. Richter. Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *Nature*, 1992.
- M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 2006.
- A. J. Ruthenburg, C. D. Allis, and J. Wysocka. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell*, 2007.
- A. L. Sanborn, S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and

- E. L. Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 2015.
- J. D. Sander and J. K. Joung. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 2014.
- L. Schermelleh, I. Solovei, D. Zink, and T. Cremer. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Methods in Cell Biology*, 2001.
- L. Schermelleh, P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, H. Leonhardt, and J. W. Sedat. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science*, 2008.
- L. Schermelleh, R. Heintzmann, and H. Leonhardt. A guide to super-resolution fluorescence microscopy. *The Journal of Cell Biology*, 2010.
- R. L. Schiltz, C. A. Mizzen, A. Vassilev, R. G. Cook, C. D. Allis, and Nakatani Y. Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *Journal of Biological Chemistry*, 1999.
- S. Selvaraj, J. R. Dixon, V. Bansal, and B. Ren. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nature Biotechnology*, 2013.
- I. Shamovsky and E. Nudler. New insights into the mechanism of heat shock response activation. *Cellular and Molecular Life Sciences*, 2008.
- D. M. Shcherbakova, M. A. Hink, L. Joosen, T. W. Gadella, and V. V. Verkhusha. An orange fluorescent protein with a large Stokes shift for single-excitation multicolor FCCS and FRET imaging. *Journal of the American Chemical Society*, 2012.

- M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B van Steensel, and W. de Laat. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, 2006.
- M. Simonis, J. Kooren, and W. de Laat. An evaluation of 3C-based methods to capture DNA interactions. *Nature Methods*, 2007.
- J. K. Sims, S. I. Houston, T. Magazinnik, and J. C. Rice. A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *Journal of Biological Chemistry*, 2006.
- D. Smeets, Y. Markaki, V. J. Schmid, F. Kraus, A. Tattermusch, A. Cerase, M. Sterr, S. Fiedler, J. Demmerle, J. Popken, H. Leonhardt, N. Brockdorff, T. Cremer, L. Schermelleh, and M. Cremer. Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics & Chromatin*, 2014.
- C. L. Smith and C. L. Peterson. ATP-dependent chromatin remodeling. *Current topics in developmental biology*, 2005.
- I. Solovei, M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, J. Guck, and B. Joffe. Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell*, 2009.
- J. A. Stamatoyannopoulos. What does our genome encode? *Genome Research*, 2012.
- S. Talwar, A. Kumar, M. Rao, G. I. Menon, and G. V. Shivashankar. Correlated spatio-temporal fluctuations in chromatin compaction states characterize stem cells. *Biophysical journal*, 2013.
- A. Tedeschi, G. Wutz, S. Huet, M. Jaritz, A. Wuensche, E. Schirghuber, I. F. Davidson, W. Tang, D. A. Cisneros, V. Bhaskara, T. Nishiyama, A. Vaziri,

- A. Wutz, J. Ellenberg, and J. M. Peters. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*, 2013.
- M. C. Thomas and C. M. Chiang. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, 2008.
- R. E. Thompson, D. R. Larson, and W. W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical journal*, 2002.
- F. Tie, J. Prasad-Sinha, A. Birve, A. Rasmuson-Lestander, and P. J. Harte. A 1-megadalton ESC/E (Z) complex from *Drosophila* that contains polycomb-like and RPD3. *Molecular and Cellular Biology*, 2003.
- B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, and W. de Laat. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell*, 2002.
- T. J. Treangen and S. L. Salzberg. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 2011.
- D. Tuan, W. Solomon, Q. Li, and I. M. London. The "beta-like-globin" gene domain in human erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America*, 1985.
- L. H. Van der Ploeg, A. Konings, M. Oort, D. Roos, L. Bernini, and R. A. Flavell. Gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature*, 1980.
- T. Vanhaecke, T. Henkens, G. E. Kass, and V. Rogiers. Effect of the histone deacetylase inhibitor trichostatin A on spontaneous apoptosis in various types of adult rat hepatocyte cultures. *Biochemical Pharmacology*, 2004.
- D. Voet and J. G. Voet. *Biochemistry*. 1995.

- P. Voigt, W. W. Tee, and D. Reinberg. A double take on bivalent promoters. *Genes & Development*, 2013.
- E. G. Volkova, S. S. Abramchuk, and E. V. Sheval. The overexpression of nuclear envelope protein Lap2 β induces endoplasmic reticulum reorganisation via membrane stacking. *Biology Open*, 2012.
- E. J. Wagner and P. B. Carpenter. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*, 2012.
- G. G. Wang, L. Cai, M. P. Pasillas, and M. P. Kamps. NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nature Cell Biology*, 2007.
- M. Wang, M. W. Mok, H. Harper, W. H. Lee, J. Min, S. Knapp, U. Oppermann, B. Marsden, and M. Schapira. Structural genomics of histone tail recognition. *Bioinformatics*, 2010.
- Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 2008.
- V. Westphal and S. W. Hell. Nanoscale resolution in the focal plane of an optical microscope. *Physical Review Letters*, 2005.
- M. Wijgerde, F. Grosveld, and P. Fraser. Transcription complex stability and chromatin dynamics in vivo. *Nature*, 1995.
- C. L. Woodcock, J. P. Safer, and J. E. Stanchfield. Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental cell research*, 1976.
- M. Woop, R. D. Schwab, J. H. Lee, and A. R. Carter. Optimizing Tethered Particle Motion to Measure DNA Compaction by Protamine. *Biophysical Journal*, 2015.

- R. Xiao, P. Tang, B. Yang, J. Huang, Y. Zhou, C. Shao, H. Li, H. Sun, Y. Zhang, and X. D. Fu. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Molecular cell*, 2012.
- C. Yang, E. Bolotin, T. Jiang, F. M. Sladek, and E. Martinez. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 2006.
- M. Yang, J. Reynoso, M. Bouvet, and R. M. Hoffman. A transgenic red fluorescent protein-expressing nude mouse for color-coded imaging of the tumor microenvironment. *Journal of Cellular Biochemistry*, 2009.
- R. Zhao, T. Nakamura, Y. Fu, Z. Lazar, and D. L. Spector. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nature Cell Biology*, 2011.
- J. Zuin, J. R. Dixon, M. I. van der Reijden, Z. Ye, P. Kolovos, R. W. Brouwer, M. P. van de Corput, H. J. van de Werken, T. A. Knoch, W. F. van Ijcken, F. G. Grosveld, B. Ren, and K. S. Wendt. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA*, 2014.
- J. A. Zuris, D. B. Thompson, Y. Shu, J. P. Guilinger, J. L. Bessen, J. H. Hu, M. L. Maeder, J. K. Joung, Z. Y. Chen, and D. R. Liu. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Biotechnology*, 2015.