# Dynamics in the Regulatory Genome

A thesis submitted for the degree of
Doctor of Philosophy of the University of Oxford




Emily Georgiades

Linacre College
University of Oxford

Wellcome Trust Doctoral Training Programme
in Genomic Medicine and Statistics

Weatherall Institute of Molecular Medicine

Michaelmas Term 2022

# Abstract

The way in which DNA is uniquely folded in the nucleus into specific conformations enables and/or reflects cell-type specific interactions of active *cis*-regulatory elements, namely promoters and enhancers, to regulate specific programs of gene expression. The cohesin protein complex is a key player in the dynamic organisation of genome folding, orchestrating chromatin looping through a process known as loop extrusion. How this process is directed in a cell-type specific manner is currently unknown. A widely accepted hypothesis is that cohesin is recruited to active *cis*-regulatory elements, however current methods are unable to assay the dynamic nature of cohesin loading onto the chromatin. To address this, in this thesis I have developed inducible cellular models that enable cohesin to be tracked in a dynamic manner, in an attempt to answer this question.

Initially, I investigated if a single active enhancer was capable of recruiting cohesin to a neutral region of the genome. Using a high-resolution chromosome conformation capture approach, I showed that a new 3D domain was formed following activation of the enhancer, and that cohesin was recruited to this locus. Whilst this provided convincing evidence that cohesin plays a key role in this process, perhaps via loop extrusion mediated domain formation, data from a single locus provides no way of testing where cohesin was first recruited upon the activation of the domain. In the next two chapters I developed two novel cellular models to investigate cohesin dynamics on a genome-wide scale by orthologous 'push' and 'pull' technologies. In doing so I aimed to investigate whether active *cis*-regulatory elements are responsible for driving the recruitment of cohesin on a genome-wide scale.

# Acknowledgements

Firstly, I would like to thank Doug Higgs for taking me on in the Higgs group for my first rotation project, and Marieke Oudelaar for showing me the ropes during this project. I would like to thank Jim Hughes for taking a leap of faith and allowing me to complete a computational rotation project in his lab and letting me stay to complete my PhD project. The COVID-19 pandemic was a curveball halfway through my PhD, Jim kept me motivated and enthusiastic about science at times when I felt like giving up. I will never forget the pat on the back I got when I got my first good result when back in the lab post-pandemic, some might say this is the equivalent to a Paul Hollywood handshake…

The WIMM has been an amazing place to do research and provided a collaborative and welcoming environment from which I have been lucky to enough to benefit from throughout my PhD. There are too many people to thank individually, but in particular I would like to give a big shout out to Philip Hublitz and the rest of the genome engineering team, Aude-Anaïs Olijnik for teaching me western blot, Jill Brown for her expert guidance on microscopy (and also moral support when things went wrong!), Damien Downes for his encyclopaedic knowledge of experimental techniques and Tom Milne for the invaluable discussions during the final year of my PhD. Last but not least, I would like to thank Nigel Roberts, who has been a tremendous help throughout my PhD both in teaching me techniques and troubleshooting them, to keeping spirits high and having a good laugh in the lab. I will look back fondly on the endless afternoons of ChIP-seq, soundtracked by Granny Tinder on Radio 1.

I also could not have survived the rollercoaster of PhD life without the weird and wonderful coffee-come-lunch breaks spent sat outside on the concrete block in the car park, soaking up the sun with my fellow PhD students in the Higgs/Hughes lab.

Finally, I want to thank my Mum, my Dad, and my amazing sisters Charlotte and Liz for always believing in me and supporting my slightly mad ambitions.

This work was made possible by generous funding from the Wellcome Trust.

# Table of Contents

**Abstract** ..... 2

**Acknowledgements** ..... 3

**Table of Contents** ..... 4

**List of Figures** ..... 9

**List of Tables** ..... 12

**List of Acronyms** ..... 13

**1 Introduction** ..... 16

**1.1 The regulatory genome** ..... 16

1.1.1 Promoter elements ..... 17

1.1.2 Enhancer elements ..... 17

**1.2 Regulation of enhancer-promoter contacts** ..... 18

i. Polymerase tracking ..... 18

ii. DNA looping ..... 18

iii. Phase separation ..... 19

iv. Loop extrusion ..... 19

**1.3 The hierarchical 3D organisation of the genome** ..... 22

1.3.1 Chromosome territories ..... 22

1.3.2 A and B compartments ..... 22

1.3.3 Topologically associated domains ..... 23

**1.4 Cohesin drives the formation of 3D structures via loop extrusion** ..... 26

1.4.1 Cohesin ..... 26

1.4.2 CTCF ..... 28

**1.5 Dynamic properties: loading, translocation and offloading** ..... 29

1.5.1 NIPBL-MAU2, a cohesin loading complex ..... 29

1.5.2 Cohesin translocation ..... 31

1.5.3 WAPL-PDS5, a cohesin unloading complex ..... 31

**1.6 Number crunching: timing cohesin dynamics** ..... 34

1.6.1 Residence times ..... 34
1.6.2 Extrusion rates ..... 36
**1.7 What can degron models tell us about the role of interphase cohesin? ..... 36**
1.7.1 Degron technologies ..... 37
1.7.2 Cohesin degron models ..... 38
1.7.3 CTCF degron models ..... 39
1.7.4 WAPL degron models ..... 40
1.7.5 NIPBL degron models ..... 41
**1.8 Limitations of existing approaches and the need for dynamics in the analysis of chromatin associated proteins ..... 41**
1.8.1 Current genomic methods do not capture dynamic processes ..... 41
1.8.2 Fixation methods may not capture protein occupancy at transient sites ..... 42
1.8.3 Incomplete degradation may invalidate conclusions ..... 42
**1.9 Thesis aims ..... 44**
**2 Cohesin recruitment at a single locus ..... 46**
**2.1 Introduction ..... 46**
2.1.1 The alpha-globin locus ..... 46
2.1.2 Methods to study 3D genome architecture ..... 49
2.1.3 Evolution of the 3D genome landscape during erythroid differentiation ..... 51
2.1.4 A single erythroid-specific enhancer can direct the formation of a regulatory domain ..... 53
**2.2 Results ..... 56**
2.2.1 High resolution Tiled-C data reveals sub-TAD-like structure ..... 56
2.2.2 Use of $RAD21^{TST}$ for reliable cohesin ChIP-seq ..... 59
2.2.3 Activation of the R2 erythroid enhancer is linked with cohesin recruitment ..... 60
**2.3 Discussion ..... 64**
**2.4 Acknowledgements ..... 68**
**3 Genome-wide methods to study dynamics using a degron system ..... 69**
**3.1 Introduction ..... 69**
3.1.1 Moving on from an equilibrium view ..... 69
3.1.2 Degron systems as a means of modelling dynamics ..... 70
3.1.3 The dTAG system for targeted degradation of RAD21 ..... 73
**3.2 Results ..... 75**
3.2.1 Genome engineering strategy ..... 75
3.2.2 System validation by subcellular fractionation and western blot ..... 78
3.2.4 Visualising the degradation at the chromatin level ..... 87
3.2.5 Monitoring $RAD21^{FLAG}$ degradation over a 4-hour time-course using Calibrated ChIP-seq ..... 97

3.2.6 How does the distribution of RAD21$^{FLAG}$ change throughout the degradation time-course? ..... 100
3.2.7 Can RAD21$^{FLAG}$ peaks be categorised based on the degradation trajectory? ..... 100
**3.3 Discussion ..... 108**
**3.4 Acknowledgements ..... 110**
**4 Genome-wide methods to study dynamics using RAPID-release ..... 111**
**4.1 Introduction ..... 111**
4.1.1 RAPID-release: a system to control cohesin influx ..... 113
4.1.2 Distinguishing between endogenous and engineered RAD21 ..... 115
**4.2 Results ..... 116**
4.2.1 RAPID-release in HeLa ..... 116
4.2.2 Optimising ChIP-seq conditions for time-course experiment ..... 127
4.2.3 RAPID-release in HEK293 ..... 131
4.2.4 A new strategy for RAPID-release in mESCs ..... 142
**4.3 Discussion ..... 148**
4.3.1 Mitochondria counts ..... 149
4.3.2 Nuclear export signal (NES) versus nuclear localisation signal (NLS) ..... 149
4.3.3 Size of RAPID-release machinery ..... 150
**4.4 Data availability ..... 150**
**4.5 Acknowledgments ..... 151**
**5 Discussion ..... 152**
**Dynamic gene regulation is evolutionarily important ..... 152**
Cohesin loading is directed to specific elements ..... 154
Cohesin is loaded in a non-directed manner, and translocates to specific regulatory elements ..... 156
A new hypothesis ..... 157
**The caveats of current methods for studying dynamic genome regulation ..... 159**
Enhancers and promoters are difficult to classify ..... 159
Cohesin dynamics is intertwined with the cell cycle ..... 160
A subset of cohesin may be resistant to depletion ..... 162
**Future perspectives ..... 166**
Alternative inducible systems ..... 166
Repurposing the cohesin-dTAG model to investigate RNA Pol II loading ..... 166
Single-cell techniques are required to truly understand dynamic systems ..... 167
**6 Materials and methods ..... 168**
**6.1 Cell culture ..... 168**
6.1.1 General culture procedure ..... 168

6.1.2 Mouse embryonic stem cells ..... 169
6.1.3 Human embryonic kidney 293(T) cells ..... 169
6.1.4 Human HeLa cells ..... 170
6.1.5 Differentiation of mESCs into embryoid bodies (EBs) for the isolation of CD71+ erythroid cells ..... 170
6.1.6 Expansion of single cells from FACS ..... 171
**6.2 Genome engineering strategies ..... 171**
6.2.1 Artificial Locus ..... 171
6.2.2 dTAG degron system ..... 172
6.2.3 RAPID-release ..... 172
6.2.4 Transient transfection of HeLa and HEK293 cell lines with RAPID-release ..... 178
6.2.5 Generation of RAPID-release stable lines in HEK293(T) cells ..... 178
6.2.6 Generation of RAPID-release stable lines in mESC E14 cells ..... 179
**6.3 Identification of positive clones ..... 179**
6.3.1 Agarose gel electrophoresis ..... 179
6.3.2 PCR screening ..... 180
6.3.3 Sanger sequencing ..... 182
**6.4 Induction methods ..... 183**
**6.5 Live cell imaging ..... 183**
6.5.1 Microscopy ..... 183
6.5.2 Image analysis ..... 184
**6.6 Protein methods ..... 184**
6.6.1 Whole-cell protein extraction ..... 184
6.6.2 Extraction of proteins from subcellular compartments ..... 184
6.6.3 Fluorescent western blot ..... 185
6.6.4 Western blot analysis ..... 186
**6.7 RNA isolation and gene expression analysis ..... 187**
6.7.1 RNA isolation and cDNA synthesis ..... 187
6.7.2 RT-qPCR expression analysis ..... 187
**6.8 Fixation condition optimisation ..... 188**
6.8.1 Testing different combinations of fixatives ..... 188
6.8.2 Testing quenching reagents ..... 188
**6.9 Chromatin Immuno-precipitation (ChIP) ..... 189**
6.9.1 Modified chromatin immunoprecipitation protocol for FLAG antibodies ..... 190
6.9.2 Chromatin immunoprecipitation using pre-conjugated magnetic FLAG beads ..... 193
6.9.2 Quality control steps ..... 193
6.9.3 qPCR ..... 194
6.9.4 Next Generation Sequencing ..... 194
**6.10 Tiled-C ..... 195**

**6.11 ATAC-seq ..... 195**
**6.12 Bioinformatic analysis ..... 196**
6.12.1 ChIP-seq and ATAC-seq ..... 198
6.12.2 Calibrated ChIP-seq ..... 198
6.12.2 Custom genome builder ..... 199
6.12.3 Tiled-C ..... 199
6.12.4 Peak calling using Lanceotron ..... 199
6.12.5 CTCF MOTIF analysis ..... 199
6.12.6 RAD21 degradation analysis ..... 200
6.12.7 Hi-C visualisation ..... 200
6.12.8 Code availability ..... 200
**Supplementary Materials ..... 201**
**References ..... 213**

# List of Figures

## Chapter 1: Introduction

Figure 1.1: General principles in gene regulation. ....21
Figure 1.2: The hierarchical organisation of interphase DNA .... 25
Figure 1.3: The vertebrate cohesin complex and its role in loop extrusion. .... 27
Figure 1.4: Model of how chromatin loop length is controlled by the acetylation cycle of cohesin. 33
Figure 1.5: Model for WAPL-dependent cohesin release from chromatin. ....34

## Chapter 2: Cohesin recruitment at a single locus

Figure 2.1: Comparison of the genomic profiles of the mouse alpha-globin locus in WT primary erythroid cells versus embryonic stem cells. ....48
Figure 2.3: Tiled Capture-C throughout differentiation.... 53
Figure 2.4: Comparison of the wild-type (WT/chr11-R2) and R2 enhancer model (chrX-R2). .... 54
Figure 2.5: Hypothetical mechanism of cohesin recruitment to active enhancers.... 55
Figure 2.6: Insertion of the R2 erythroid-specific enhancer into a neutral region of the genome initiates the formation of a TAD-like structure. .... 57
Figure 2.7: The unedited (WT) neutral region of the genome initiates is devoid of the TAD-like structure seen in the R2 insertion model.... 58
Figure 2.8: ChIP-seq of the cohesin subunits around the Dppa5a locus. .... 59
Figure 2.10: Comparison of RAD21 ChIP-seq in WT mES E14 cells with those containing the inserted R2 enhancer.... 61
Figure 2.11: Insertion of the R2 enhancer element directs cohesin (RAD21) recruitment to the locus. ....63
Figure 2.12: Comparison of the chromatin landscape following the insertion of the R2 element into chromosome X.... 67

## Chapter 3: Genome-wide methods to study dynamics using a degron system

Figure 3.1: RAD21 sites may differ in the ability to retain stably associated RAD21, which should be reflected in the ChIP-seq tracks during a degradation time-course.... 72
Figure 3.2: Hijacking a cell's intrinsic degradation machinery: the mechanism of action of dTAG-13 mediated degradation.... 74
Figure 3.3: RAD21-FKBP$^{F36V}$ (dTAG)-4xFLAG donor vector map. .... 75
Figure 3.4: A single RAD21$^{FLAG}$ homozygous clone identified by PCR screening and validated by Sanger sequencing.... 77
Figure 3.5: Simplified diagram of the subcellular isolation strategy. .... 79
Figure 3.6: Subcellular fractionation of WT mESC cells.... 80

Figure 3.7: Degradation of RAD21$^{FLAG}$ over a 24-hour time-course following treatment with 500 nM of dTAG-13. ..... 82
Figure 3.8: Titration of dTAG-13 concentration for a 4-hour degradation. ..... 84
Figure 3.9: RAD21 does not rapidly reaccumulate following dTAG-13 treatment. ..... 86
Figure 3.10: Comparison of different antibodies for chromatin immunoprecipitation of RAD21$^{FLAG}$. ..... 91
Figure 3.11: ChIP-seq comparing different antibodies for RAD21-FLAG ChIP. ..... 92
Figure 3.12: ChIP-seq comparing different antibodies for RAD21-FLAG ChIP. ..... 93
Figure 3.13: RAD21$^{FLAG}$ ChIP-seq pre and post dTAG-13 treatment at three active loci in mouse embryonic stem cells. ..... 95
Figure 3.14: RAD21 is degraded over a 4hr period. ..... 96
Figure 3.16: Heatmap of genome-wide RAD21$^{FLAG}$ signal during a 4-hour degradation time-course. 99
Figure 3.17: Quantification of peaks during RAD21 degradation time-course. ..... 104
Figure 3.18: Different classes of degradation dynamics, are enriched for peaks that intersect with different classes of *cis*-regulatory elements (CREs). ..... 105
Figure 3.19: Intersection of mESC Hi-C data (Bonev et al.) with RAD21 and CTCF ChIP-seq peaks, demonstrating the alignment of RAD21 persistent peaks with TAD structures. ..... 107

**Chapter 4: Genome-wide methods to study dynamics using RAPID-release**

Figure 4.1: Prediction of the calibrated ChIP-seq time-course for induction of RAD21 if cohesin is loaded at the enhancer. ..... 112
Figure 4.2: Schematic representation of the RAPID-release technique. ..... 114
Figure 4.3: Plasmid map of the RAD21$^{TST}$ RAPID-release dual expression vector. ..... 117
Figure 4.5: Comparison of cells transfected with circular H3.1, RAD21$^{TST}$ and empty vector (EV) RAPID-release plasmids by live cell imaging (20x magnification) in HeLa cells. ..... 123
Figure 4.6: Comparison of cells transfected with circular H3.1, RAD21$^{TST}$ and empty vector (EV) RAPID-release plasmids by live cell imaging (63x magnification) in HeLa cells. ..... 124
Figure 4.7: Live cell imaging of HeLa cells transiently expressing RAPID-release (H3.1, RAD21$^{TST}$ and empty vector (EV)) before and after the addition of Rapamycin (63x magnification). ..... 125
Figure 4.8: Image analysis showing the change in fluorescence profiles of the nucleus and mitochondria throughout the RAPID-release experiment. ..... 126
Figure 4.9: Assessing the efficiency of different fixation conditions using ChIP-qPCR. ..... 129
Figure 4.10: Comparison of the quenching efficiency of tris and glycine using ChIP-qPCR. ..... 130
Figure 4.11: Live cell imaging of HEK293 cells transiently expressing RAD21 RAPID-release 48hrs post transfection (10x magnification). ..... 132
Figure 4.12: FACS gating strategy for sorting populations of HEK293 cells based on expression levels of mCherry and EGFP. ..... 133
Figure 4.13: HEK293 RAD21$^{TST}$ RAPID-release stable lines without induction (63x magnification). .. 135
..... 138

Figure 4.14: Subcellular fractionation of HEK293 cells. ..... 138
Figure 4.15: Comparison of RAD21TST ChIP-seq in RAPID-release HEK293 cells F9 and D11 before and after the addition of Rapamycin. ..... 141
Figure 4.16: Ct RT-qPCR analysis to investigate copy number variation in mESC RAPID-release clones compared to HEK293 clone. ..... 143
Figure 4.17: the CRISPR-cas9 homology directed repair (HDR) strategy for the insertion of the RAPID-release machinery into a landing pad that lies within a neutral region of chromosome X... 145
Figure 4.18: Plots showing the fluorescence activated cell sorting (FACS) gating strategy for the CRISPR-cas9 RAPID-release mES E14 cells 24hrs post transfection. ..... 146
Figure 4.19: Colony screening by PCR strategy for RAD21 RAPID-release (EF1a) in mESCs. ..... 147

## Chapter 5: Discussion

Figure 5.1: Is cohesin loading directed or random? ..... 154
Figure 5.2: A proposed mechanism for cohesin can be insulated from proteasomal degradation. 159
Figure 5.3: Genome architecture changes throughout the cell cycle. ..... 161
Figure 5.4: Most published studies do not achieve complete depletion of the target protein. ..... 164

## Chapter 6: Materials and methods

Figure 6.1: In vitro erythroid differentiation. ..... 170
Figure 6.2: Quality control of ChIP-seq sonication efficiency. ..... 193

## Supplementary Materials

Figure S1: Defining a neutral region of the genome. ..... 201
Figure S2: Comparison of the wild-type (WT/chr11-R2) and R2 enhancer model (AL2R2). ..... 202
Figure S3: Raw greyscale western blot membrane images for Figure 3.6. ..... 203
Figure S4: Raw greyscale western blot membrane images for Figure 3.7 ..... 204
Figure S5: Raw greyscale western blot membrane images for Figure 3.8. ..... 205
Figure S6: Raw western blot membrane images for Figure 3.9. ..... 206
Figure S7: Plasmid map of the original dual vector H3.1 RAPID-release vector from Andrew Bowman's lab (University of Warwick)[188] ..... 207
Figure S8: Plasmid map of the single plasmid (dual vector) empty RAPID-release vector from Andrew Bowman's lab (University of Warwick)[188] ..... 208
Figure S9: Plasmid map of the single vector empty RAPID-release vector from Andrew Bowman's lab (University of Warwick)[188] ..... 209
Figure S10: Raw western blot membrane images for Figure 4.14A. ..... 210
Figure S11: Raw western blot membrane images for Figure 4.14B. ..... 211
Figure S12: Raw greyscale membrane images of TOMM20 and VDAC western blots ..... 212

# List of Tables


Table 6.1 Guide sequences for CRISPR-cas9 editing. ....173
Table 6.2 Donor sequences for CRISPR-cas9 editing. .... 174
Table 6.3 Plasmid used for RAPID-release ....177
Table 6.4 DNA extraction lysis buffer for colony screening....180
Table 6.5 Screening primers for RAPID-release EF1a HDR targeting to chrX....181
Table 6.6 Screening primers for dTAG HDR targeting to chrX. .... 182
Table 6.7 Antibodies used for ChIP and western blot experiments.... 186
Tables 6.8 RT-qPCR primers for copy number analysis of RAPID-release clones.... 188
Table 6.9 Buffer substitution for 'gentle' ChIP protocol ....191
Table 6.10 RIPA wash buffer ....191
Table 6.11 ChIP dilution Buffer. ....191
Table 6.12 SDS Lysis/Elution Buffer. .... 192
Table 6.13 Gentle lysis buffer. .... 192
Table 6.14 Gentle wash buffer. .... 192
Table 6.15 Primer sequences for RAD21 ChIP-qPCR ....194
Table 6.16 Packages and versions used for bioinformatic analysis ....196
Table 6.17 Publicly available genomic data used in analyses .... 197

# List of Acronyms

| | |
|---|---|
| 3C | Chromosome conformation capture |
| 3D | Three-dimensional |
| ATAC-seq | Assay for transposase-accessible chromatin sequencing |
| ATCC | American type culture collection |
| ATP | Adenosine triphosphate |
| CD71 | Cluster of Differentiation 71 |
| cDNA | Complementary DNA |
| ChIP-seq | Chromatin immunoprecipitation-sequencing |
| CID | Chemical induced dimerization |
| CRE | *Cis* regulatory element |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CTCF | CCCTC-binding factor |
| CTD | C-terminal domain |
| DMEM | Dulbecco's modified eagle medium |
| DMSO | Di-methyl sulfoxide |
| DNA | Deoxyribose nucleic acid |
| DSG | Disuccinimidyl glutarate |
| dTAG | Degradation TAG |
| EB | Embryoid body |
| EDTA | Ethylenediaminetetraacetic acid |
| EGFP | Enhanced green fluorescent protein |
| eRNA | Enhancer ribonucleic acid |
| ESC | Embryonic stem cells |
| EtBr | Ethidium bromide |
| EV | Empty vector |
| FA | Formaldehyde |
| FACS | Fluorescence-activated cell sorting |
| FBS | Foetal bovine serum |
| FIMO | Find individual motif occurrences |
| FISH | Fluorescence in situ hybridisation |
| FKBP | FK506 binding protein |

| | |
|---|---|
| FRB | FKBP12-Rapamycin-binding |
| GMEM | Glasgow's minimal essential medium |
| HDR | Homology directed repair |
| HEPES | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| IDR | Intrinsically disordered regions |
| IAA | Indole-3-acetic acid |
| LP | Landing pad |
| MACS | Magnetic-activated cell sorting |
| MCC | Micro-Capture-C |
| MEM | Minimum essential medium |
| mESC | Mouse embryonic stem cell |
| MOPS | 3-morpholinopropane-1-sulfonic acid |
| mTOR | Mechanistic target of rapamycin |
| NEAA | Non-essential amino acids |
| NES | Nuclear export signal |
| NLS | Nuclear localisation signal |
| NTD | N-terminal domain |
| OMM | Outer mitochondrial membrane |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PIC | Protease inhibitor complex |
| POI | Protein of interest |
| qPCR | Quantitative PCR |
| $RAD21^{FLAG}$ | FLAG-tagged RAD21 |
| $RAD21^{TST}$ | Twin Strep tagged RAD21 |
| RE | Restriction endonuclease |
| RNA | Ribonucleic acid |
| ROI | Region of interest |
| RPKM | Reads per kilobase of exon per million reads mapped |
| RPM | Revolutions per minute |
| RT | Room temperature |
| SDS | Sodium dodecyl sulfate |
| TBE | Tris borate EDTA |
| TRIS | Tris(hydroxymethyl)aminomethane |
| TST | Twin Strep tag |

| | |
|---|---|
| WCL | Whole cell lysis |
| WIMM | Weatherall Institute of Molecular Medicine |
| WT | Wild type |

# 1 Introduction

One of the most remarkable observations in mammalian biology is the ability for an entire organism, made up of 200 different cell-types[1], to develop from a single fertilised egg. The instruction book for this process is encoded in the DNA, which contains the entire catalogue of ~20,000 protein-coding genes required in these cells[2]. For each unique cell-type to express only the proteins necessary for their individual functions, genes are tightly regulated to ensure only specific genes are 'switched-on' and that others remain silenced. The fundamental question underlying this thesis is to understand the role of a protein called cohesin in this complex process. To address this question, the various stages of gene regulation will first be reviewed.

## 1.1 The regulatory genome

The regulation of gene expression is governed by DNA elements in the non-protein-coding regions of the genome, which is often referred to as the 'non-coding', or 'regulatory' genome. Given that these regulatory elements do not directly encode protein sequences, their importance was previously overlooked. However, over recent years the significance of the non-coding genome has become increasingly appreciated, particularly with the rise of genome-wide association studies (GWAS), in which the majority of disease associated

variants reside in these non-coding, regulatory regions [3,4]. These regulatory elements form the grammar of the genome, and through unique combinations of regulatory elements, cell-type specific gene expression patterns are instructed (Figure 1.1C). Broadly speaking, there are three types of regulatory elements: enhancers and promoters, which act to positively regulate transcription, and insulator elements which frequently bind the CCCTC-binding factor (CTCF).

### 1.1.1 Promoter elements

As the name suggests, the role of the promoter elements is to enable transcription initiation. They are typically short sequences (100bp–1kb) located proximally to the transcription start site and are orientated depending on the direction of transcription. Active promoters, i.e., those associated with actively transcribed genes, reside in regions of open chromatin and are associated with tri-methylation of histone H3 lysine 4 residues (H3K4me3)[5,6] (Figure 1.1B). They provide a platform on which general transcription factors can bind and the pre-initiation complex can assemble. To bring together the full compendium of cell specific and general transcription factors required to initiate cell-type specific gene expression, many promoters interact with specific enhancers, or clusters of enhancer elements (Figure 1.1A).

### 1.1.2 Enhancer elements

Enhancers regulate spatiotemporal, cell-type specific patterns of gene expression, through the recruitment of cell-type specific transcription factors. When active, chromatin associated with these elements is enriched for mono-methylation of histone H3 lysine 4 residues (H3K4me1) and acetylation of histone H3 lysine 27 residues (H3K27ac)[7,8] (Figure 1.1B). Compared to promoters, enhancers are located more distally from the TSS,

up to distances of 1Mb in some cases[9], and can operate independently of the orientation of the target gene.

## 1.2 Regulation of enhancer-promoter contacts

For enhancer and promoter elements to interact, the associated chromatin must adopt a specific 3D conformation that brings them into close proximity. Next, I will briefly discuss the current models that could explain these interactions.

### i. Polymerase tracking

This model describes the action of RNA Polymerase II binding initially to an enhancer from which it moves unidirectionally along the DNA until it finds the target promoter [10-12]. This mechanism is evidenced by chromatin immunoprecipitation experiments which can detect the unidirectional movement of DNA-TF-enhancer complexes progressing towards the promoter[11], as well as the presence of polyadenylated RNAs transcribed from enhancer DNA that occasionally span the region between the enhancer and the promoter[13-15].

### ii. DNA looping

This mechanism suggests that enhancers and promoters are brought into close proximity through interactions between transcription factors that are dynamically bound to these regulatory elements, facilitating loop formation [16-18]. In a landmark study performed by the Blobel lab[19], they experimentally induced contacts between the mouse beta-globin promoter and the locus control region (LCR) and showed that transcription could be initiated in the absence of the tissue specific transcription factor GATA1. They also showed in a follow-up study that forced looping has the capacity to reactivate transcriptionally silenced genes[20].

### iii. Phase separation

One widely discussed concept in the field is that of phase separation, specifically liquid-liquid phase separation which has been suggested to play a role in facilitating enhancer-promoter contacts. It is thought that intrinsically disordered regions (IDRs) of proteins, such as the enhancer/promoter-bound TFs, can attract IDRs of other proteins by engaging in weak ionic interactions to establish a phase condensate. Phase separation could help to explain the establishment of non-direct contact between enhancers and promoters, as well as the ability for a single promoter to interact with numerous enhancers[21,22]. In a recent study by Crump *et al.*[23] they showed that the dissolution of phase condensates caused a reduction in gene transcription, but had no effect on enhancer-promoter contacts. This provides evidence that phase separation is more likely a consequence of enhancer-promoter contact, rather than a driver of interactions.

### iv. Loop extrusion

Currently, the most widely accepted hypothesis for how DNA looping occurs is the model known as 'loop extrusion', which was first proposed by Fudenberg and colleagues in 2016[24]. In this model chromatin is bound by a loop extruding factor (cohesin) which is thought to entrap chromatin within its ring structure and utilising its ATP driven motor activity, reels DNA through the ring until an insulator element bound by CTCF is reached and further extrusion is blocked (Figures 1.3C and 1.3D). Sites enriched for cohesin and CTCF are commonly found near enhancers and promoters[25-27], which is suggestive of their role in facilitating the physical interaction and communication between enhancers and promoters. Evidence for this was also shown in a recent publication by Hua and colleagues[28], in

which they linked the increased loading of cohesin to the activation of enhancers and promoters. They also showed that a lack of cohesin results in a subsequent lack of interactions between CTCF bound sites, which directly supports the loop extrusion hypothesis.

In summary, there are several models of interaction; most likely, enhancer-promoter contact is regulated by a complex combination of mechanisms. There have been a number of studies showing that the disruption of enhancer-promoter proximity, either through mutations in the regulatory element or via genome rearrangements, can give rise to developmental abnormalities or disease phenotypes[9,29-32]. This suggests that the 3D organisation of chromatin *per se* plays a role in moderating biological processes such as long-distance gene regulation[33] and gene transcription[34].

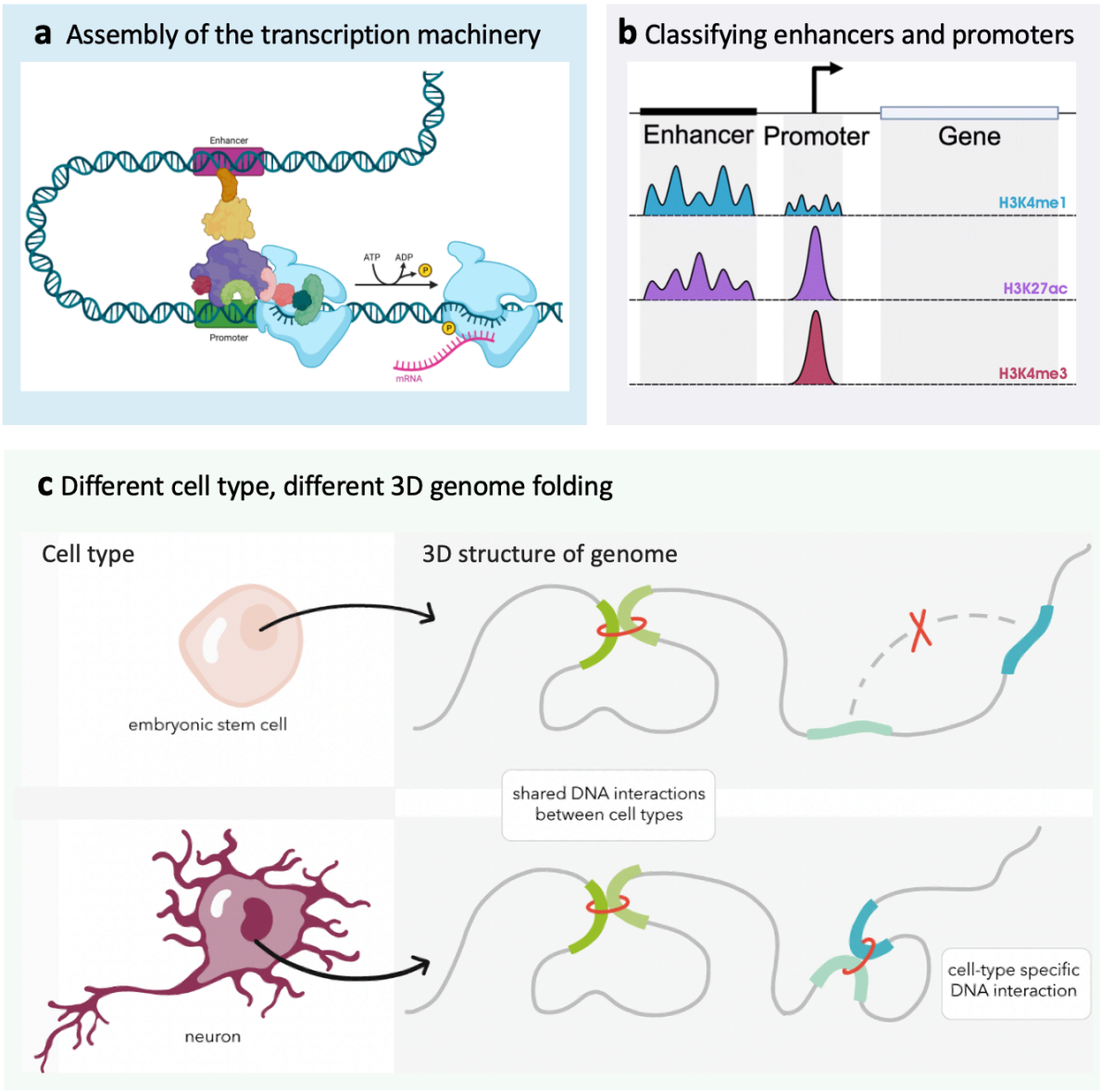



**Figure 1.1: General principles in gene regulation.**

**a** | By DNA folding, the enhancer and promoter are brought into close proximity, bringing together their associated transcription factors and co-factors to assemble the transcription complex that regulates tissue-specific gene transcription. Once assembled, transcription is initiated, and RNA Polymerase II can begin to transcribe DNA into RNA in an ATP dependent process. **b** | Diagram shows the basic principle underlying enhancer and promoter classification. Promoters are found proximal to the transcription start site (TSS) and are commonly marked with punctate peaks in H3K4me3 and H3K27Ac, and often a background H3K4me1 signal. Enhancers are located more distally from the TSS and are marked with broad peaks of H3K4me1 and H3K27Ac, and sometimes background levels of H3K4me3. **c** | Illustration of how genome folding can influence which regulatory elements interact in a cell-type specific manner. Figure adapted from Harvard University blog[35].

## 1.3 The hierarchical 3D organisation of the genome

The spatial organisation of the genome can be described in a hierarchical manner, first by compartmentalisation of the genome into chromosome territories which contain A and B compartments, and then by considering DNA looping which gives rise to topologically associated domains (TADs), sub-TADs and chromatin loops.

### 1.3.1 Chromosome territories

The first level of organisation are the chromosomes, which during interphase have been shown to occupy distinct nuclear territories[36] (Figure 1.2A), a phenomenon first discovered by Rabl almost 120 years ago[37]. The development of modern *in situ* hybridisation techniques such a fluorescence *in situ* hybridisation (FISH) has allowed individual chromosomes to be distinctly visualised in an intact nucleus[38]. Within these territories, chromosomes predominantly interact with themselves rather than other chromosomes[36]. This chromosomal level of organisation does not directly influence gene regulation.

### 1.3.2 A and B compartments

Within these chromosome territories, chromatin segregates into active (A) and inactive (B) compartments (Figure 1.2B). The A compartment generally comprises gene-rich euchromatin that exhibits histone marks characteristic of actively transcribed genes (e.g., H3K4me3 and H3K36me3) residing close to nuclear speckles[39-41]. The B compartment is typically located at the periphery of the nucleus, associated with the nuclear lamina or nucleolus and contains transcriptionally repressed, gene-poor regions of inactive heterochromatin[42]. The A and B compartments were first discovered using Hi-C, a chromosome conformation capture (3C) technique that maps all-vs-all interactions[42]. The characteristic plaid pattern seen on a typical Hi-C map distinguishes the A and B

compartments showing contacts within the same compartment are enriched, and between compartments is depleted. One explanation for this pattern could be that chromatin has a preferential affinity for chromatin with similar chromatin modifications, which can give rise to a phase-separation type formation of compartments[43]. Given that these compartments can be distinguished by their activity, this has sparked significant discussion in the field as to whether this level of organisation is a cause or effect transcriptional state[44-46].

### 1.3.3 Topologically associated domains

The A and B compartments can be further sub-divided into domains of self-interacting chromatin at the mega-base (Mb) scale, with distinct boundaries demarcated by CTCF-bound sites (Figure 1.2C) and active promoters. In the active A compartments, these domains are referred to as topologically associated domains (TADs); in the inactive B compartments these are referred to as lamina associated domains (LADs). TADs were first identified through next generation sequencing of chromosome conformation capture (3C) approaches[47-50], revealing a two- to three-fold enrichment of chromatin contacts within the TAD compared to outside regions[51]. Gene regulation has been shown to be restrained by TADs[48,52]: most often specific enhancers and their target genes reside within the same TAD[53], which acts to define the spatial limits within which regulatory elements such as enhancers can interact[32,54]. CTCF not only functions to simply demarcate the TAD boundaries, it also functions to insulate the TAD from interactions outside of the TAD structure.

Nested within TADs are sub-TADs, a finer-scaled architectural feature[55] that requires higher resolution techniques to detect them such as deep Hi-C[54], Tiled-C[56] or Region Capture Micro-C[57]. These are regions in which chromatin contact frequencies within the domain are even more enriched than in the larger parent TAD structure and are again

delimited by CTCF sites. To enable the interaction of specific regions of chromatin within TADs and sub-TADs, and for CTCF bound elements to act as boundaries to such structures, there needs to be a mechanism capable of bringing specific regions of chromatin together to interact within the limits of the CTCF sites. A protein of central importance in the regulation of this process is cohesin (Figure 1.2D).

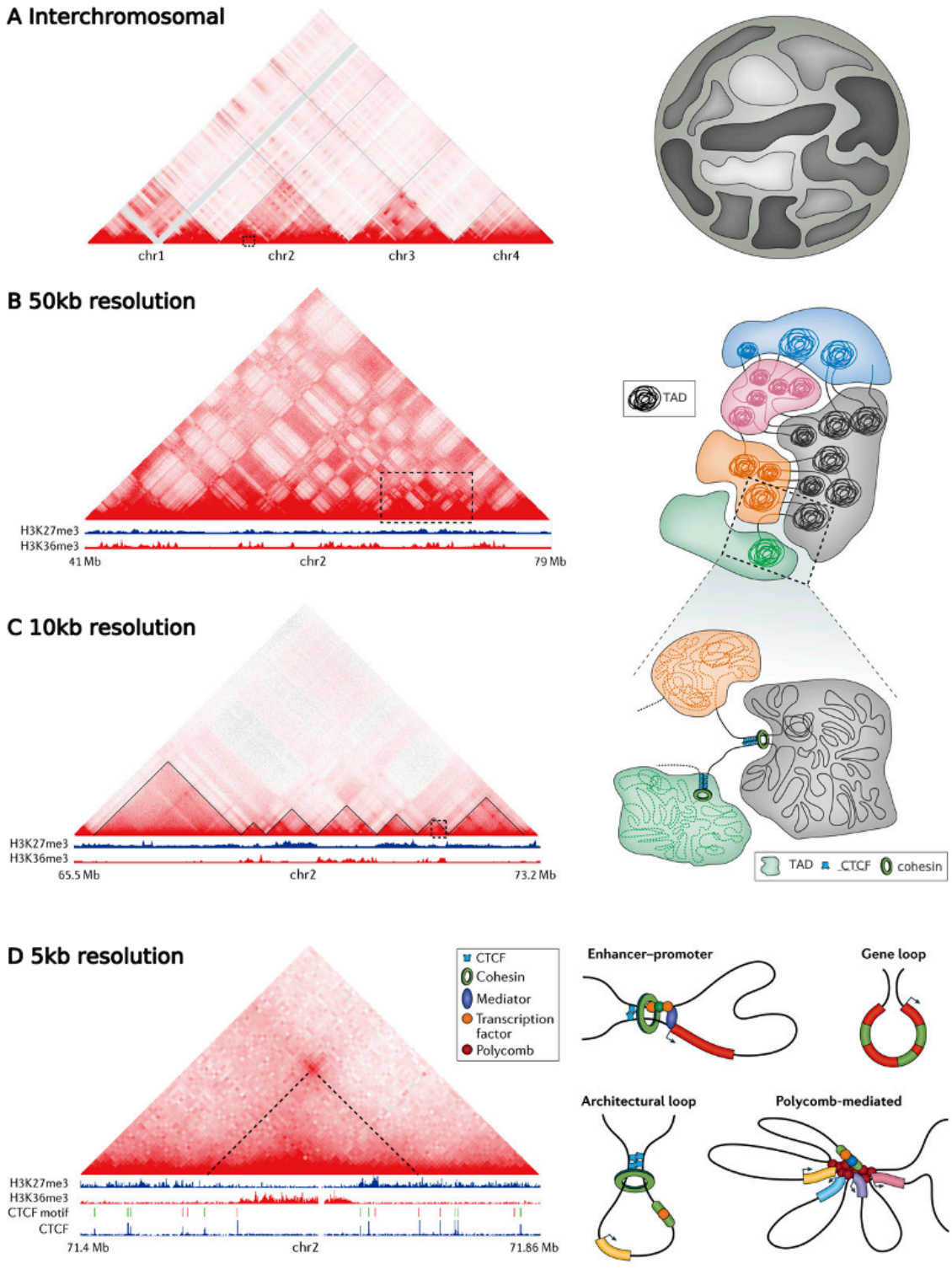



**Figure 1.2: The hierarchical organisation of interphase DNA**

Figure adapted from Review paper by Bonev and Cavalli[34]. Each level of organisation is represented by Hi-C data from GM12878 cells[58] on the left and a diagrammatic representation on the right. Dashed rectangles on the Hi-C data indicate the regions shown at higher magnification and increased resolution moving from panels d-a. **a** | Individual chromosomes occupy distinct territories within the nucleus (grey circle), and trans interactions are rare. **b** | Topological domains

with similar chromatin marks are characterised by stronger inter-domain interactions and are organised into active (A; blue/grey) and inactive (B; green/orange/red) compartments. **c** | A ~8 Mb region containing multiple topologically associating domains (TADs) which are annotated with solid lines. Diagram shows three different TADs: enriched for active marks (H3K4me3 and H3K36me3; grey), Polycomb (H3K27me3; green) or heterochromatin (H3K9me3; orange). **d** | 5kb resolution Hi-C enables the visualisation of loop domains. Beneath the Hi-C profile is the CTCF-binding profile and corresponding motif orientation (green represents forward and red represents reverse). Several different types of chromatin loop can potentially reside within a domain e.g., enhancer–promoter loop, Polycomb-mediated loop, gene loop or architectural loop.

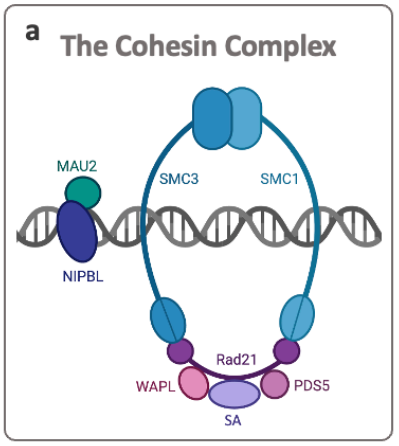
## 1.4 Cohesin drives the formation of 3D structures via loop extrusion

The mechanism of loop extrusion was introduced in this chapter as a mechanism to facilitate enhancer-promoter interactions. Whilst this is now a widely accepted hypothesis, the precise details of the timing and order of events during this dynamic process are yet to be determined[59]. I will next review the key proteins involved in this process before taking a closer look at the dynamics of the system.

### 1.4.1 Cohesin

The cohesin protein is a member of the Structural Maintenance of Chromosomes (SMC) family of protein complexes. It is a multi-subunit complex comprised of four core subunits that are present in a 1:1:1:1 stoichiometry[60]: two SMC proteins (SMC1 and SMC3), RAD21 and either of the stromal antigens STAG1/STAG2 [61]. The homologous ATPases, SMC1 and SMC3 dimerise through their hinge-domain and are linked by their head domain by the kleisin subunit, RAD21 [62-64] to form the characteristic ring-shaped complex as shown in Figure 1.3A. In the literature there are several aliases for the cohesin subunits that are used interchangeably, these are listed in the table in Figure 1.3B for clarity.

Cohesin is an essential and highly conserved regulatory protein complex which has two important and distinct functions: the first, and the one it was first known for, involves the tethering of sister chromatids; the second, and of central importance in this thesis, is in regulating 3D genome organisation by mediating DNA looping during interphase. The cohesin protein itself does not have a defined DNA binding motif and is instead loaded onto chromatin by the NIPBL-MAU2 loading complex which will be discussed later in further detail.




| b Subunit | Alias |
|---|---|
| RAD21 | SCC1 |
| STAG1 | SA1, SCC3A |
| STAG2 | SA2, SCC3B |
| SMC1 | SMC1a |
| SMC3 | - |
| NIPBL | SCC2 |
| MAU2 | SCC4 |
| WAPL | - |
| PDS5A/PDS5B | PDS5 |

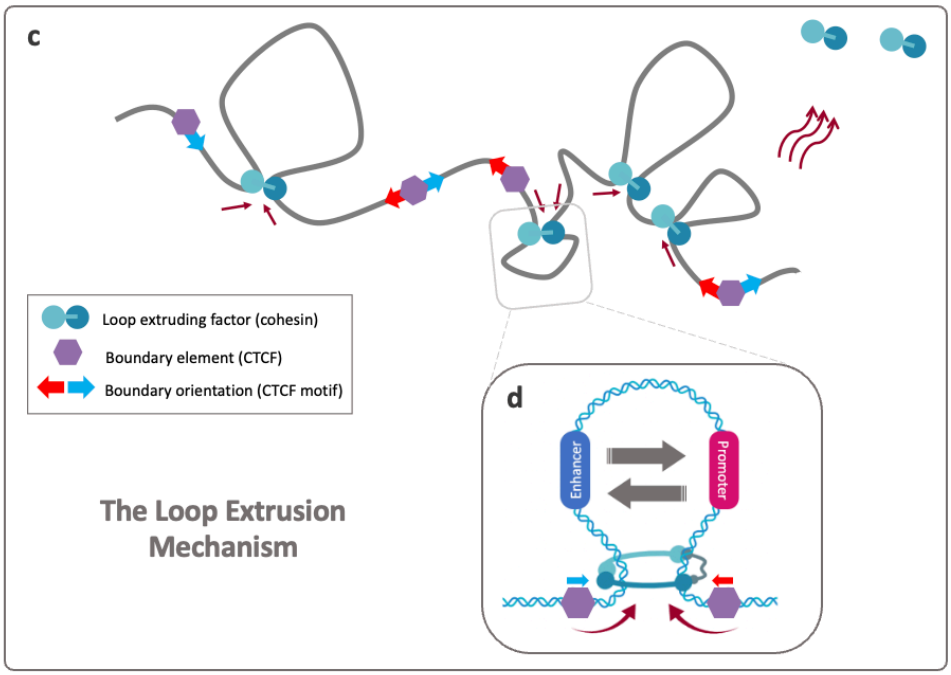



**Figure 1.3: The vertebrate cohesin complex and its role in loop extrusion.**

**a** | Schematic showing the subunits that make up the cohesin complex assembled on DNA, along with the cohesin loading complex (NIPBL/MAU2) and the unloading complex (WAPL/PDS5). Figure generated using BioRender. **b** | Table lists the subunits in this complex, along with the common

aliases seen in the literature. **c** | The loop extrusion mechanism. Along the strand of chromatin shown in grey, there are CTCF boundary elements (purple hexagons) which have specific directionality which is depicted by the red and blue arrows. Cohesin acts as a loop extruding factor by entrapping the chromatin and reeling it through the complex, forming a continuously extruding loop structure. This extrusion is halted when a boundary is met. Figure adapted from original loop extrusion paper by Fudenberg *et al.*[24]. **d** | Chromatin looping can bring regulatory elements, such as promoters and enhancers into close proximity, allowing them to interact with one another. Shown in an example of the CTCF convergence rule in which two convergently orientated CTCF motifs are found at the loop anchor. Figure created using BioRender

### 1.4.2 CTCF

The boundary elements which block the progression of cohesin along the chromatin commonly contain CCCTC-binding factor (CTCF) binding sites. CTCF is a highly conserved 11 zinc-finger protein. It acts as a transcription factor central to the establishment of 3D genome organisation, forming a physical block to cohesin by directly interacting with the RAD21 and STAG2 subunits of the cohesin complex[65]. Unlike cohesin, the binding sites of CTCF are well-defined and orientation dependent; it has also been shown that CTCF is only able to block the movement of cohesin via its N-terminal domain[66,67]. This has given rise to the 'CTCF convergence rule', which states that CTCF motifs orientated towards each other (see Figure 1.3D) are enriched at the boundaries of long-range interacting chromatin domains, where they prevent further translocation of cohesin [24,68,69] and thereby block further loop extrusion [70-72]. The binding of CTCF has been shown to be very similar between cell-types and yet 3D genome organisation varies between cell-types. This suggests that the map of TAD boundaries is pre-defined based on sequence, and that CTCF does not play a primary role in conferring cell-type specific gene regulation. Given cohesin's fundamental role in the formation of loops, it could be that sites at which cohesin is loaded dictate which boundary elements are utilised, and consequently which promoters and enhancers are topologically enabled to interact, therefore driving tissue-specific gene expression. In other words, whilst CTCF is largely

invariant, it could be utilised in a cell specific manner due to the cell specific loading sites for cohesin.

## 1.5 Dynamic properties: loading, translocation and offloading

Thus far, I have introduced the non-coding genome and the regulatory elements harboured within it, our current understanding of 3D organisation, and alluded to its importance in cell-type specific gene expression. However, to my knowledge, there is no convincing evidence to determine the order of events: does 3D-genome architecture arise independently and function to regulate cell-type specific gene expression, or is the 3D structure an emergent property of cell-type specific expression which is directed by the regulatory elements? Given that cohesin plays a key role in this process, understanding the dynamic regulation of this protein, i.e., how, and precisely where cohesin is loaded, off-loaded and moves along the chromatin, may provide important insights into this order of events.

### 1.5.1 NIPBL-MAU2, a cohesin loading complex

As previously mentioned, cohesin, unlike CTCF, does not directly bind to specific sequences in the genome. Instead, it has been shown that the loading of cohesin onto chromatin is regulated by the cohesin loading complex which is comprised of Nipped-B-like protein (NIPBL) bound at its N-terminus by MAU2 chromatid cohesion factor homolog (MAU2). Evidence suggests that MAU2 is required both to ensure the correct folding of the NIPBL N-terminus, and to enable the association of NIPBL to chromatin and direct the localisation of cohesin loading[73,74]. However, the localisation of cohesin loading by the NIPBL complex has caused much debate in the field, in general there have been two schools of thought. The first is that NIPBL is preferentially associated with enhancers, where it loads cohesin onto the chromatin [28,75-79]. In this way active enhancers act as

cohesin entry sites [80], one postulated mechanism for this is that tissue-specific transcription factors at the enhancer enable the recruitment of cohesin [81] and the binding of NIPBL at these sites is facilitated by the histone acetyltransferase, p300 [78]. An alternative is that NIPBL binds and recruits cohesin at active promoters [77,82]. One explanation for discrepancies in the data supporting these two ideas, is the lack of a universal classification for regulatory elements. It has been suggested that there is perhaps no clear-cut way to define enhancers and promoters, and they should instead be considered as a spectrum of regulatory elements[43]. The specificity of NIPBL antibodies has also been questioned, which could provide another explanation for the conflicting results[83]. Regardless of this debate, what these studies do agree on is that NIPBL is recruited to activated regulatory elements, which are themselves tissue specific. Given that the activation of regulatory elements is highly cell-type specific, this could provide a mechanism to explain how cohesin is recruited in a controlled manner to coordinate the formation of cell-type specific domains.

Other important functions of NIPBL that have been investigated in the literature include its ability to activate cohesin's ATPase domain[84-86], and its role in translocating cohesin along the chromatin[87-89]. Additionally, it has been suggested that NIPBL remains associated with cohesin after it has been deposited onto chromatin[90]. These additional functions suggest a role for the NIPBL in initiating and enabling cohesin to perform loop extrusion[91]. This could also provide a counter argument for cohesin being recruited to active regulatory elements. Given that NIPBL can track along the chromatin with cohesin, it could be the case that cohesin is not in fact directly recruited to active elements. Instead, this could simply be an artefact arising because current experimental techniques, such as ChIP-seq, do not allow for protein occupancy to be captured in a dynamic manner. It could be that active elements become enriched with cohesin after it has become associated with chromatin, and this is what is captured by ChIP-seq experiments.

### 1.5.2 Cohesin translocation

Cohesin dynamics can be divided into two branches: the loading and unloading by NIPBL and WAPL respectively, and the translocation by cohesin's innate ATPase activity[92]. The ATPase activity of cohesin enables it to act as a molecular motor, which can extrude DNA to form loops[76,93,94]. Cohesin is proposed to translocate along the DNA by a 'swing and clamp' mechanism[95], by which there is a spontaneous swing of the cohesin hinge, followed by clamping of cohesin to the DNA by NIPBL. Specifically, interactions between NIPBL and cohesin's SMC3 subunit initiates a conformational change which allows the transfer of DNA from the hinge domain to the SMC3 head, enabling the 50nm step, or 'swing'. This mechanism provides an explanation for how DNA can be progressively extruded to form loops and confirms the role of NIPBL in the translocation of cohesin. The ability of cohesin to translocate along the DNA has also been directly linked to factors such as acetylation levels, as shown in Figure 1.4. There is evidence that the acetylation of lysine residues in SMC3 by the acetyltransferase ESCO1, may hinder the ATPase activity of cohesin and subsequently reduce its ability to translocate along the DNA[86].

### 1.5.3 WAPL-PDS5, a cohesin unloading complex

The unloading of cohesin from chromatin is regulated by the complex of Wings apart-like protein homolog (WAPL) and Sister chromatid cohesion protein PDS5 homolog A (PDS5A)[96-98]. This complex is thought to enable the release of cohesin from the chromatin by opening up the cohesin ring at the SMC3/RAD21 connection point[99]. Structural studies suggest that the N-terminus of WAPL associates with PDS5A and binds to RAD21/STAG2, leaving the C-terminus to contact SMC1/SMC3 and mediate the dissociation of cohesin from the chromatin[100] (Figure 1.5). A key regulator of this process is the acetylation status of SMC3. When SMC3 is acetylated: PDS5A switches to bind Sororin, a protein involved in

sister chromatid cohesion, this leads to a displacement of WAPL which stabilises cohesive cohesin[101,102]. It has also been shown that PDS5 competes with NIPBL for cohesin binding and is favoured over NIPBL when SMC3 is acetylated[85,103]. In a recent publication, a 'PDS5A brake mechanism' was proposed for the control of chromatin loop length: levels of SMC3 acetylation determine whether NIPBL or PDS5A binds cohesin, therefore directly affecting the stability of cohesin on chromatin and subsequent loop length[104] (Figure 1.4).

Through loop extrusion, cohesin is thought to play a significant role in establishing regulatory domains to control tissue specific gene regulation. Whilst WAPL and NIPBL, are key regulators of cohesin dynamics, other factors such as acetylation have a significant impact on their functions.

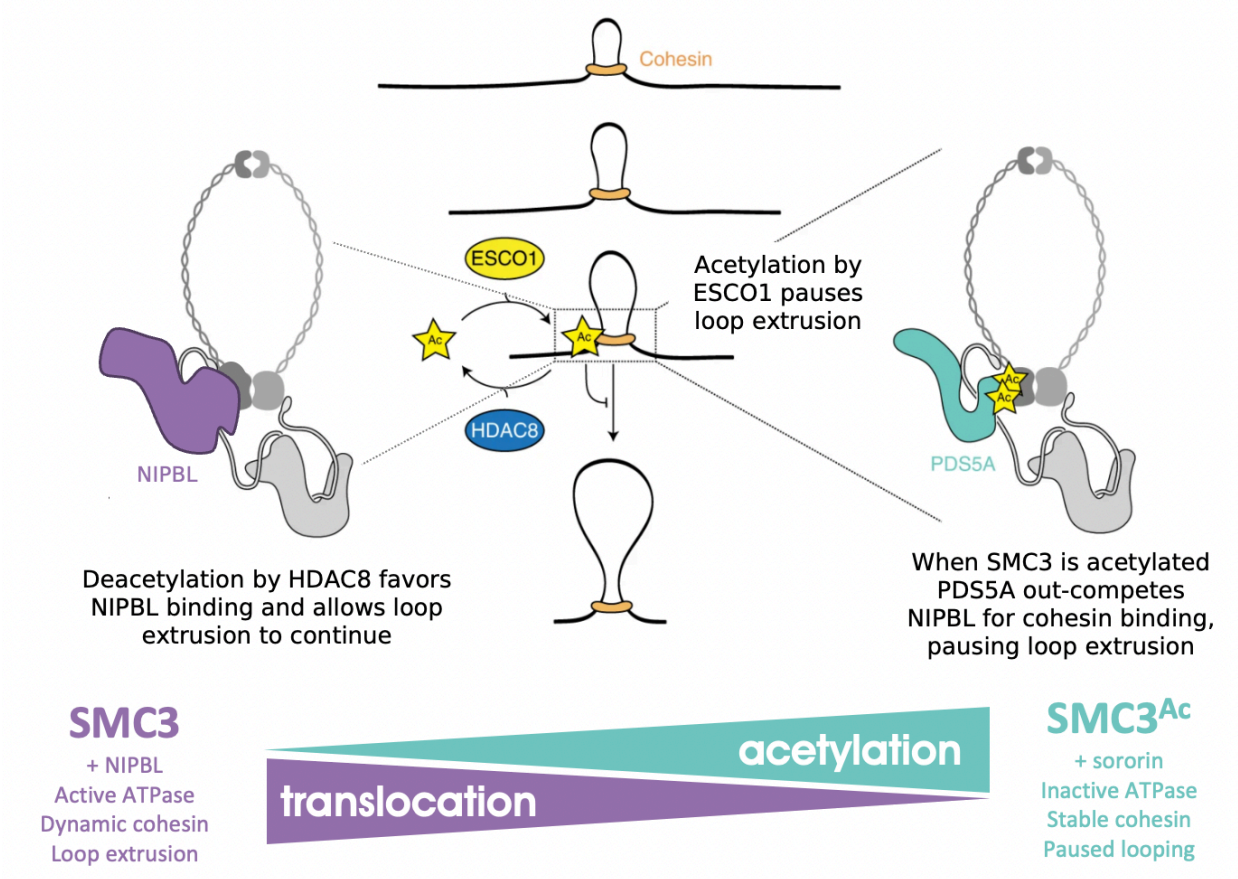



**Figure 1.4: Model of how chromatin loop length is controlled by the acetylation cycle of cohesin.** Data presented in a recent publication by van Ruiten and colleagues[104] supports a model in which cohesin mediated looping is controlled through SMC3 acetylation by ESCO1 and deacetylation by HDAC8. When cohesin is acetylated, PDS5A is thought to inactivate cohesin's ATPase, resulting in a pause in the looping process. Figure adapted from van Ruiten *et al.*[104].

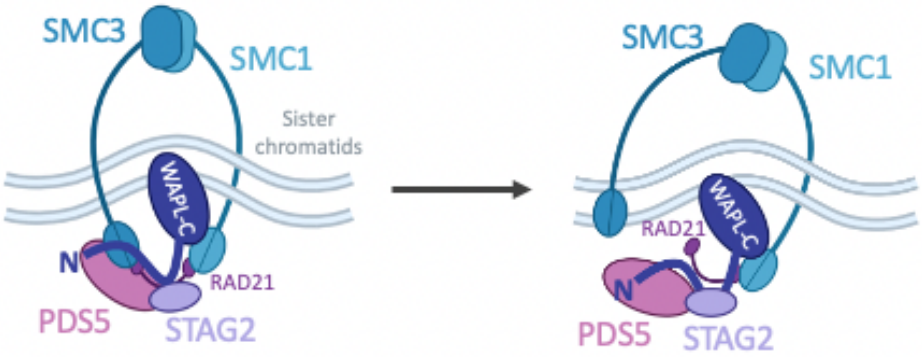



**Figure 1.5: Model for WAPL-dependent cohesin release from chromatin.**
The N-terminus of WAPL binds to PDS5 and RAD21-STAG2 via two interfaces, whereas the C-terminus has been proposed to contact SMC1-SMC3 or other associated proteins. It has been suggested that WAPL might be responsible for indirectly opening the cohesin ring at the SMC3-RAD21 junction by stimulating the ATPase activity of chromatin bound cohesin. Figure adapted from paper by Ouyang and colleagues[100] and created using BioRender.

## 1.6 Number crunching: timing cohesin dynamics

It is important to have an idea of the timescales by which various processes occur to understand which methods may have the appropriate temporal resolution to study cohesin dynamics. Whilst imaging techniques can capture rapid processes that take place at sub-second speed, genomic techniques that require time (minutes) for processes such as fixation will not have such fine-scale temporal resolution. Next, I will briefly review the current estimations for the timings of dynamic cohesin processes.

### 1.6.1 Residence times

There have been several studies that have investigated the residence time of cohesin on chromatin[105-110]. In work by Hansen and colleagues[110] they found that cohesin had a mean residence time of ~22 minutes, which is in agreement with previously published estimates[96,97,105]. In this study, they also showed that the search time, i.e., the time required for cohesin to re-associate with the chromatin is ~33 minutes, which they suggest

is reflective of what is deemed an optimal ratio of 3D diffusion and 1D sliding according to the theory of facilitated diffusion[111].

The residence time of cohesin has also been shown to differ depending on which STAG subunit it is associated with. Cohesin$^{STAG1}$ has been shown to preferentially associate with CTCF for TAD organisation[112,113] and compared to cohesin$^{STAG2}$, has a longer residence time, and is less abundant[108,109]. Cohesin$^{STAG2}$, on the other hand, is the most abundant form and is known to localise to enhancers and mediate short-range cell-type specific interactions[113,114]. This evidence suggests that interphase cohesin could be considered as two populations depending on the associated STAG protein: one that maintains TAD structure through the preferential interactions with the N-terminus of CTCF[112,113] (STAG1), and another that mediates cell-type specific looping (STAG2). Whilst it is known that STAG1 and STAG2 share 70% sequence homology in the central regions of the proteins[114], the N and C terminal domains only share 30-50% homology[115], which suggests that it is these domains which contribute to their specific functions.

The residence times of CTCF and NIPBL have also been investigated. In the same study by Hansen and colleagues as mentioned above, they showed that CTCF is more transient and dynamic than cohesin, with an average residence time of ~1-2 minutes and a search time of ~1 minute[110]. Describing CTCF as a boundary element creates connotations of stability when in fact the binding at cognate CTCF sites is relatively short-lived. They suggest that the highly dynamic occupancy at CTCF sites provides a potential mechanism for boundary permeability which could explain why competing loop domains are seen in their data. It could also explain why most loops assayed by DNA-FISH are only seen in a subset of cells at any given time[116,117]. Similarly short-lived occupancy has been demonstrated for NIPBL. This was seen as two distinct populations: a residence time of ~1 minute was suggested to apply to approximately half of the population of NIPBL which associates with bound

cohesin complex, whilst the other half binds stochastically on a sub-second time-scale[90]. The depletion of cohesin further supported this finding, showing increased mobility of NIPBL in the nucleus and a reduction in the chromatin bound fraction[90].

### 1.6.2 Extrusion rates

In vitro studies have shown that cohesin is able to extrude naked DNA loops at a rate of 1kbs$^{-1}$ [118], which with a mean residence time of 22 minutes would mean that if cohesin was unobstructed it has the ability to extrude 1.32Mb of DNA per binding cycle. However, the average loop length demarcated by CTCF, has been shown to be approximately 0.3Mb[24,119]. It is likely that the extrusion rate of chromatinised DNA will be slower given that the nucleosomes and other chromatin associated proteins will create some resistance to the translocating cohesin.

## 1.7 What can degron models tell us about the role of interphase cohesin?

There are a number of elegant ways to study protein function and their role in gene regulation, such as using techniques to prevent or reduce the expression of a protein by genetic knock-out or degradation of mRNA using RNAi [120]. However, when the protein plays an instrumental role in essential cellular processes such as cell division and the cell cycle, a complete knock out of the gene is lethal. Additionally, there is a significant lag-time before gene function is abolished [121], making these techniques impractical for studying dynamic systems. It has also been shown that these techniques often do not result in a homogenous and synchronous depletion across the cell population [122], which can be problematic in cases where the downstream analysis relies on population-based assays. For this reason, a number of inducible methods have been developed that can rapidly and reversibly degrade or sequester the target protein to a different sub-

compartment of the cell, thus allowing the short-term study of cellular functions upon the abrupt absence of the protein.

Given that the cycle of cohesin loading and unloading is a dynamic process, inducible systems lend themselves nicely to the study of proteins involved in this process. Each step of this loading, translocation and unloading cycle has been extensively studied, both in the normal, unperturbed state and using depletion methods to investigate their role in the regulation of 3D architecture and gene expression. In the next section, I will briefly discuss the current degron technologies and a few of the significant studies that have used these techniques to investigate the cohesin and the regulation of genome organisation.

### 1.7.1 Degron technologies

The most common inducible degradation techniques rely on molecules (also known as 'degrons') that can bridge the protein of interest (POI) with an E3 ligase, consequently targeting the POI for degradation by the proteasome. Until recently, one of the most widely known and utilised degron systems was the auxin inducible degron (AID)[123]. This is an example of a three-part system, comprised of: (1) an exogenous E3 ligase: OsTir1 (*Oryza sativa* Tir1), (2) a degron tag: *Arabidopsis thaliana* IAA7 auxin-inducible-degron (AID) (3) an inducer molecule: Auxin. However, this method is not without limitations: namely leaky degradation, the high, and slightly toxic doses of auxin required for degradation [124], and the complex genome engineering strategy required to successfully introduce both the OsTir1 and AID subunits into the target cell line. Since the development of the AID system, there have been numerous developments and improvements to the system, notably those reducing the number of components required and therefore decreasing the amount of genetic engineering required. For example, the two-part FKBP12$^{F36V}$/dTAG system, designed to circumvent the use of an exogenous E3 ligase by making use of either of the endogenous E3 ligases: Cereblon (CRBN) or von Hippel–Lindau (VHL) [125]. Another

degron system that has gained more attention in recent years are the heterobifunctional small molecules known as Proteolysis-targeting chimeras (PROTACs). These are a one-part degron system, the PROTAC molecule is comprised of two ligands joined by a linker: the endogenous POI is bound by one ligand and the endogenous E3 ligase to the other. Whilst PROTACs are an elegant approach for targeted and inducible protein degradation that requires no genome engineering, PROTAC design comes with the familiar challenges faced in medicinal chemistry when designing small molecule drug targets. Thus whilst the E3 ligase binding ligand can be reused, the POI has to be 'PROTAC-able' and consequently most of the PROTACs developed to date are for POIs that have existing ligands[126]. To my knowledge, there have been no PROTACs designed to target any of the cohesin subunits. Thus, whilst this would be an exciting strategy to explore in future work, the AID[127-130] and dTAG[131] methods are the two most popular degron technologies that are currently used in the field. I will next discuss some of the key studies that have used degrons to investigate the role of cohesin in genome organisation.

### 1.7.2 Cohesin degron models

Results across studies have tended to agree that the loss of cohesin results in a general decompaction of chromatin, as shown by analysis of both fixed[132-134] and live-cell imaging[135,136]. In addition, cohesin degradation seems to strengthen A/B compartmentalisation and increase polycomb interactions across TAD boundaries[130]. This suggests that cohesin is not required for very long-range interactions forming compartments, and may play a role in disrupting long-range polycomb dependent chromosome interactions via loop extrusion[137]. These studies have also shown that the preferential interactions between convergent CTCF sites no longer holds true in the absence of cohesin, further supporting the idea that CTCF directionality is an essential component of cohesin mediated looping. As a result, it is not surprising that TAD structure and loop domains disappear in Hi-C data following acute depletion of

cohesin[127,128]. However, despite the disruption of TAD structures, which define the limits and enhancer action and gene regulation[32,54,138], enhancer-promoter interactions are minimally affected following loss of cohesin[135]. Whilst at first this seems surprising; this could simply be explained by the fact that most enhancers and promoters are located close to one another and therefore do not need loop extrusion to mediate their interaction, although this would raise the question of why this system evolved in the first place. Alternatively, this could be due to an incomplete degradation of cohesin, which could leave a proportion of the enhancer-promoter interactions intact. It could be the case that different segments of the genome have different affinities for cohesin, perhaps due to the presence of proteins that associate with cohesin and are therefore able to retain cohesin for longer during degradation. A method capable of tracing dynamic cohesin loss throughout the degradation would be insightful, as this could determine if cohesin loss is ubiquitous, or whether the behaviour of cohesin differs from one site to another.

### 1.7.3 CTCF degron models

Given that CTCF has a well-defined, directional binding site, they lend themselves nicely to genetic perturbation studies. Targeted deletion, inversion, or silencing can be used to investigate the role of CTCF in a specific regulatory context[48,70,116,139]. Studies have shown that ablation of individual CTCF sites using methods such as these can have severe pathological implications in processes such as cancer progression[140] and limb development[141]. Genome-wide loss of CTCF has little effect on compartmentalisation, whilst TAD boundaries become less distinct, and most loops disappear[127,142]. It has also been demonstrated that genome-wide acute depletion of CTCF changes the pattern of chromatin accessibility across the genome[142]. These findings are consistent with the loop extrusion hypothesis: the removal of CTCF, the protein that demarcates the TAD boundaries and forms the loop anchors, allows cohesin to continue unimpeded to extrude DNA which results in less distinct TAD boundaries. Given the significant effect of targeted

CTCF loss, it could be expected that genome-wide loss of CTCF would have a significant widespread impact on gene expression. Conversely, in the study by Hyle and colleagues[143] they state that following acute depletion of CTCF in a human paediatric B-ALL cell line (SEM), only minimal effect on gene expression is observed. To support this statement, they show 330 differentially expressed genes between WT and CTCF AID depleted cells. However, supplementary data from this study suggests that CTCF was incompletely degraded, which perhaps explains this surprising result.

### 1.7.4 WAPL degron models

Upon WAPL depletion, it has been shown that cohesin is loaded, but not released and therefore cohesin turn-over is ablated[96,97]. This reduces the free pool of cohesin that is available for re-loading by WAPL, creating a genome-wide state of cohesin stabilisation and increases chromatin loop size. In a recent study by Liu and colleagues[137], they showed that the loss of WAPL resulted in a global redistribution of cohesin from cell-type specific transcription factor binding sites to CTCF binding sites and a consequent reduction in cell-type specific gene expression, which was reversible upon indole-3-acetic acid (IAA) wash-out. Through Hi-C analysis they showed that the depletion of WAPL did not affect the number and size of the TADs, however they did report an increase in inter-TAD (1-10 Mb) contacts, and a decrease in intra-TAD (< 1 Mb) contacts. WAPL degradation models are somewhat analogous to cohesin depletion models, in that by retaining cohesin on the chromatin, the pool of free cohesin that would ordinarily be maintained by off-loading is depleted. The finding that cohesin accumulates at CTCF sites following WAPL depletion, indicates that cohesin translocates from its loading site until it reaches these boundary elements where it is usually off-loaded, but instead accumulates. The absence of cohesin at cell-type specific TF binding sites could be explained by the lack of free cohesin available to either load at these sites, or to load and then translocate via these sites before accumulating at the CTCF boundaries.

### 1.7.5 NIPBL degron models

In a similar vein to WAPL ablation, when NIPBL is depleted, cohesin can no longer be loaded onto chromatin. In this case however, there is an increase in the pool of unbound cohesin, and a decrease in chromatin associated cohesin. This has been shown to cause a genome-wide loss of chromatin loops and TADs[144,145]. Using RNAi, Luppino and colleagues[91] showed that a degradation of NIPBL alters approximately 90% of TAD structures, similar to that seen following cohesin depletion.

## 1.8 Limitations of existing approaches and the need for dynamics in the analysis of chromatin associated proteins

At which sites within chromatin is cohesin first loaded? This is a question that requires a method capable of tracking the dynamic behaviour of cohesin on a genome-wide scale. In the next section I will briefly discuss a few limitations and caveats of the existing methods to set the scene for how a new approach could help to answer this question.

### 1.8.1 Current genomic methods do not capture dynamic processes

Whilst single-cell approaches are becoming more wide-spread, most genomic methods including 3C techniques and ChIP-seq are ensemble approaches and report the average protein occupancy from the analysis of millions of nuclei. These techniques have been pivotal in developing our current understanding of the 3D genome, but they are not informative for dynamic processes. These techniques take a 'snapshot' of the genomic landscape which represents an average across all cells, which in systems testing the effect of perturbing specific elements or proteins, are only usually taken before and after the perturbation. Therefore, the resulting data only provide evidence for stable ensemble

status at the start and end of a process, not the path by which the end result was achieved. Single-cell techniques have the ability to deconvolve the population average effect, which can potentially mask significant but infrequent trends, but again these still only achieve a snapshot of the overall process. High-resolution imaging techniques can provide a solution, allowing the tracking of events in single live cells in real-time. However, these imaging techniques are limited in their throughput, and at present do not have the ability for genome-wide scale-up.

### 1.8.2 Fixation methods may not capture protein occupancy at transient sites

ChIP-seq and 3C methods alike, rely on chemical fixation to 'freeze' the chromatin environment, it is then assumed that the protein occupancy or 3D interactions measured should reflect the status at the time of fixation. For proteins such as cohesin, positions on the chromatin such as the loading sites may only be transiently occupied and are therefore less likely to be fixed than regions such as CTCF sites where cohesin is often stalled. However, if cohesin is continuously loaded at a particular site, even if the amount of time each individual cohesin spends at the site is minimal, the amount of time in which the site is occupied by cohesin will be the above background average and so detectable. An important technical caveat to consider is that chemical fixation may not be instantaneous, using methods that rely on this to take a 'snapshot' might more likely capture a window, rather than a precise point in time.

### 1.8.3 Incomplete degradation may invalidate conclusions

In many of the degradation studies reviewed in the previous section, there is no available, or convincing evidence to confirm that the entire population of protein was depleted. In my work, I have found that a protein may appear to be degraded by western blot yet can still show significant levels bound to the chromatin by ChIP-seq. Even if a small population

of the protein remains, this significantly alters the validity of any conclusions drawn: a population of protein resistant to degradation, or that degrades at a slower rate may perhaps be enriched at sites where the protein is the most stably bound or has the longest residence time. Furthermore, these sites may correlate with specific architectural structures such as TAD boundaries; to draw a conclusion that the structures are independent of the (incompletely) degraded protein would therefore be incorrect.

## 1.9 Thesis aims

Earlier I introduced the conundrum of whether genome architecture is reflective or instructive of the regulation of gene expression. One of the questions that remains to be answered in this field is at which sites is cohesin first loaded onto chromatin. At present, the most popular hypothesis is that cohesin is recruited to active enhancers, yet this remains to be experimentally confirmed. If this is the case, one might argue that the activation of regulatory elements pre-determines gene expression patterns, thus the recruitment of cohesin to already active elements may suggest that the subsequent 3D organisation orchestrated by cohesin is reflective of gene expression. However, it could be argued that if cohesin was not recruited to these sites, the regulation of gene expression would be disrupted and therefore it also plays a role in instructing gene regulation. Considering this from an evolutionary perspective, it could be the case that the distribution of enhancers in the genome has mandated a requirement to recruit cohesin to enable them to interact with one another.

This thesis aims to address this long-standing question by interrogating cohesin dynamics. To preliminarily test the hypothesis that active elements direct cohesin recruitment, in Chapter 2 I use an artificial system developed in the lab to determine whether cohesin can be recruited to an isolated active regulatory element in a neutral region of the genome. This model however is not a dynamic one and will therefore be unable to determine whether cohesin is directly loaded at the active element itself. The remainder of the thesis is consequently dedicated to the development of novel inducible systems, with an aim to directly measure the distribution and rate of change of cohesin across the genome by either a pulse-chase or degradation pathway. From the genomic profiles captured throughout the time-course of these dynamic processes, it can be inferred whether trends across sites in these profiles segregate with specific regulatory

elements. The overall aim of this thesis is to determine whether cohesin dynamics differ across sites in the regulatory genome. Whilst the focus in this thesis is specifically on cohesin, the need for understanding the dynamics of chromatin proteins is a general one. Therefore, the techniques developed in this thesis were designed in a way that would allow them to be easily adapted and generally deployed for other key components of the regulatory genome.

# 2 Cohesin recruitment at a single locus

## 2.1 Introduction

It is known that enhancer-promoter interactions are important for eliciting cell-type specific transcriptional regulation; one of the most widely accepted mechanisms for how these interactions are orchestrated is loop extrusion. Here, cohesin plays a central role in mediating the chromatin looping that brings the regulatory elements into close enough proximity to interact. Disentangling the involvement of individual proteins and elements in the formation of regulatory domains is often challenging given the complex and tightly linked relationships between them. In this chapter I make use of a system that was previously developed in our lab, to investigate the capability of a single erythroid-specific regulatory element to recruit cohesin and form a 3D domain when isolated away from its native genomic context within the alpha-globin locus.

### 2.1.1 The alpha-globin locus

One of the best characterised loci where the regulatory landscape has been extensively studied for many years as a model system for gene regulation, is the mouse alpha-globin locus[146]. This relatively small 200kb region contains a cluster of regulatory elements that

are specifically activated in erythroid cells, including the five erythroid-specific enhancers: R1, R2, R3, R4 and Rm. The beauty of this locus is the ability to use it to study how 3D architecture changes throughout lineage differentiation: from an inactive state in the pluripotent embryonic stem cells (ESC), into an increasingly active and ordered domain as alpha-globin gene expression begins in the terminal stages of erythroid differentiation. Figure 2.1 demonstrates the stark contrast in genomic signatures when the alpha-globin locus is inactive (undifferentiated mESCs) versus active (terminally differentiated erythroid cells). For example, in erythroid cells the R2 enhancer (highlighted in turquoise in Figure 2.1) displays the characteristic signature of an active enhancer (H3K4me1, H3K27Ac and open chromatin); whereas in an inactive state in mESCs, this element is devoid of these marks. Interestingly, cohesin is also present at these elements when the locus is active. Given that changes in genomic signatures coincide with the upregulation of alpha-globin gene expression, it is thought that they play a significant role in regulating this process.

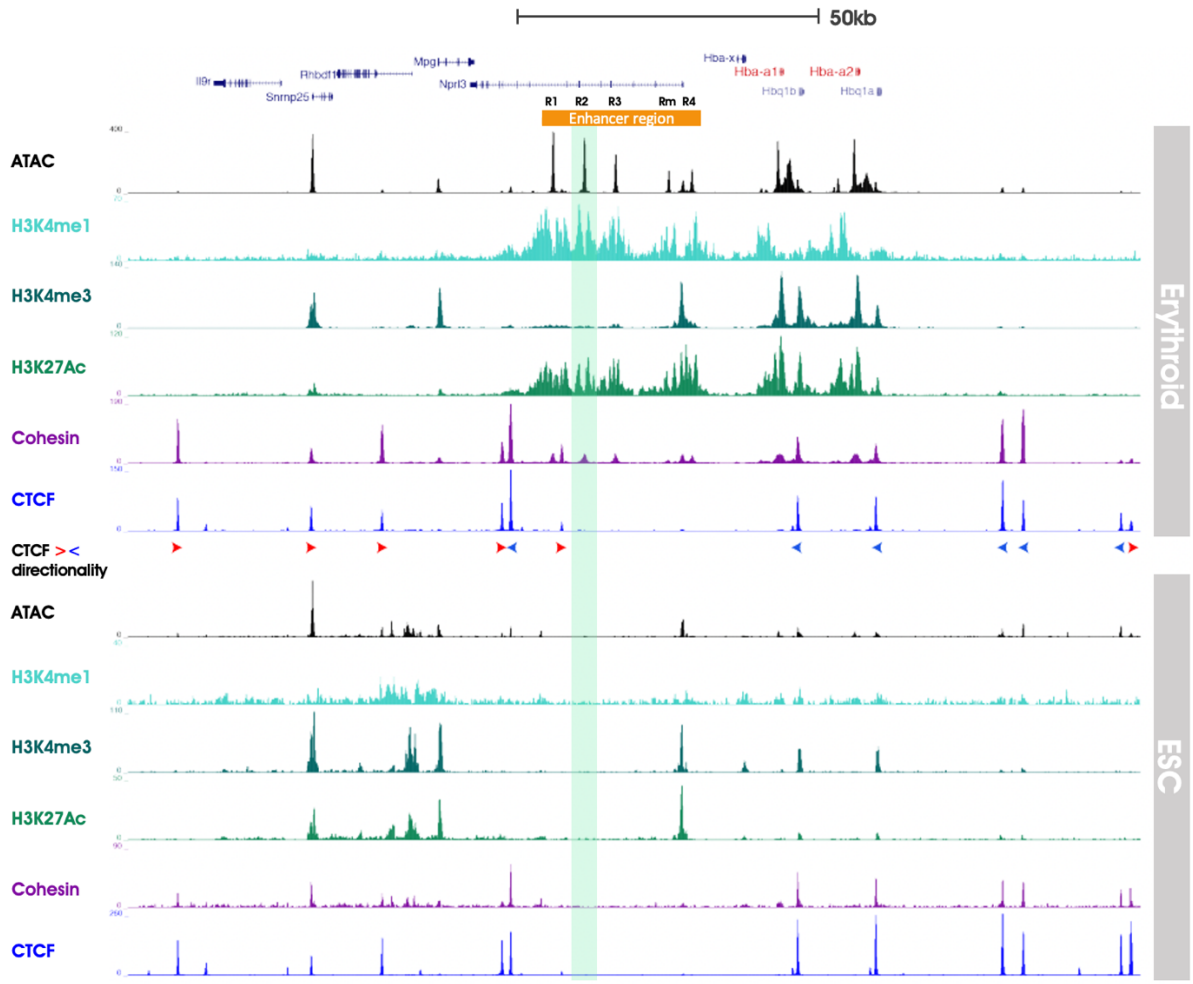



**Figure 2.1: Comparison of the genomic profiles of the mouse alpha-globin locus in WT primary erythroid cells versus embryonic stem cells.**

From top to bottom: chromatin accessibility profile (ATAC-seq), ChIP-seq of histone modifications (H3K4me1, H3K4me3, H3K27Ac), cohesin ChIP-seq (Rad21 subunit in erythroid cells, Smc1 subunit in ESCs), CTCF ChIP-seq, red and blue arrows indicate CTCF motif orientation (red = forward; blue = reverse). The enhancer region is highlighted with an orange bar and individual enhancer elements are labelled R1, R2, R3, R4 and Rm. Gene annotation are from Refseq with the alpha-globin genes highlighted in red. Region highlighted in turquoise is the region containing the R2 enhancer. Profiles shown are normalised (RPKM). Figure from Caz Harrold's thesis, where full disclosure of data source and analysis can be found [147].

### 2.1.2 Methods to study 3D genome architecture

In recent years there have been significant advancements in our understanding of the 3D genome, driven by the expansion of techniques available to study the interaction landscape of the genome. These techniques are designed to detect the spatial proximity of DNA sequences that are physically close together, but not necessarily nearby in the DNA sequence itself. This family of techniques is referred to as 'Chromosomal Conformation Capture', or '3C', and is achieved through a process of biochemical crosslinking, followed by fragmentation and re-ligation. Segments of DNA that are commonly in close proximity, are more frequently ligated together therefore the read-out of these techniques is a report of the frequency of "contacts" or "interactions", between regions of the DNA (Figure 2.2A). 3C methods can be further subdivided into categories depending on the interactions they assay: methods such as 3C[148] and 4C[149,150] assay interaction frequencies at a single locus (one-to-many); Capture-C[151,152], 5C[153], Tiled-C[56] or Micro-Capture-C (MCC)[28] assay a set of loci (many-to-all); and finally, Hi-C which assays interactions on a genome-wide scale (all-to-all)[154]. As these techniques have become more advanced, many now rely on high-throughput sequencing technologies that have been revolutionary in enabling unprecedented discoveries within the field of genome architecture. The two predominant techniques for which data are presented in this thesis are Capture-C, and the Capture-C based method, Tiled-C. These two techniques assay the interaction frequencies from a particular viewpoint, and at viewpoints tiled across a region of interest respectively. A summary of the data representation of these techniques is shown in Figure 2.2B.

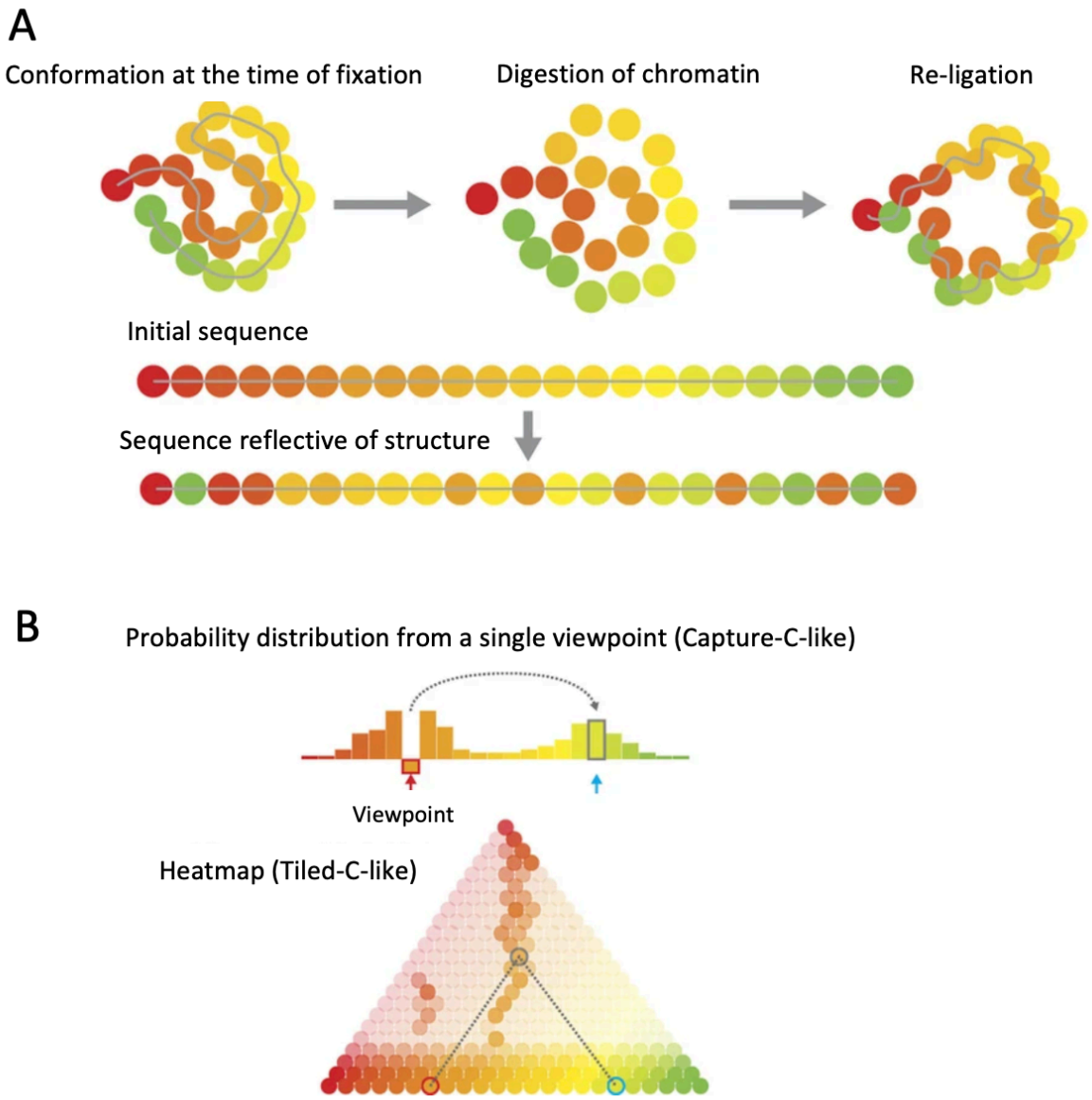



**Figure 2.2: Overview of the 3C-based techniques.**

**a** | The first step is to fix the chromatin using a chemical fixative, and then to digest the chromatin fibre into short restriction fragments (shown as circles). These fragments are then ligated back together to create a re-shuffled version of the chromatin fibre in which the order of the fragments reflects the 3D structure at the time of fixation. **b** | The contact frequencies determined from the 3C-based methods can be represented as a probability distribution, or interaction profile from a single viewpoint (one-to-many). Contact data generated by approaches such as Tiled-C or Hi-C (many-to-many and all-to-all) are typically represented in 2D-heatmaps, in which the colour intensity represents the contact frequencies between two loci. Figure adapted from Review by Davies *et al.*[155].

### 2.1.3 Evolution of the 3D genome landscape during erythroid differentiation

3C technologies have been instrumental in developing the understanding of gene regulation beyond one-dimensional chromatin modifications and protein occupancy, to determine the 3D interactions among specific elements within a locus, or on a genome-wide scale. A recent paper published by our group [56] has shown for the first time how the 3D architecture evolves throughout the stages of *in vivo* erythroid differentiation, at high resolution using Tiled Capture-C (Figure 2.3). This technique captures interactions at the resolution of Capture-C[151,156], but with the all-vs-all interaction information that can be yielded from Hi-C[42] by tiling capture probes over the region of interest, enabling the complex interactions between multiple regulatory elements within the region to be interrogated simultaneously. The added advantage of this technique is that it can be carried out using relatively low cell numbers (as few as 2000 cells), making it a feasible technique to assay low abundance primary cell types which would not have been possible using existing techniques[56].

Figure 2.3 demonstrates an example of how the 3D genome landscape changes as cells progress through erythroid differentiation from stages S0-low to S3[56]. As the alpha-globin locus becomes increasingly activated, there in an increase in open chromatin signal (assayed by ATAC-seq), which is mirrored by an increase in 3D interactions (assayed by Tiled-C) within this region to form a sub-TAD structure. Meanwhile the occupancy of the boundary elements by CTCF in this region remains stable, a phenomenon that has previously been described[157-159]. In the original publication[56] it is suggested that the formation of the sub-TAD is likely driven by cohesin-mediated loop extrusion[116,160].

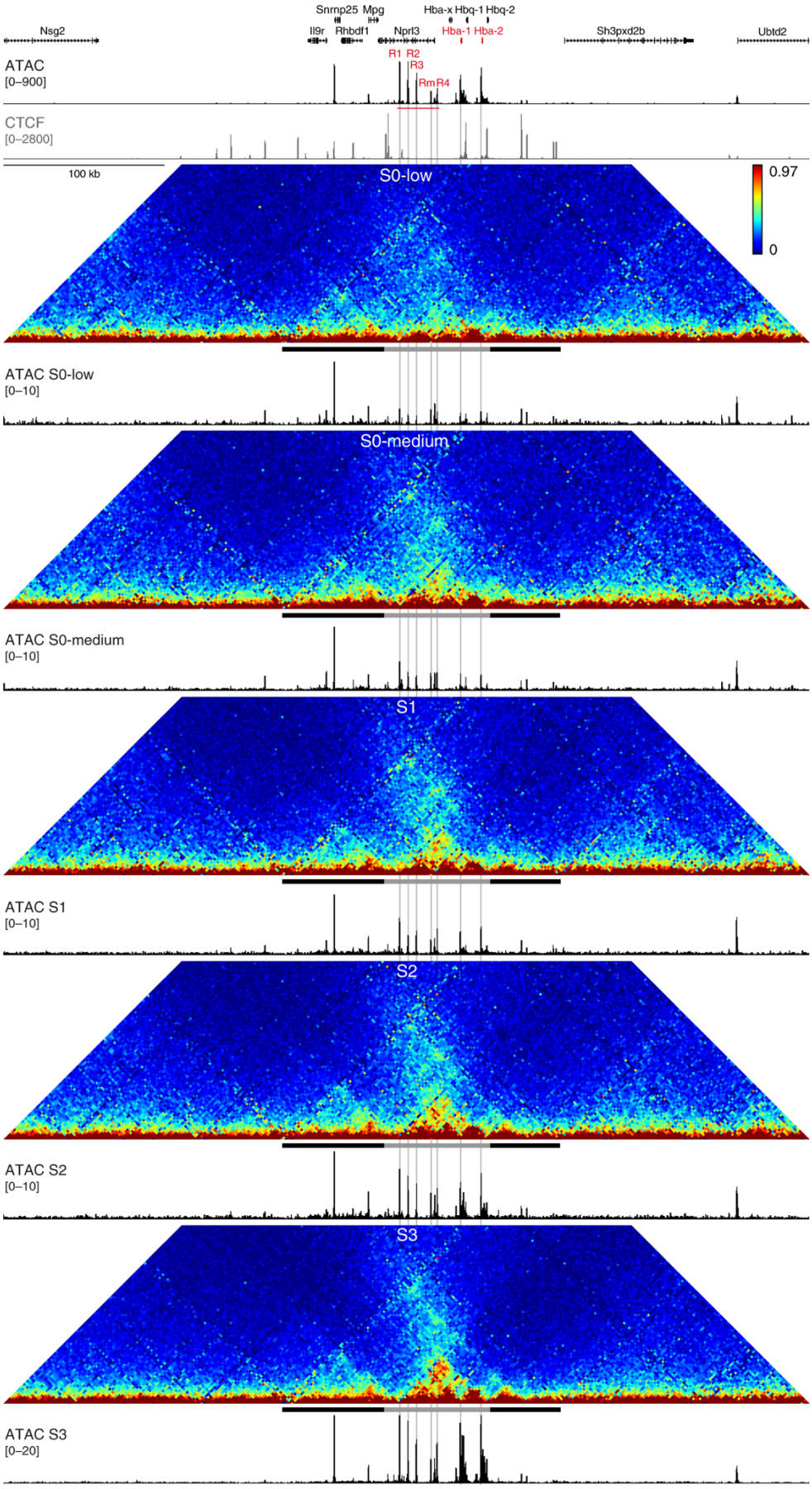

Snrnp25 Mpg
Hba-x Hbq-1 Hbq-2
Nsg2
Il9r Rhbdf1
Nprl3
Hba-1 Hba-2
Sh3pxd2b
Ubtd2
R1 R2
R3
Rm R4
ATAC
[0–900]
CTCF
[0–2800]
100 kb
S0-low
0.97
0
ATAC S0-low
[0–10]
S0-medium
ATAC S0-medium
[0–10]
S1
ATAC S1
[0–10]
S2
ATAC S2
[0–10]
S3
ATAC S3
[0–20]

**Figure 2.3: Tiled Capture-C throughout differentiation.**
As the alpha-globin locus becomes increasingly active throughout the stages of erythroid differentiation, there is an increase in chromatin accessibility and an increase in enhancer-promoter interactions within the alpha-globin locus. 500kb Tiled-C contact matrices, with corresponding ATAC profiles covering the mouse alpha-globin locus throughout the stages of erythroid differentiation at 2 kb resolution. Coordinates: chr11:31,900,000–32,400,000 (mm9). Figure adapted from Oudelaar and Beagrie[56], full details available in original publication.

### 2.1.4 A single erythroid-specific enhancer can direct the formation of a regulatory domain

Another intriguing finding from our group came from the thesis work of Caz Harrold[147], in which she showed that a single, isolated active enhancer was sufficient to direct the formation of an interaction domain (Figure 2.4). For this work, a 325bp sequence containing the erythroid-specific alpha-globin R2 enhancer element was inserted into a neutral region in chromosome X and an *in vitro* erythroid differentiation system was used to obtain erythroid cells in which the enhancer was active[161] (characterisation of this locus is provided in Supplementary Materials, Figure S1). The isolation of the R2 enhancer element in a neutral region of the genome removes any confounding effects produced by the *in-situ* alpha-globin environment, such as the influence of other regulatory elements or proteins. Therefore, this model provides an excellent platform for testing the hypothesis that cohesin-mediated loop extrusion drives the formation of the erythroid specific sub-TAD seen in the Tiled-C data. If cohesin is seen to be recruited at this locus following activation, this would provide convincing evidence that it is involved in *de novo* domain formation.

I hypothesise that the sub-TAD structure that is formed progressively throughout erythroid differentiation is a result of the specific activation of the erythroid enhancer elements (the mechanism of which is currently unknown), which enables tissue specific transcription factor binding. Cohesin is then recruited to these transcription factor hubs

at the enhancers and loop extrusion is initiated, which functions to regulate cell-type specific transcription. A diagrammatic summary of this hypothesis is shown in Figure 2.5. With this hypothesis in mind, there are two overarching goals of this chapter:

(1) To further define the 3D structure detected by Capture-C to determine if a sub-TAD structure is formed in the presence of the active R2 element.

(2) To investigate cohesin recruitment to the locus upon activation of the R2 element, to determine if the formation of the domain is linked to recruitment of cohesin and is therefore likely a product of loop extrusion,

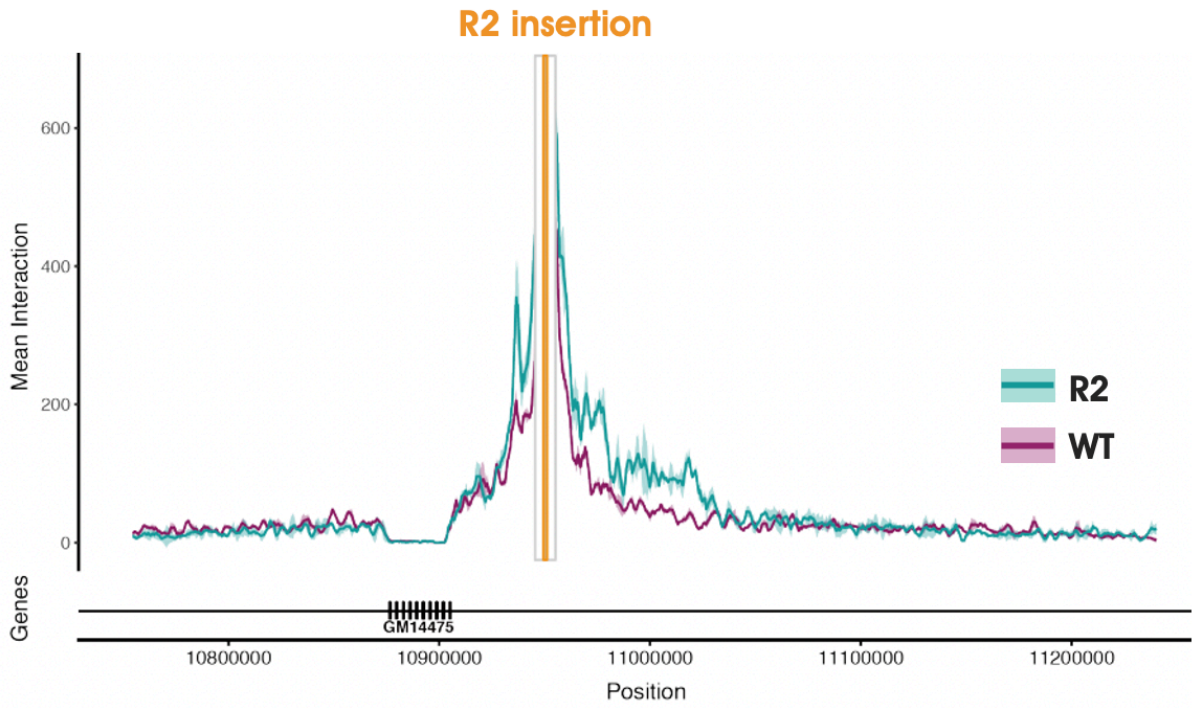



**Figure 2.4: Comparison of the wild-type (WT/chr11-R2) and R2 enhancer model (chrX-R2).** NG Capture-C from the R2 viewpoint (R2 insertion), with a 1 kb exclusion zone around the viewpoint. Comparison of interaction profiles from CD71+ erythroid cells containing the R2 insertion (magenta), and WT CD71+ erythroid cells (turquoise). Interaction profiles represent normalised unique interactions, averaged from three biological replicates/independently targeted clones. Standard deviation of a sliding 3 kb window is shown as a halo around the mean. Genes and genomic position below interaction profiles. Figures adapted from Caz Harrold's thesis [147].

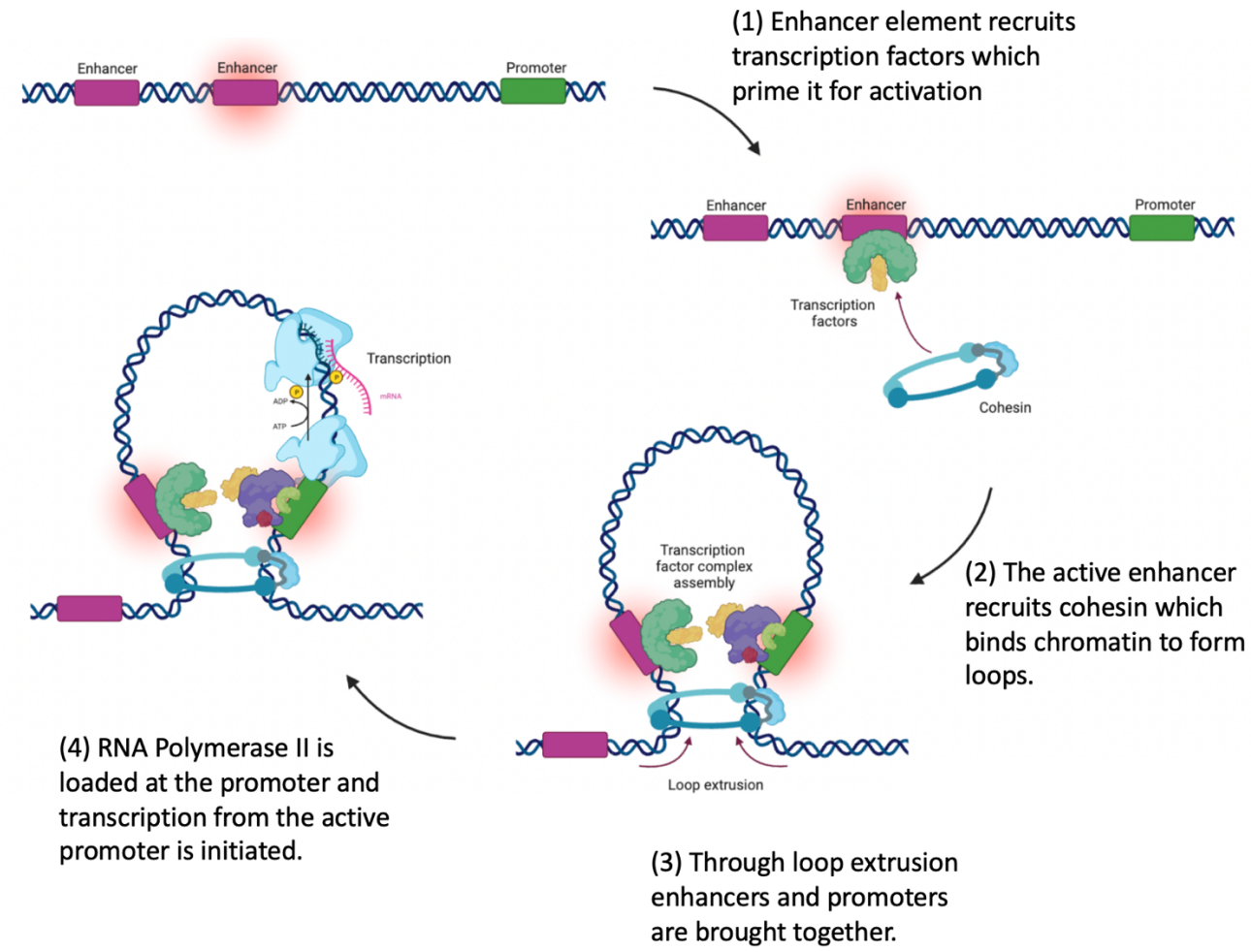



**Figure 2.5: Hypothetical mechanism of cohesin recruitment to active enhancers.**
Enhancer elements are activated in a cell-type specific manner. These active enhancer elements are then able to recruit tissue specific transcription factors, which themselves recruit the cohesin complex. The recruitment of cohesin to the chromatin initiates loop extrusion, which brings the active enhancer and promoter elements into close proximity, allowing transcription to occur. Figure produced using BioRender.

# 2.2 Results

### 2.2.1 High resolution Tiled-C data reveals sub-TAD-like structure

The NG Capture-C data presented in the chapter introduction shows that an isolated enhancer, when activated, can initiate an increase in 3D interactions. Whilst this is suggestive of domain formation, this technique only assays one-way interactions from the R2 viewpoint and cannot be used to make inferences about higher-order structures and configurations such as TADs. I wanted to investigate the 3D interactions across the whole of this locus in an unbiased manner using a many-vs-all approach to determine whether the interaction profile across this region was representative of a TAD-like structure. To achieve this, Tiled-C was performed across a ~3.3Mb locus spanning the insertion site in cells containing the R2 insertion (Figure 2.6) and in WT cells (Figure 2.7). In agreement with the NG Capture-C data, Tiled-C data revealed an increase in the interaction frequencies of the R2 insertion site with the surrounding locus compared to WT, and was reflective of a TAD-like structure. As was the case in the erythroid differentiation data shown in Figure 2.3, there was also an increase in open chromatin sites compared to WT within this domain (assayed by ATAC-seq, Figure 2.6-2.7). I next wanted to test whether cohesin occupancy correlated with this domain formation.

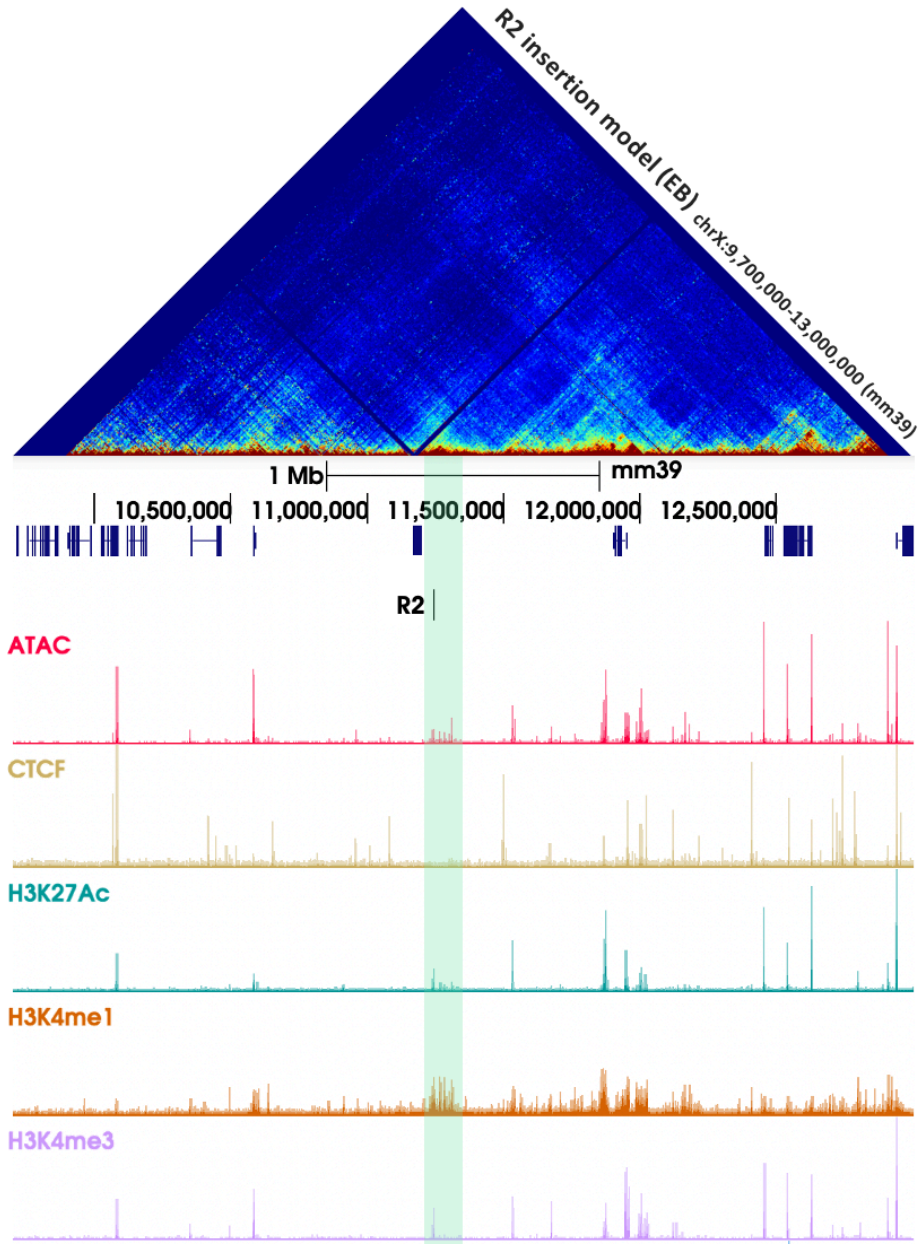



**Figure 2.6: Insertion of the R2 erythroid-specific enhancer into a neutral region of the genome initiates the formation of a TAD-like structure.**

Approx. 3.3 Mb region spanning the R2 insertion site of chromosome X. All data generated from CD71+ erythroid cells. Heatmap shows Tiled-C contact matrix at 5 kb resolution in mouse CD71+ erythroid cells containing the R2 insertion in chrX. Contact frequencies represent normalised, unique interactions in three replicates, original data generated by Marieke Oudelaar (Max Plank Institute). Open chromatin (ATAC-seq), Histone ChIP-seq (H3K27Ac, H3K4me1 and H3K4me3) tracks each represent merged data from 3 clones, original data generated by Caz Harrold. CTCF data is generated in WT cells (2 replicates merged), original data generated by Caz Harrold. GENCODE VM30 annotations are plotted in navy blue [162,163]. R2 insertion shown in purple. Region plotted: chrX:9,700,000-13,000,000 (mm39).

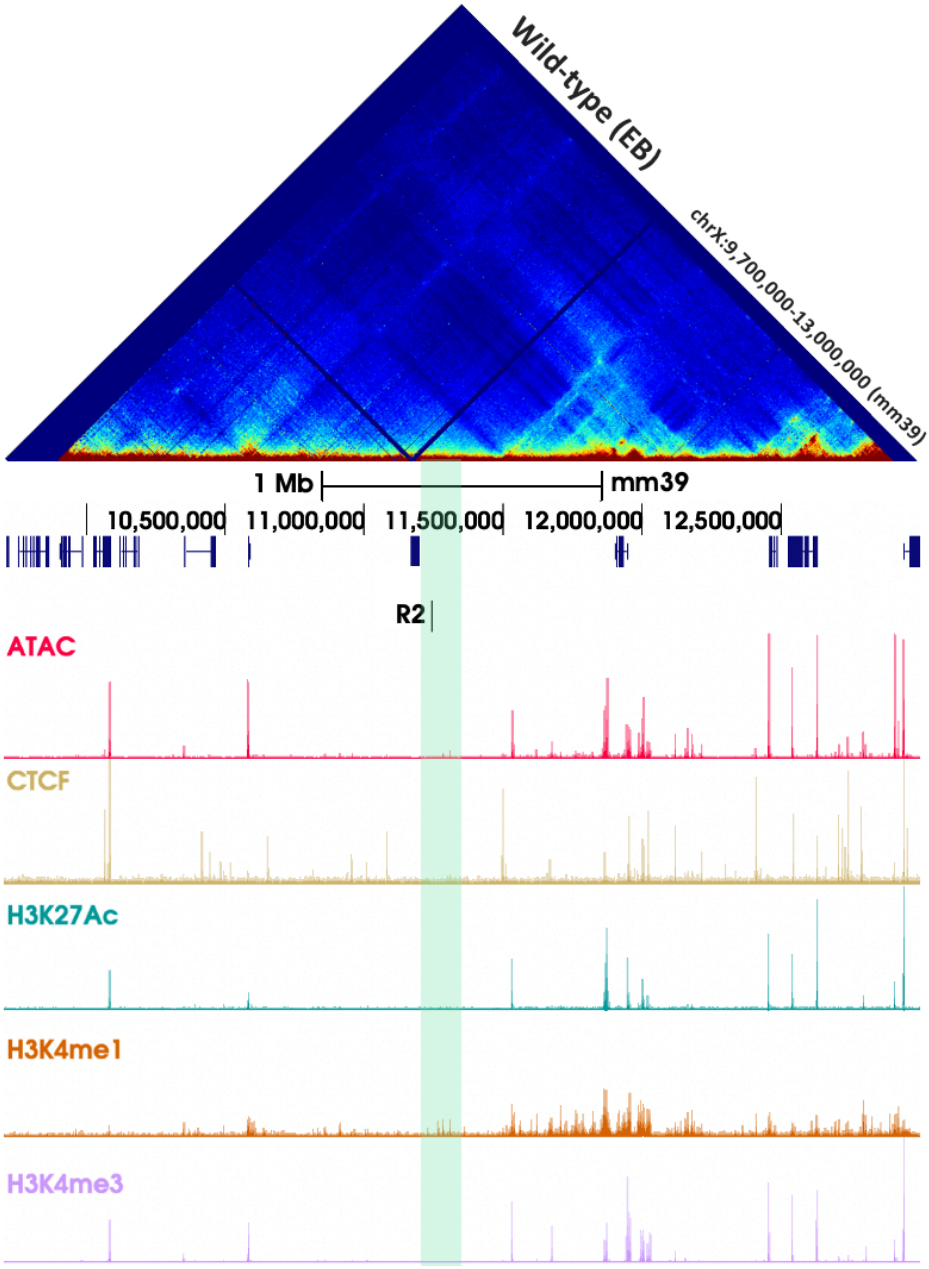



**Figure 2.7: The unedited (WT) neutral region of the genome initiates is devoid of the TAD-like structure seen in the R2 insertion model.**

Approx. 3.3 Mb region spanning the R2 insertion site of chromosome X. All data generated from wild-type CD71+ erythroid cells. Heatmap shows Tiled-C contact matrix at 5 kb resolution in mouse CD71+ erythroid cells. Contact frequencies represent normalised, unique interactions in three replicates, original data generated by Marieke Oudelaar (Max Plank Institute). Original data generated by Caz Harrold: open chromatin (ATAC-seq) from a single replicate; Histone (H3K27Ac, H3K4me1 and H3K4me3) and CTCF ChIP-seq tracks each represent merged data from 2 replicates. GENCODE VM30 annotations are plotted in navy blue [162,163]. R2 insertion shown in purple. Region plotted: chrX:9,700,000-13,000,000 (mm39).

### 2.2.2 Use of RAD21$^{TST}$ for reliable cohesin ChIP-seq

It is known that the subunits of the cohesin complex exist in a 1:1:1:1 stoichiometry[60]. I analysed the available ChIP-seq data for these 4 subunits (SMC1, SMC3, RAD21 and STAG1/2) to confirm that the profiles were all approximately representative, as shown in Figure 2.8. I decided to proceed with all further experiments using RAD21 as a representative for the cohesin complex. This is the most commonly studied subunit of the cohesin complex in the context of 3D genome organisation and would allow me to make use of existing tagged RAD21 lines generated in our lab.

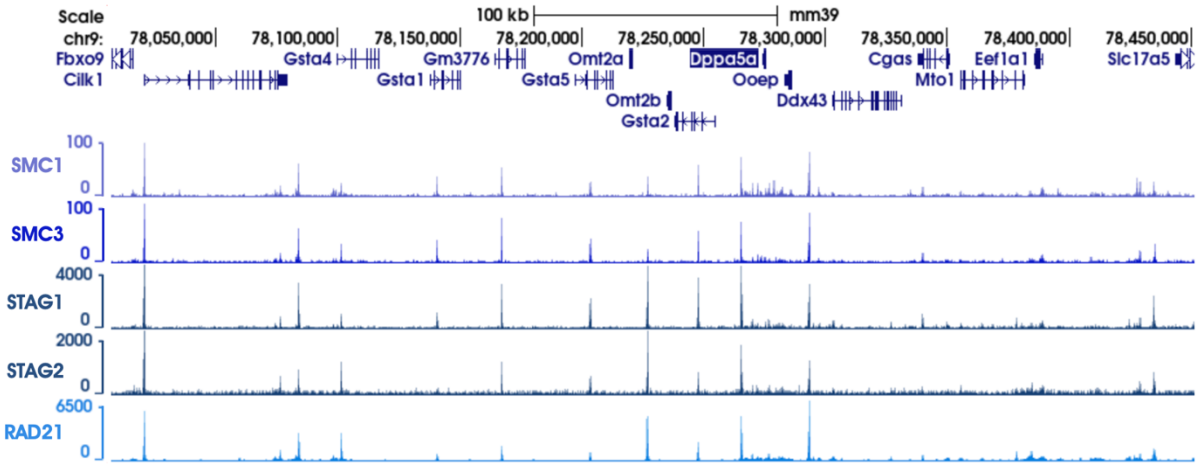



**Figure 2.8: ChIP-seq of the cohesin subunits around the Dppa5a locus.**
ChIP-seq data for SMC1[75], SMC3[75], the STAG subunits[164] (STAG1 and STAG2), and RAD21 (my data). Region shown is chr9:78,006,557-78,454,132 (mm39). See Materials and Methods for further details on data sources.

To facilitate reliable and efficient RAD21 ChIP, endogenous RAD21 was tagged at the C-terminus with a twin Strep tag (TST) in the pre-existing R2 insertion model using CRISPR-cas9 mediated HDR. From this point tagged RAD21 will be referred to as RAD21$^{TST}$. The TST was chosen as it is a small, inert tag which has been used in several other projects in the lab, including the generation of WT RAD21$^{TST}$ HEK293 and mESC lines. As a lab, we have been able to reliably generate high quality StrepTag ChIP-seq data in the RAD21$^{TST}$ WT cell

lines that recapitulates RAD21 ChIP-seq, thus validating the RAD21 tagging (Figure 2.9). Throughout this chapter the mESC RAD21$^{TST}$ line was used as a positive control for ChIP, and the HEK293 RAD21$^{TST}$ as a spike-in control for calibration purposes.

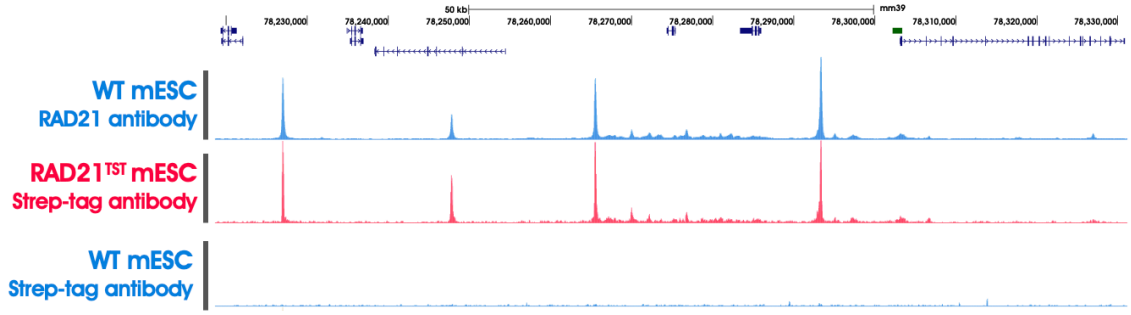



**Figure 2.9: Immunoprecipitation of RAD21 and RAD21TST gives identical profiles.**
ChIP-seq was performed in WT and RAD21$^{TST}$ mES E14 cell lines to compare the use of the anti-RAD21 antibody versus anti-Strep-tag antibody to immunoprecipitate the RAD21 protein. Top blue track shows ChIP-seq data in a WT cell line using an anti-RAD21 antibody (ab992, abcam). Middle pink track shows ChIP-seq data in the RAD21$^{TST}$ cell line using an anti-Strep-tag antibody (34850, Qiagen). The bottom blue track shows a lack of protein-pull down when the anti-Strep-tag antibody (34850, Qiagen) is used for ChIP-seq in a WT cell line. Region shown is chr9:78,218,597-78,331,196 (mm39).

### 2.2.3 Activation of the R2 erythroid enhancer is linked with cohesin recruitment

To ensure that the R2 RAD21$^{TST}$ line did not behave differently to WT cells in the undifferentiated state, StrepTag ChIP-seq was performed in undifferentiated mES E14 cells and compared with WT RAD21$^{TST}$ line. No visually observable difference was seen between the two cell lines (Figure 2.10). I therefore concluded that the tagging of RAD21 in the R2 insertion model did not interfere with normal cohesin recruitment, and that there is no cohesin recruitment to the inserted R2 enhancer when it is inactive. Next, I proceeded to differentiate these lines using the erythroid differentiation system and isolated a pure population of erythroid cells by CD71 selection.

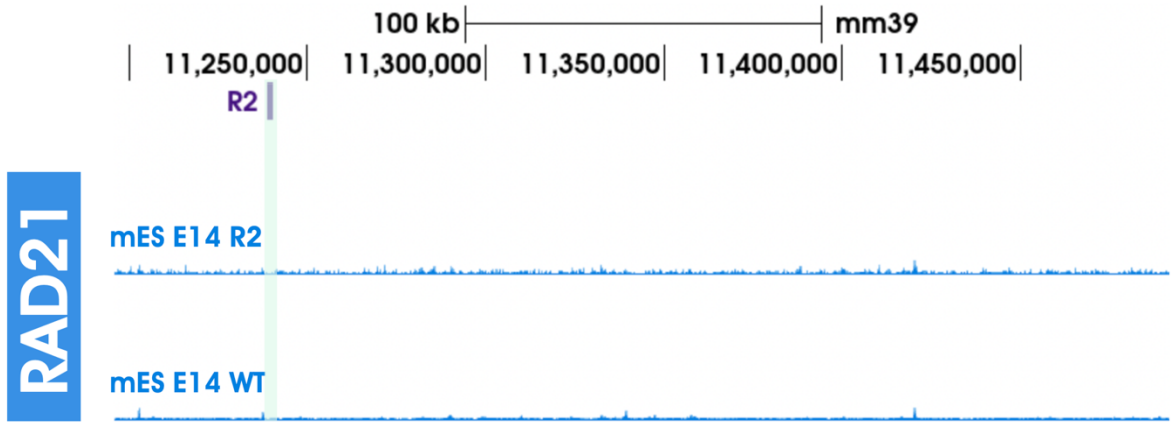



**Figure 2.10: Comparison of RAD21 ChIP-seq in WT mES E14 cells with those containing the inserted R2 enhancer.**
Region shown: chrX:11,195,902-11,492,154(mm39). R2-insertion mES E14 cells (two replicates merged), WT mES E14 cells (multiple replicates merged).

In the CD71+ erythroid cells the R2 enhancer is active. I therefore repeated the StrepTag ChIP-seq in WT erythroid and in erythroid cells containing the R2 insertion and intersected this data with the pre-existing ATAC-seq data generated in these cells (Figure 2.11). Upon activation of R2, a peak of open chromatin appears at this element, as well as a small peak of cohesin. Of the additional ATAC-seq peaks that appear downstream in the R2 EB cells, only the last peak (~71 kb downstream from the enhancer) also coincides with a large peak of cohesin. Several studies have shown evidence for cell-type specific CTCF mediated interactions[157,165], indicating that CTCF binding can vary extensively across cell types, specifically the sites associated with domain boundaries[166]. Whilst the CTCF ChIP-seq data in both WT EB cells and WT mESC did not show any significant peaks in this domain (Figure 2.11), it could be the case that the presence of a strong enhancer instates CTCF binding within the vicinity. Therefore, I performed CTCF motif analysis across this region using FIMO[167], the results are shown in the bottom track of Figure 2.11. Through this analysis I found that a putative CTCF binding site overlaps the cohesin peak at the most downstream open chromatin site. The fact that this CTCF site is on the negative strand suggests that CTCF may bind at that motif in the R2 EB line where it could act as a

boundary to block cohesin-mediated loop extrusion extending outside of this domain and result in the accumulation of cohesin seen at this site. If this was the case, it would suggest that cohesin is loaded from within the domain, perhaps at the enhancer, and translocates in the downstream direction towards the open chromatin marked CTCF boundary. This mechanism is supported by the domain structure seen in the Capture-C and Tiled-C data. Overall, this data suggests that the presence of the active R2 enhancer is somehow directing cohesin recruitment to this locus to set up a regulatory domain. Whilst this finding is one step closer to understanding cohesin's role in 3D architecture, it does not probe the dynamics of this mechanism as there is no way of pinpointing the exact site to which cohesin is initially recruited compared to where it translocates and stalls.

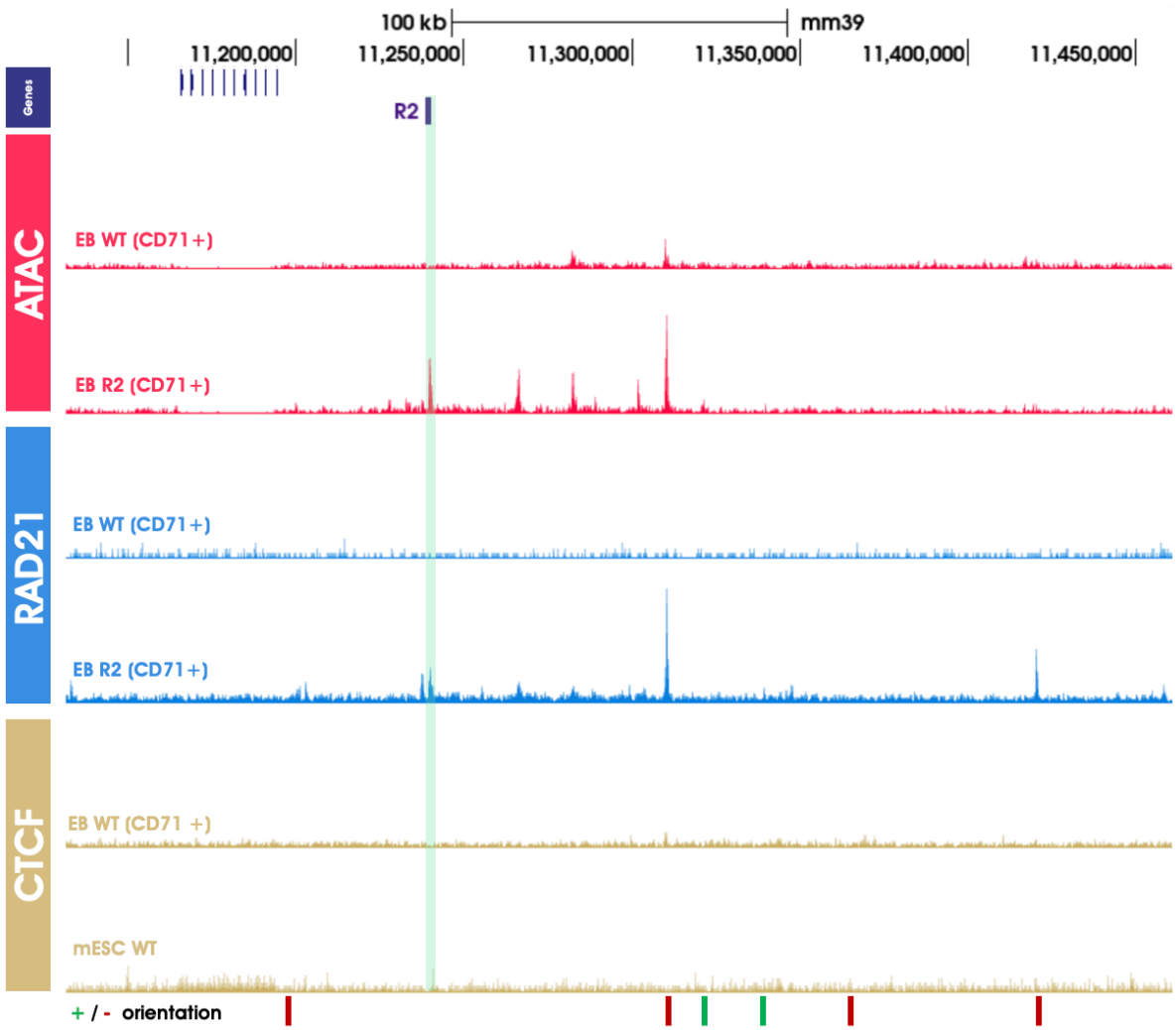



**Figure 2.11: Insertion of the R2 enhancer element directs cohesin (RAD21) recruitment to the locus.** Region shown is ~3Mb surrounding the R2 insertion site in chromosome X (chrX:11,131,442-11,461,132, mm39). R2 insertion site is labelled in purple and highlighted in green. GENCODE VM30 annotations are plotted in navy blue [162,163]. ATAC-seq* R2-insertion and WT EB (three replicates merged), RAD21 ChIP-seq R2-insertion (two replicates merged), RAD21 ChIP-seq WT EB* (1 replicate), CTCF ChIP-seq WT EB* (two replicates merged), CTCF ChIP-seq WT mESC (one replicate). CTCF orientation analysis is shown at the bottom: motifs on the positive strand in green and the negative strand in red. Motif analysis using FIMO, see Materials and Methods for details. *Indicates data generated by Caz Harrold[147].

## 2.3 Discussion

The results from this chapter build on the previous findings by Caz Harrold, showing that a single active regulatory element can set up a regulatory domain even when isolated outside of its native genomic context. By Tiled-C, I have shown that this regulatory domain is reminiscent of a TAD structure. Furthermore, I have shown that peaks of cohesin at the R2 element as well as at a downstream site appear in concordance with the formation of the domain. I therefore speculate that this domain formed as a result of loop extrusion, particularly given the presence of the putative CTCF binding motif that overlaps the accumulation of downstream cohesin. To corroborate these findings, I plan to perform CTCF ChIP-seq in the R2 EB model to determine whether CTCF does indeed bind at this motif in the R2 engineered line.

The Tiled-C data presented in the introduction of this chapter shows evidence for the dynamic nature of genome architecture throughout cellular differentiation. Whilst it is informative that the changes in 3D structure seem to be reflected by a change in cohesin occupancy within the locus, it is unclear if this is a cause or effect. The peaks of RAD21 shown in the R2-insertion ChIP-seq data provide information only on where cohesin resides at equilibrium: these peaks may indicate sites where cohesin is loaded, or sites where cohesin has accumulated after translocating from a different loading site. The site of recruitment therefore remains an open question, with two potential solutions: (1) the active element itself recruits cohesin, which can then translocate to open chromatin sites, (2) the active element is responsible for opening the chromatin so that it is more permissible to cohesin recruitment, allowing cohesin to be loaded both at the active element and the downstream open chromatin sites.

Recent studies evidence suggests that enhancers have the ability to recruit cohesin[80], which would suggest that perhaps option (1) is the most feasible. It has been shown that some of the erythroid specific transcription factors that bind the enhancers within the alpha-globin cluster can prime the enhancers for activation by initiating changes in chromatin modifications prior to globin gene upregulation[168]. Figure 2.12 contains a summary of the histone modification ChIP-seq data presented in this chapter, aligned with the RAD21 ChIP-seq, ATAC-seq data and CTCF motif analysis to further assess this idea. The striking novel peaks that appear in the ATAC-seq data following the activation of R2 (Figure 2.12, peaks 1-5) are also marked by histone modifications characteristic of active enhancers (H3K27Ac and H3K4me1). Whilst this region was originally defined as 'neutral' prior to the development of the model, H3K4me1 ChIP-seq showed that there were small peaks visible, particularly at peaks 2-5 (Figure 2.12), in WT cells which is indicative of a primed enhancer. This would provide an explanation as to why the insertion and activation of the R2 element results in the formation of these novel enhancer-like elements. However, if these open chromatin sites were acting as enhancers, I would expect to also see cohesin recruitment at these sites. By visual inspection of the RAD21 ChIP-seq data there is perhaps a slight enrichment of RAD21 occupancy at these sites, but these peaks are not significant in comparison to the major cohesin peak seen at the furthest downstream open chromatin site (Figure 2.12, peak 5). This does not rule out the possibility of cohesin loading at these sites, however it does suggest that there is another factor at play which I believe may be CTCF binding. To validate this, CTCF ChIP-seq would need to be performed in these edited cells to determine if this putative CTCF site is indeed occupied in the erythroid lineage. Additionally, it is important to note that peak 5 is marked by H3K4me1 in both the WT and R2 cell lines which is not indicative of a classical CTCF site. It could be the case that this histone mark is informative on tissue specific regulation of CTCF binding, which then drives domain formation. Previous studies have suggested that dynamic CTCF binding can occur at regulatory regions, particularly

enhancers, whose activity positively correlates with CTCF recruitment[169]. Hence, the H3K4me1 could act as a place-marker: by which as a primed enhancer, when activated enables CTCF binding in a tissue specific manner.

There are existing models which show evidence that a regulatory domain can be formed in the absence of promoters[170], which supports the idea that enhancers are the key drivers of regulatory domain formation. Intriguingly however, the R2 enhancer element (Figure 2.12, peak 1) seemed to switch from displaying enhancer-like chromatin marks (H3K4me1 and H3K27Ac), to displaying promoter-like marks (H3K4me3). Several studies have shown that enhancers have the ability to act as promoters and vice-versa [171-174]. This evidence, together with that presented in this chapter, suggests that there is perhaps a context dependency on the chromatin modifications and function that a *cis*-regulatory element displays. If this is the case, perhaps domain formation is driven more broadly, by the presence of active *cis*-regulatory elements, irrespective of whether these are promoter- or enhancer-like in nature.

From the data presented in this chapter, as well as from other studies[56,175,176], we know that upon the activation of a locus, there is a change in genome architecture which correlates with the pattern of cohesin occupancy. However, currently there is no way of knowing whether cohesin recruitment precedes the formation of the 3D genome architecture at active loci. In the next chapter I have developed a method to assay genome-wide cohesin dynamics so that I can hone in on the recruitment of cohesin on a global scale.

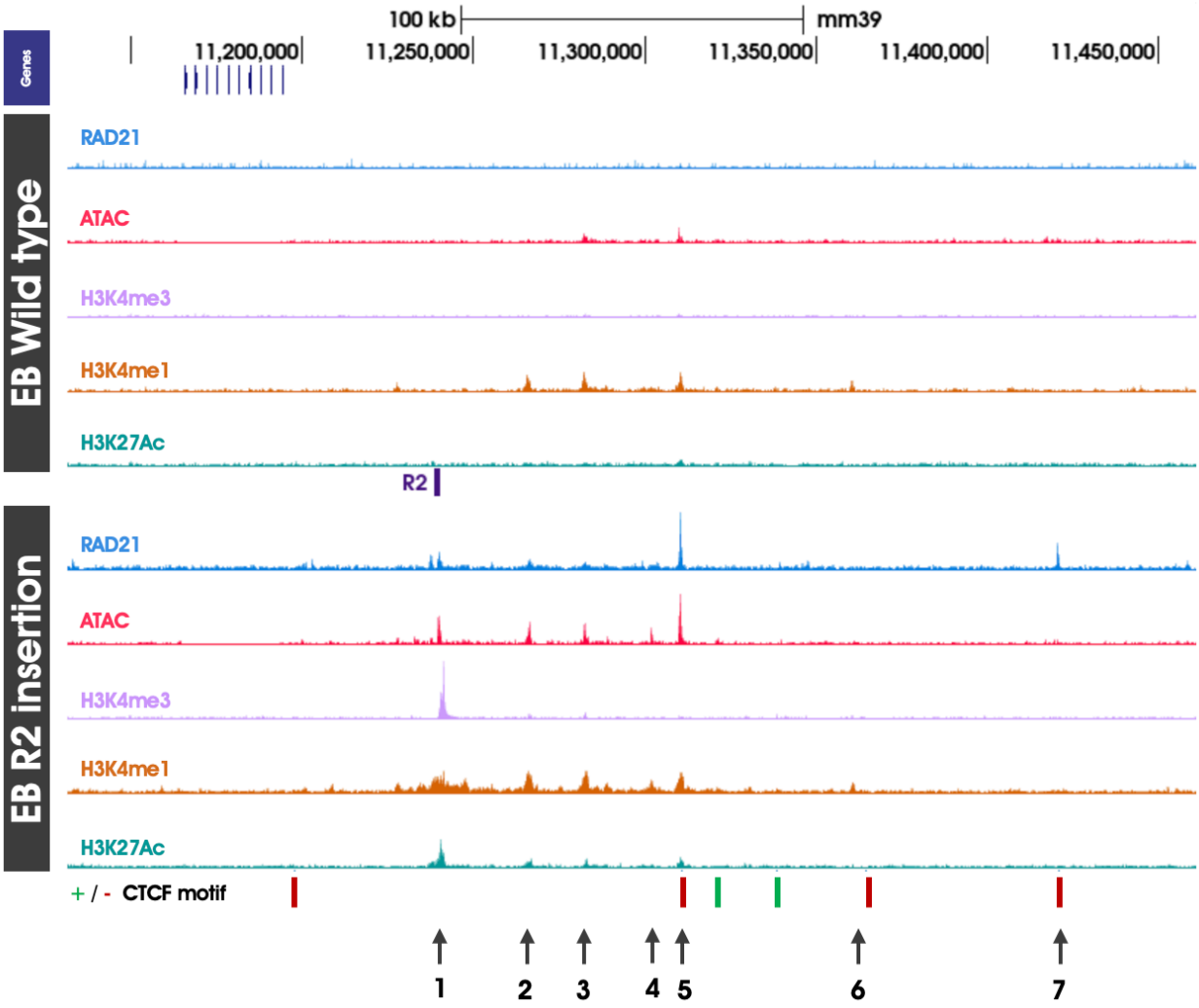



**Figure 2.12: Comparison of the chromatin landscape following the insertion of the R2 element into chromosome X.**

All tracks shown have been previously shown in the context of the larger ~3.3 Mb region shown in Figures 6 and 7, with the addition of the RAD21 ChIP-seq data from Figure 2.11 and CTCF motif orientation. Arrows and numbers indicate peaks of interest for reference. Region shown: chrX:11,131,442-11,461,132 (mm39).

## 2.4 Acknowledgements

This chapter is built on the foundations of the work done by Caz Harrold (PhD student, Hughes Group), who designed and developed the R2 insertion model. Philip Hublitz (WIMM, Genome Engineering Services) designed and produced the constructs for tagging RAD21 with TST. Nigel Roberts (WIMM, Hughes lab) carried out the genome engineering to tag native RAD21 with TST in the R2 model and WT cell lines. Due to the laboratory access restrictions in place at the University during the pandemic, the Tiled-C Capture reaction was performed for my analysis by Marieke Oudelaar (Max Planck Institute).

# 3 Genome-wide methods to study dynamics using a degron system

## 3.1 Introduction

### 3.1.1 Moving on from an equilibrium view

The findings from the previous chapter (Chapter 2) suggest that active elements, such as promoters and enhancers, may recruit cohesin which facilitates 3D interactions between these elements to initiate transcription. In Chapter 2 I took a focused approach, investigating recruitment at a single locus using ChIP-seq and combining these data with Tiled-C and ATAC-seq data to build an understanding of how the 3D domain is built *de novo*. The caveat of the ChIP-seq data shown in Chapter 2 is that it represents equilibrium cohesin occupancy at this locus, thus I cannot assume that cohesin is loaded directly to the sites where we see the peaks. Instead, these peaks show where cohesin accumulates via recruitment at these sites, or by migration to them. In this chapter I aimed to investigate the cohesin dynamics on a genome wide scale using a well-established degron system.

### 3.1.2 Degron systems as a means of modelling dynamics

As discussed in the main introduction (Chapter 1), there have been a number of studies using degron systems to investigate the role of regulatory proteins in 3D genome architecture, in particular: cohesin, CTCF and the cohesin unloading factor WAPL. The majority of these studies have used next-generation sequencing coupled 3C methods such as Hi-C to assay loop length, cohesin stability and TAD structure. In this chapter I want to use degradation as a means of identifying the regulatory elements that recruit cohesin. Whilst it may seem counterintuitive to investigate the recruitment of a protein by a degradation method, I am working on the assumption that the chosen method of degradation depletes the pool of unbound, freely diffusing protein. With no freely diffusing cohesin, the elements which recruit cohesin will no longer be able to do so, therefore on a ChIP-seq track we would expect to see that the peaks where cohesin is recruited reduce most rapidly during a degradation time-course experiment. It is known that cohesin can translocate along the DNA, stall, and accumulate at boundary elements, namely CTCF sites. It could therefore be expected that the regions in which cohesin is translocating will reduce in signal gradually throughout the time-course, with the peaks at boundary elements to be the last to be depleted. As with all models, there are a few assumptions and caveats that must be considered when designing these experiments, these are discussed below.

#### Degron tagging RAD21 does not affect normal cohesin recruitment.

Whilst we cannot be certain that this is the case, there are a number of ways in which this can be tested. Firstly, given that RAD21 is an essential part of the cohesin complex that is required for both sister chromatid cohesion and 3D architecture, cell survival is a good indication that RAD21 functionality and recruitment to chromatin is unaffected. If degron-tagged cells behave comparably to WT cells, by dividing and progressing through the cell

cycle as normal, it can be assumed that the tagging is not inhibiting the normal cell cycle function of RAD21, though there may be subtle changes that are not easily detectable. ChIP-seq can be used to assay the chromatin association of RAD21 in degron-tagged cells in the uninduced state compared to WT cells to determine whether the tagging affects the ability of RAD21 to bind to chromatin. Additionally, the subcellular localisation of tagged RAD21 can be assayed by western blot and compared to WT cells to ensure that the tagging does not impact subcellular localisation which would directly impact cohesin recruitment.

### Chromatin associated RAD21 is not directly degraded, it must be dissociated from chromatin in order to be degraded.

One of the main assumptions made about the degradation process is that only unbound cohesin is degraded. To my knowledge there is no evidence in the literature to support either the bound or unbound degradation mechanisms. To test this, I aimed to measure the degradation efficiency across the subcellular compartments to ascertain this information.

### Cohesin is actively loaded at specific sites, rather than randomly across the genome.

There have been a number of studies showing that cohesin is actively loaded onto the chromatin by the NIPBL/MAU2 complex, disruption of which results in loss of cohesin from the chromatin[86] and suggests that cohesin is indeed actively loaded at specific sites. As mentioned in the introduction, however, there has been much debate over the exact sites at which cohesin is loaded and is one of the key questions addressed in this thesis. Cohesin has innate ATPase activity which allows it to move along the chromatin, there is also evidence that NIPBL can activate the ATPase domain[84-86] and transverses along the chromatin with cohesin[177]; with current experimental methods such as ChIP-seq and 3C

techniques being population average approaches, it is difficult to disentangle the loading site of cohesin from other sites along the chromatin at which cohesin stalls or accumulates.

## Genome-wide variability in RAD21 binding dynamics

Sites which can retain cohesin for longer, i.e., the sites that exhibit a longer RAD21 residence time, should demonstrate a degree of resistance to degradation as opposed to the highly dynamic sites with shorter RAD21 residence time. Performing ChIP-seq over the time-course of the degradation should reflect this by the highly dynamic sites showing the most rapid degradation and vice versa (Figure 3.1).

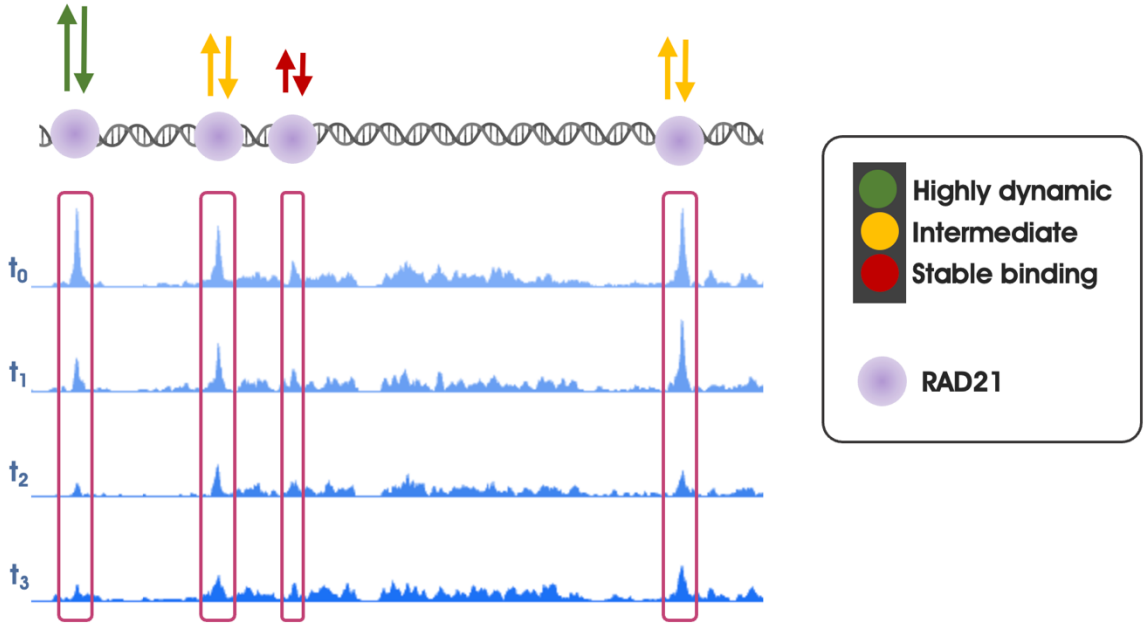



**Figure 3.1: RAD21 sites may differ in the ability to retain stably associated RAD21, which should be reflected in the ChIP-seq tracks during a degradation time-course.**
RAD21 is represented as a purple sphere at specified sites along the chromatin. Theoretical ChIP-seq tracks are shown below in blue from pre-degradation ($t_0$), and at regular intervals throughout the degradation. Highlighted regions show examples of how RAD21 occupancy might change throughout the degradation depending on the stability of chromatin association.

### 3.1.3 The dTAG system for targeted degradation of RAD21

In this chapter the degron tag (dTAG) method was selected as the method of choice to achieve targeted and acute RAD21 degradation. As briefly mentioned in the introduction chapter, the dTAG method is a two-part system, it requires: (1) the expression of a mutant form of FKBP12 ($FKBP12^{F36V}$) in-frame with the RAD21 gene and (2) the treatment with a heterobifunctional degrader molecule, dTAG-13. Upon treatment with the degrader molecule, an E3 ligase is engaged which catalyses the ubiquitination of RAD21, and subsequent proteasomal degradation (Figure 3.2). Genetic knockouts/knockdowns require a significant delay before loss of transcription translates into impactful reduction at the protein level; protein loss in these cases is not abrupt and instead relies on the natural rates of protein turnover which could make tracking sites of protein loss more challenging. The dTAG method in comparison, was chosen given the convincing evidence from previous publications on the ability to degrade the target protein rapidly and acutely, which was essential in order to trace degradation by time-course ChIP-seq experiments. Additionally, the relatively simple genome engineering strategy in which a single tag of $FKPB12^{F36V}$ is added to RAD21 should increase the likelihood of generating positive and homozygous clones[125,137].

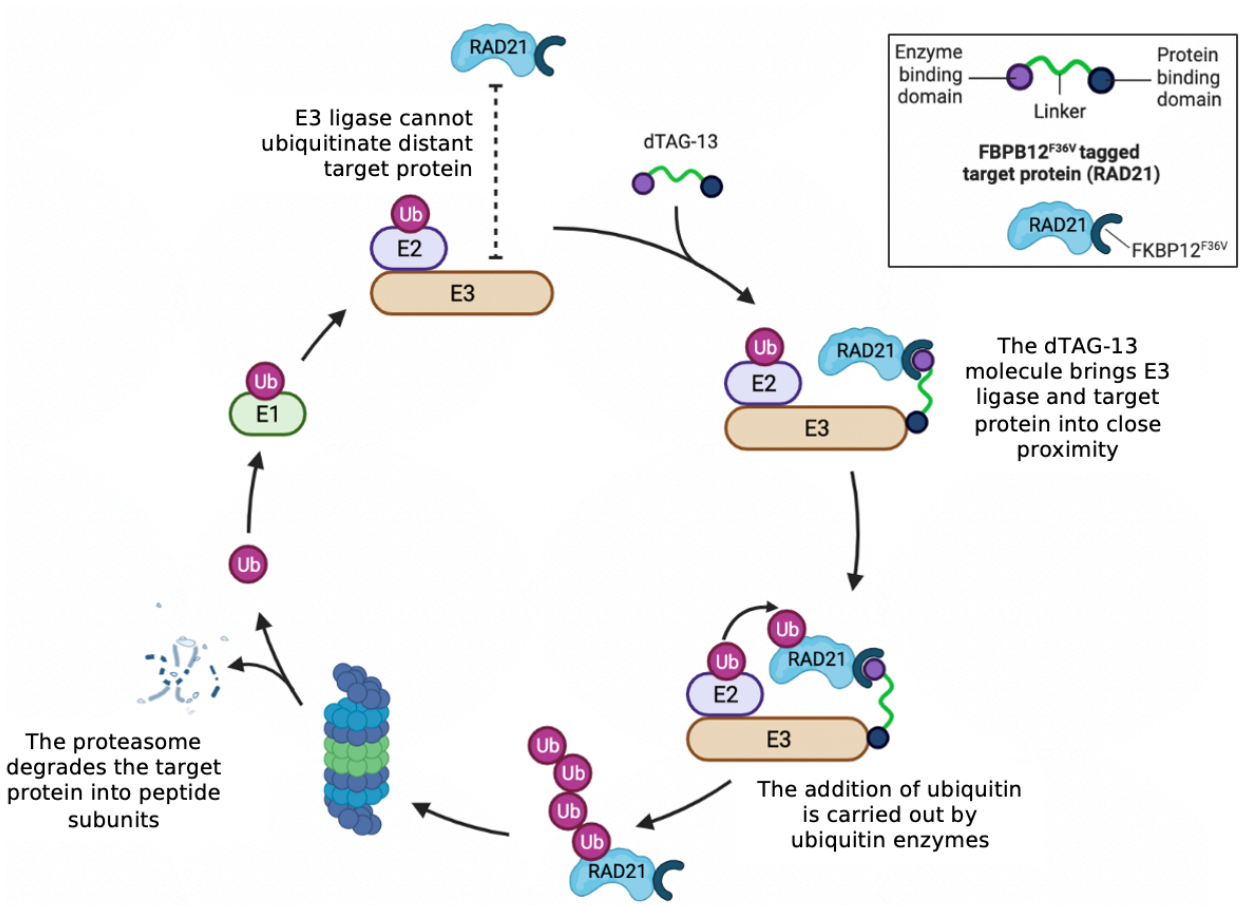



**Figure 3.2: Hijacking a cell's intrinsic degradation machinery: the mechanism of action of dTAG-13 mediated degradation.**

The protein to target for degradation, in this case RAD21, is tagged with a mutant version of FKBP ($FKBP12^{F36V}$). When the dTAG-13 molecule is administered, it binds one end to $FKBP12^{F36V}$ and the other end to an endogenous E3 ubiquitin ligase which initiates the ubiquitination process and subsequent degradation by the proteasome[125]. Figure made using BioRender.

# 3.2 Results

## 3.2.1 Genome engineering strategy

Philip Hublitz from the WIMM Genome Engineering Facility designed the cloning strategy and built the FKBP$^{F36V}$-FLAG4x construct for tagging of the C-terminal domain (CTD) of native RAD21 locus using CRISPR-cas9 editing (Figure 3.3).

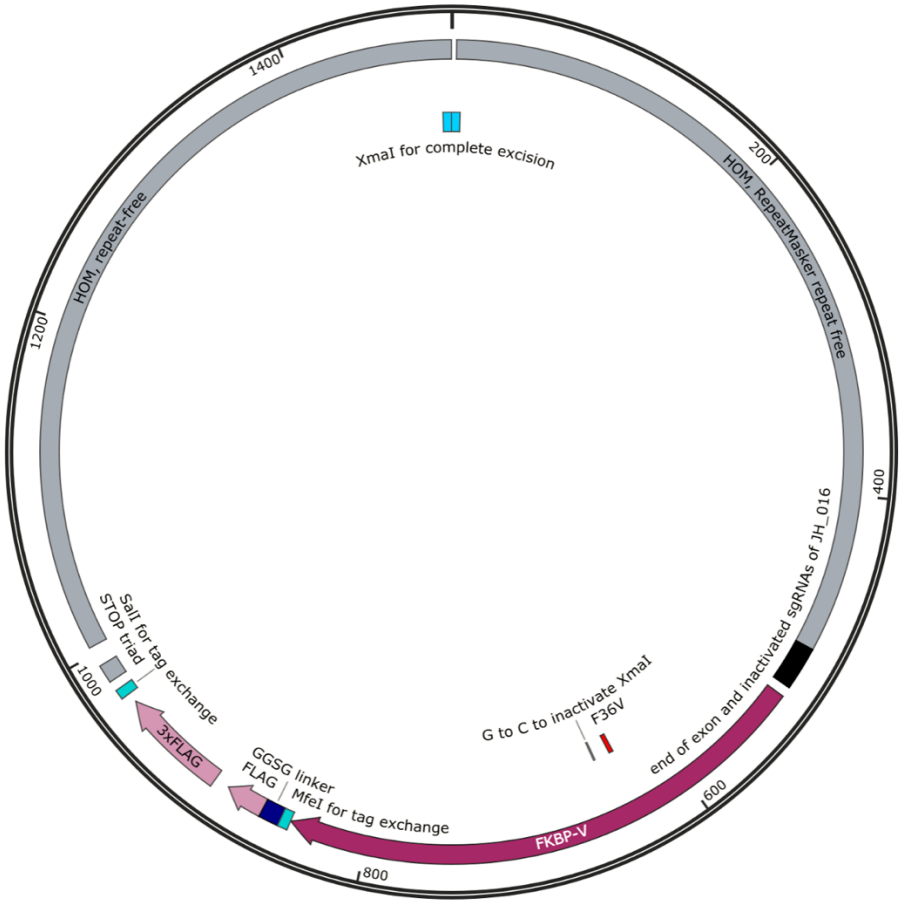



**Figure 3.3: RAD21-FKBP$^{F36V}$(dTAG)-4xFLAG donor vector map.**
Total vector size is 1496 bp, containing the FKBP$^{F36V}$-4xFLAG components flanked by ~500bp homology arms. Vector design includes the insertion of several unique REs for easy exchange of homology arms, tags (FLAG/TST), flexible linkers and potential to remove TAG stop codon if used for N-terminal domain tagging. To ensure the tag is at the extreme C-terminus and thus optimally available for ChIP, a flexible GSGG$_{4x}$ linker has been included between the FKBP$^{F36V}$ and FLAG-tag.

We chose to include the FLAG (DYKDDDDK) affinity tag in the design to account for the possibility of successful clones being heterozygotes, which would require a means of distinguishing the edited from the WT allele. Previous projects in the Milne lab (WIMM) had used a tandem repeat of four FLAG sequences (DYKDDDDK-GTR-DYKDHD-G-DYKDHD-I-DYKDDDDK-TR) to achieve high quality ChIP-sequencing data, here the same strategy was applied. Also included in the design were multiple restriction endonuclease (RE) sites to make the construct as adaptable and versatile as possible for use in other related projects, also flanking of the stop codon with a unique RE for possible N-terminal domain (NTD) tagging if required. CRISPR-Cas9 mediated homology-directed repair (HDR) was used to tag the C-terminal domain (CTD) of RAD21 in WT mES E14 cells. Of the edited cells, one clone (C8) successfully passed the PCR screening strategy (Figure 3.4A-B). By Sanger sequencing, the clone was confirmed to be homozygously tagged and did not contain any large deletions in the tag sequence (Figure 3.4C). Given that the positive clone was homozygous for tagged RAD21 we have the option of using either RAD21 or FLAG antibodies in the downstream experiments. All subsequent experiments in this chapter use mESC clone C8. For the remainder of this chapter, I will refer to the C8 tagged RAD21 as $RAD21^{FLAG}$.

**A**

| Primer set | Forward primer | Reverse primer | Size (bp) WT | Size (bp) Edited | Notes |
|---|---|---|---|---|---|
| External | RAD21 ExF | RAD21 ExR | 1137 | 1630 | Both bands present in heterozygotes |
| Internal #1 | JH_025 FKBP F | JH_025 FLAG R | - | 99 | Present in random insertion |
| Internal #2 | JH_025 HOM F | JH_025 FLAG R | - | 526 | Present in random insertion |
| Internal #3 | JH_025 FKBP F | JH_025 HOM R | - | 560 | Present in random insertion |
| Spanning #1 | JH_025 FKBP F | RAD21 ExR | - | 774 | Position dependent |
| Spanning #2 | RAD21 ExF | JH_025 FLAG R | - | 955 | Position dependent |

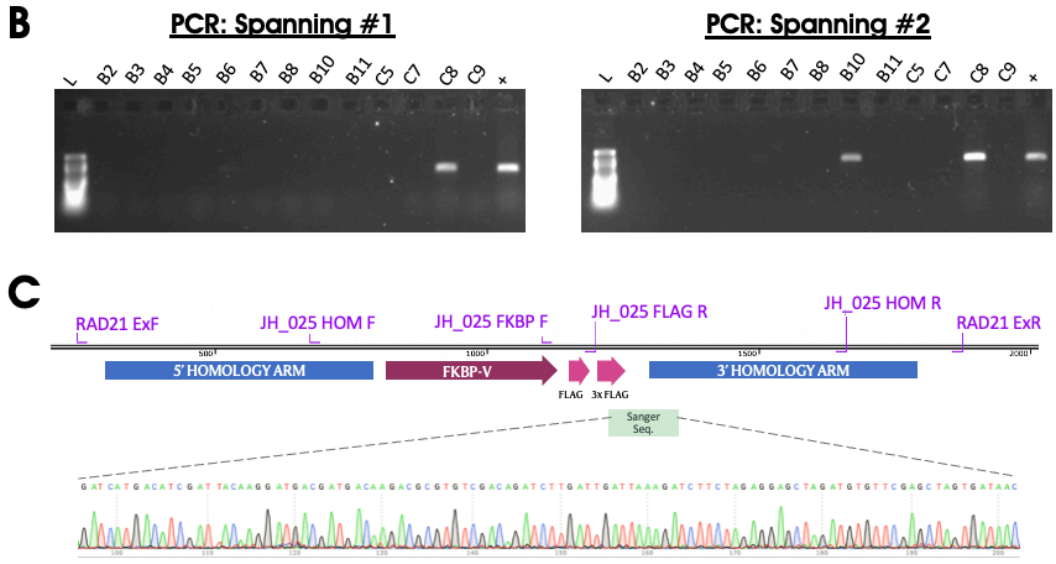



**Figure 3.4: A single RAD21$^{FLAG}$ homozygous clone identified by PCR screening and validated by Sanger sequencing.**

Surviving clones were screened using the above screening strategy (for full details see Materials and Methods). **a** | Table showing the primer pairs and expected band sizes in WT and edited clones. External primers can distinguish between wild type and edited clones by band size. Internal primer sets 1-3 will only be present in edited clones but will not distinguish between random insertions of the plasmid versus correctly targeted clones. Spanning primer sets 1-2 will only give bands if the clone has been correctly targeted, agarose gels for these primer sets are shown in b. **b** | Agarose gel images of spanning primer PCR products. Clone C8 gave correctly sized bands in sets 1 and 2 and was therefore selected for further validation using Sanger sequencing. + = positive control, L = ladder. **c** | Linear vector map with annotated primers. Green bar indicates region shown in Sanger sequence trace. Sanger sequencing of clone C8 confirmed homozygosity and did not identify any deletions or rearrangements.

### 3.2.2 System validation by subcellular fractionation and western blot

Western blot is an excellent technique for visualising cellular changes in protein levels. I chose to use fluorescent western blot for all western blot experiments for a number of reasons. Firstly, fluorescent western blot is a direct detection method (compared to indirect chemiluminescent methods) which means that the signal is directly proportional to the amount of protein, making quantification more accurate. Secondly, this technique has the widest dynamic range of western blot detection[178] enabling both high and low abundance proteins to be detected in the same experiment, which is useful when the relative abundances of proteins across samples is unknown. There is also the option to multiplex antibodies when using fluorescent western blot, which can be useful for visualising tagged ($RAD21^{FLAG}$) and untagged ($RAD21^{WT}$) proteins on the same blot.

To further increase the detection capacity, I performed subcellular fractionation this was particularly important for the nuclear fraction as nuclear proteins as often difficult to detect by western blot, thus isolating a nuclear specific fraction allowed for enrichment of these proteins that are usually diluted by the more abundant cytoplasmic proteins. The subcellular fractionation was performed using the ThermoFisher subcellular protein fractionation kit for cultured cells, slight modifications were made to the manufacturer's protocol to achieve optimal fractionation (Figure 3.5, full details in Material and Methods). Using this protocol, cytoskeletal (CS), nuclear chromatin (NC), nuclear soluble (NS), cytoplasm (C) and membrane (M) fractions were obtained from WT mESCs, in parallel with a whole cell lysate (W) extraction. A selection of well characterised antibodies was used to probe the fractions to ensure that each fraction was representative of the expected subcellular compartment, and that the method was able to achieve a clean fractionation with minimal spill-over into the other fractions (Figure 3.6). TopoII-beta was chosen as a control for the nuclear fractions, with Histone-2B as a control for the nuclear

chromatin fraction specifically. Both Golgin-97 and Alpha-tubulin were specific to the cytoplasmic fraction, with alpha-tubulin being more abundant and therefore also detectable in the whole cell lysate.

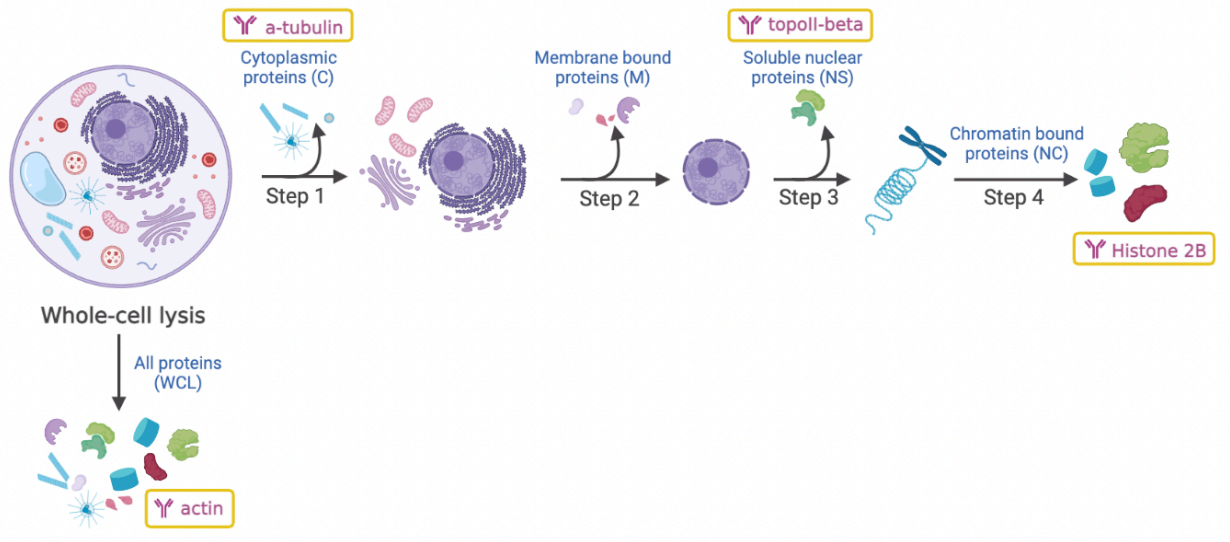



**Figure 3.5: Simplified diagram of the subcellular isolation strategy.**
Step 1: selective permeabilization of the plasma membrane and isolation of cytoplasmic proteins, step 2: recovery of intact nuclei by centrifugation and isolation of membrane bound and organelle proteins, step 3: permeabilization of the nuclear membrane and isolation of soluble nuclear proteins, step 4: addition of MNase to release chromatin bound protein. Fractions obtained from each step are labelled in blue, with the abbreviations in brackets that are used throughout this chapter. Control antibodies for each fraction are shown in yellow boxes. Figure created using BioRender.

**A** Table of antibodies used as fraction controls

| Antibody | Localisation | MW (KDa) |
|---|---|---|
| TopoII-beta | Nucleus | 180 |
| RAD21 | Nucleus | 130 |
| Golgin-97 | Cytoplasmic face of Golgi apparatus | 97 |
| Alpha-tubulin | Cytoplasm | 50 |
| Actin | Ubiquitous | 42 |
| Histone 2B | Nuclear, chromatin associated | 17 |

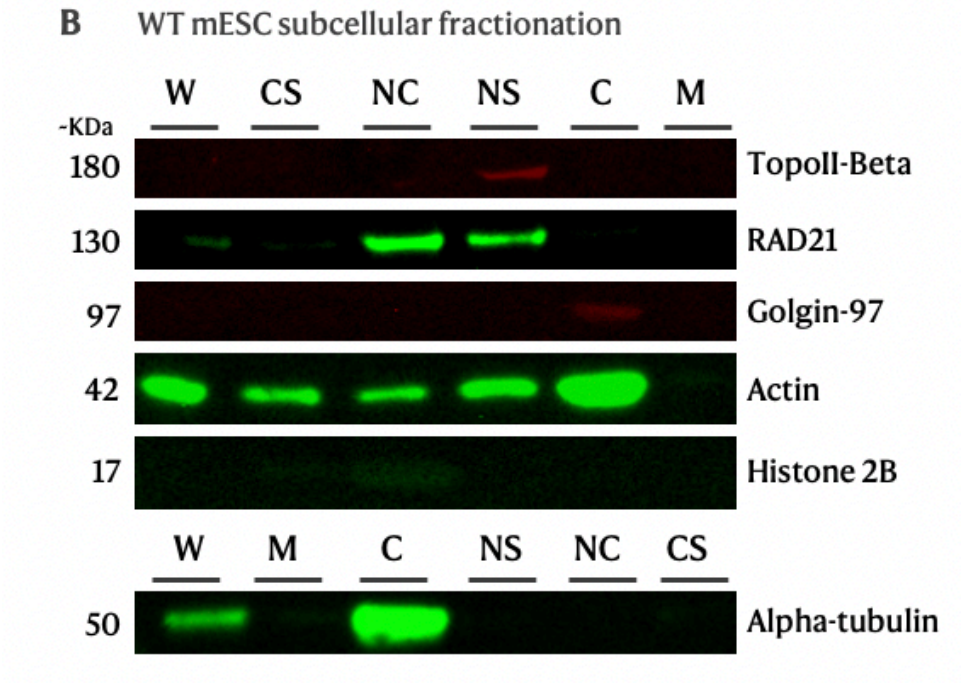



**Figure 3.6: Subcellular fractionation of WT mESC cells.**

**a** | Fraction specific antibodies for use as fraction controls in subcellular fractionation western blot experiments. Table shows the antibodies used, along with their localisation and expected molecular weight. **b** | Fluorescent western blot showing subcellular fractionation in WT mES cells. Whole cell lysate (W), cytoskeletal (CS), nuclear chromatin (NC), nuclear soluble (NS), cytoplasm (C) and membrane (M) fractions are shown.

## How long does it take to achieve complete protein degradation?

A key consideration in this set of experiments was the degradation timing. There are a number of published studies that use the dTAG system for protein degradation, although none thus far have been used to track degradation dynamics over a time-course. In my experiments, I aimed to achieve full degradation of RAD21 over a period of time that would be experimentally feasible and allow sufficient time to harvest cells at regular intervals during the degradation process. As a starting point, I chose to use a 500 nM concentration of dTAG-13 as this had been used as a working concentration in the original dTAG paper[125] and shown to work in other published studies[137]. I administered 500 nM dTAG-13 to adherent $RAD21^{FLAG}$ mES E14 cells, incubated them at 37°C, then at regular intervals throughout a 24-hour period I harvested the cells and immediately proceeded with protein extraction. Protein degradation efficiency throughout the time-course was monitored by western blot analysis (Figures 3.7A-B). From Figure 3.7A-B it is clear to see that the majority of protein has been degraded within a 2-hour time window, suggesting that at this concentration, dTAG-13 is able to rapidly degrade RAD21. Interestingly, at 0 hours (i.e., no dTAG-13 added) there appears to be a greater amount of soluble (NS) RAD21 as opposed to chromatin bound RAD21 (NC), which is in agreement with published data which shows that soluble RAD21 in human cells is approximately twice as abundant as chromatin-bound RAD21[179]. After 2 hours of dTAG-13 treatment, RAD21 is no longer visible in the nuclear soluble (NS) fraction, which is followed by an increasingly faint band of chromatin-bound (NC) RAD21 during the proceeding timepoints. This suggests that the unbound RAD21 is degraded more rapidly than the chromatin bound RAD21, thus supporting the hypothesis that the dTAG method of degradation targets unbound protein, preferentially to chromatin bound protein. Another point to note is the rise in nuclear chromatin signal seen at 4 hours, however given that this data only represents a single replicate, it is difficult to draw any conclusions from this.

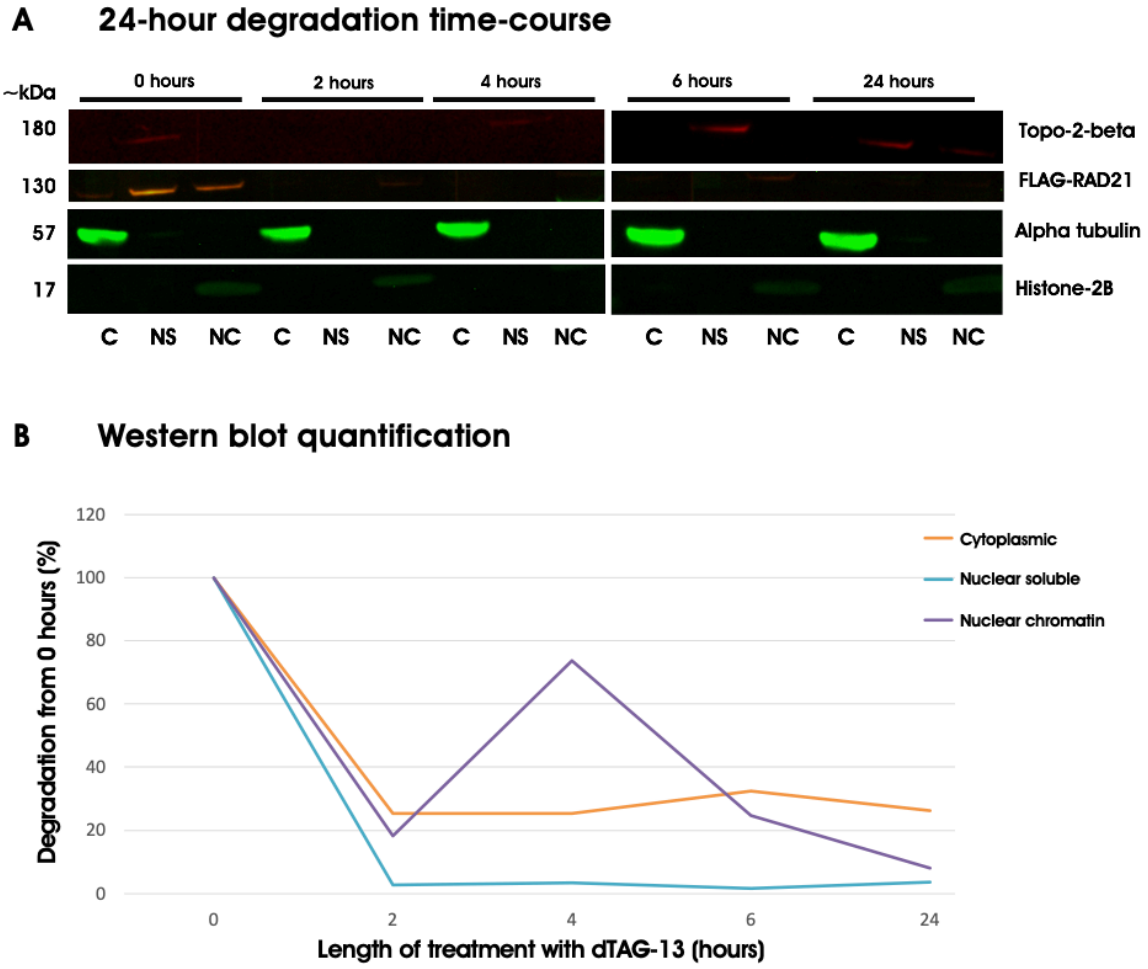



**Figure 3.7: Degradation of RAD21$^{FLAG}$ over a 24-hour time-course following treatment with 500 nM of dTAG-13.**

**a** | Fluorescent Western blot images. Timepoints were split over two separate blots. Loading controls for subcellular fractions are as follows: Alpha-tubulin: cytoplasmic (C), TopoII-beta: nuclear soluble (NS), Histone-2B: nuclear chromatin (NC). The FLAG-RAD21 panel shows the simultaneous detection of RAD21 (800nm, green-channel) and FLAG (700nm, red-channel). **b** | Quantification of Western blot data. Signal was quantified for RAD21$^{FLAG}$ bands at each timepoint. Undegraded (0 hours) band was taken as 100%, and percentage degradation from this level was plotted for each fraction. See Materials and Methods for full details of analysis, n=1.

### What is the optimal concentration of dTAG-13 to use over the time-course?

Next, I wanted to titrate the concentration of dTAG-13 molecule used for the degradation to determine what would be the optimal concentration to use in order to capture the degradation progressing throughout the time-course (Figure 3.8A-B). For this experiment dTAG-13 was added at varying concentrations (1 nM to 10uM) for a 4-hour treatment duration. In the previous experiment, 2-hours was sufficient to see over 70% degradation of $RAD21^{FLAG}$. Therefore, by choosing a slightly longer timeframe of 4-hours, I was aiming to capture degradation using lower concentrations of dTAG-13 which would perhaps require more time to achieve similar levels of degradation. Both 1 nM and 5 nM concentration were not sufficient to degrade $RAD21^{FLAG}$ within a 4-hour timeframe, whilst all higher concentrations tested appeared to show full degradation at this point. As a result, I decided to maintain the dTAG-13 concentration for subsequent experiments at 500 nM. From these results I was confident that the dTAG degradation of $RAD21^{FLAG}$ was functioning as expected at the cellular level.

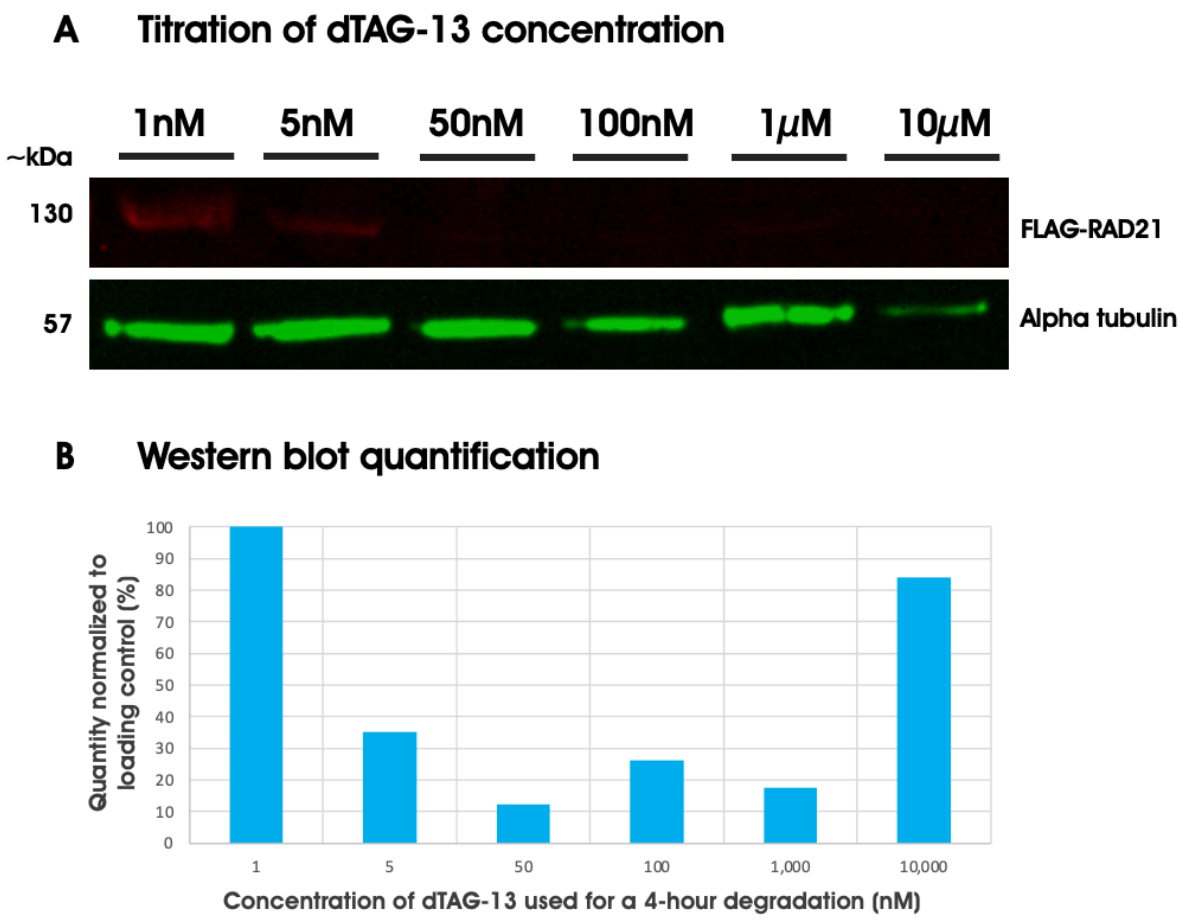



**Figure 3.8: Titration of dTAG-13 concentration for a 4-hour degradation.**

RAD21$^{FLAG}$-dTAG cells were subject to 4hrs degradation using dTAG-13 concentrations varying from 1nM to 10uM in order to determine the optimal concentration to use. **a** | Fluorescent western blot showing the decrease in RAD21$^{FLAG}$ with increasing dTAG-13 concentration in whole cell lysates. The FLAG-RAD21 panel shows FLAG signal (700 nm, red-channel) only, alpha-tubulin is used as a loading control. **b** | Bar graph shows the quantity of RAD21$^{FLAG}$ protein across the concentrations. Data is normalised to the alpha-tubulin loading control. Analyzed using iBright™ Analysis Software.

### Can RAD21 levels be restored following acute degradation?

If the re-accumulation of RAD21 occurred in the same order of time as the degradation, this could simulate a pulse-chase style experiment in which the wash out of dTAG-13 would be the equivalent of adding RAD21 back into the system. I therefore wanted to determine the time required for RAD21 to reach WT levels following acute depletion. To do so, cells were exposed to dTAG-13 for 24 hours to ensure full degradation, after 24 hours the culture media was removed and fresh culture media without dTAG-13 was added, and cells were harvested at 30 minutes, 1-hour, and then at 1-hour intervals for 6 hours. After 6 hours, the levels of $RAD21^{FLAG}$ remained extremely depleted compared to WT (Figure 3.9A-B), I therefore concluded that this would not be a feasible method to study RAD21 recruitment given the long timeframe required for re-accumulation.

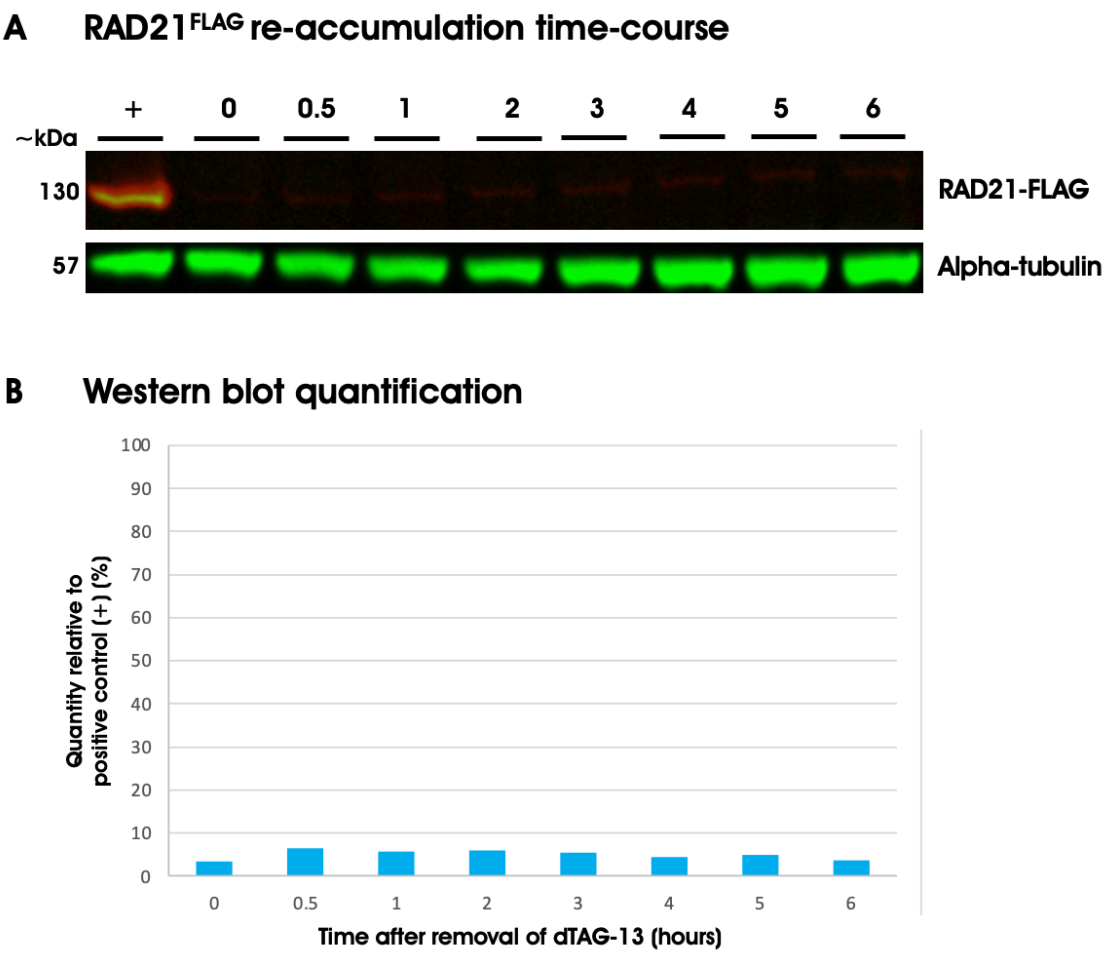



**Figure 3.9: RAD21 does not rapidly reaccumulate following dTAG-13 treatment.**

To assess RAD21 re-accumulation following degradation, $RAD21^{FLAG}$-dTAG cells were kept in media containing 500 nm dTAG-13 for 24 hours. After which, fresh media was added without dTAG-13 and whole cell lysates were extracted at timepoints from 0 to 6 hours. Positive control (+) is the $RAD21^{FLAG}$-dTAG cell line that has not been incubated in dTAG-13. Alpha tubulin was used as a loading control. **a** | Western blot shows that over a 6-hour time-course $RAD21^{FLAG}$ does not reach anywhere near physiological levels as shown in the positive control. **b** | Chart shows the quantity of $RAD21^{FLAG}$ protein throughout the re-accumulation time-course relative to the positive control and normalised to the a-tubulin loading control. Analyzed using iBright™ Analysis Software.

### 3.2.4 Visualising the degradation at the chromatin level

Ultimately, the aim of these experiments was to assay RAD21 occupancy throughout the degradation time-course on the chromatin level, therefore I needed to test if the pattern of degradation seen at the protein level by western blot was reflected at the chromatin level by ChIP-seq.

#### Testing the chromatin immunoprecipitation efficiencies of different antibodies

The homozygosity of the tagged cell line gives the ability to use either a RAD21 antibody or a FLAG antibody for immunoprecipitation of $RAD21^{FLAG}$, I therefore wanted to compare the different commercially available antibodies to determine which would achieve the most efficient immunoprecipitation. I tested a ChIP-grade RAD21 antibody (abcam, ab992) that we routinely use in the lab and that has been widely used in the literature, alongside two batches of FLAG-M2 antibody (Merck, F1804) that is widely used in the literature for IP experiments but not in ChIP-seq, and an anti-DYKDDDDK (FLAG) antibody (Bethyl, A190-102A) (Figure 3.10A). A routine step of using magnetic beads to purify the antibody-bound chromatin is applied in our lab's standard ChIP-seq protocol; magnetic beads pre-conjugated to FLAG-M2 are commercially available, therefore I also developed a modified protocol in which I used these beads to replace the individual antibody and bead steps (Figure 3.10A). In combining these two steps using pre-conjugated beads, I aimed to make the protocol more efficient as the number of required incubation steps was reduced, which should reduce the loss of material in the wash steps and from inefficient binding of the beads and antibody. The fully optimised protocol of both the regular ChIP-seq and magnetic FLAG bead ChIP are detailed in the Materials and Methods chapter.

### Quantifying chromatin immunoprecipitation efficiencies by qPCR

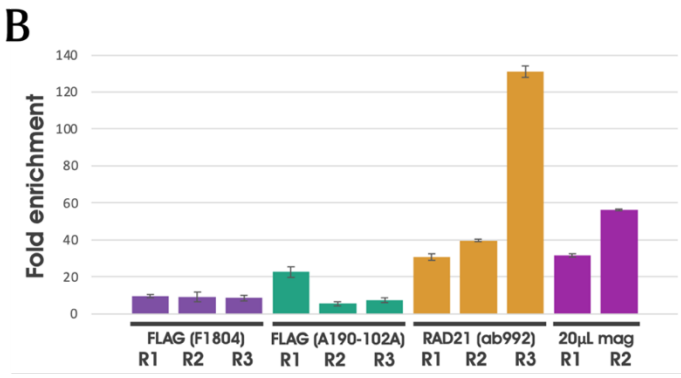
As a preliminary check to determine if the ChIP had worked before proceeding with next generation sequencing, I used ChIP-qPCR and measured fold enrichment of RAD21 at a site in the Nanog locus (chr6:122,646,563-122,646,667 (mm39)) compared to an intergenic background control (Figure 3.10B-C). Figure 3.10B shows the fold enrichment analysis from ChIP-qPCR experiment testing three replicates of each: anti-RAD21 (ab992), anti-FLAG M2 (F1804) and anti-FLAG (A190-102A), compared to two replicates of the 20 µL magnetic bead ChIP. Whilst the anti-RAD21 antibody showed the highest enrichment on average, across the 3 replicates there was also the greatest variation. This inconsistency in RAD21 ChIP-qPCR using this antibody has also been previously reported by other members of the group (data not shown). The highest fold enrichment was achieved using the 20 µL magnetic bead protocol, but again this was somewhat inconsistent across replicates. The anti-FLAG M2 (F1804) antibody gave the most consistent results, with 9-fold enrichment. The alternative anti-FLAG antibody was inconsistent among replicates ranging from a 5 to 11-fold enrichment. Replicates of the magnetic bead ChIP protocol (Figure 3.10C) showed a higher degree of consistency among replicates, apart from the first replicate of 20 µL beads for which there was a large standard error. From this set of ChIP-qPCR results, the anti-FLAG M2 antibody gave the highest and most reliable enrichment score of approximately 34-fold. Overall, this data confirmed that the ChIP of the $RAD21^{FLAG}$ was feasible and was able to achieve enrichment above background levels. However, this set of ChIP-qPCR experiments assayed enrichment at a single locus that is known to give a strong signal and may not be reflective of all RAD21 loci.

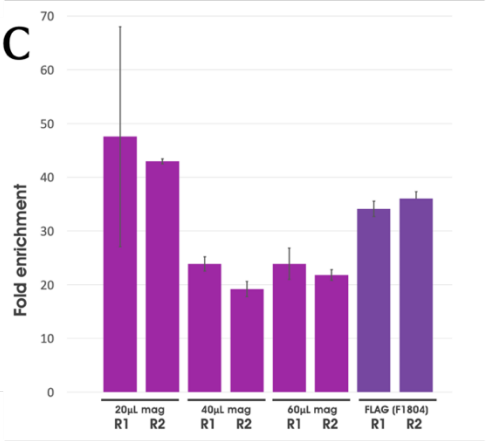
## Quantifying genome-wide chromatin immunoprecipitation efficiencies

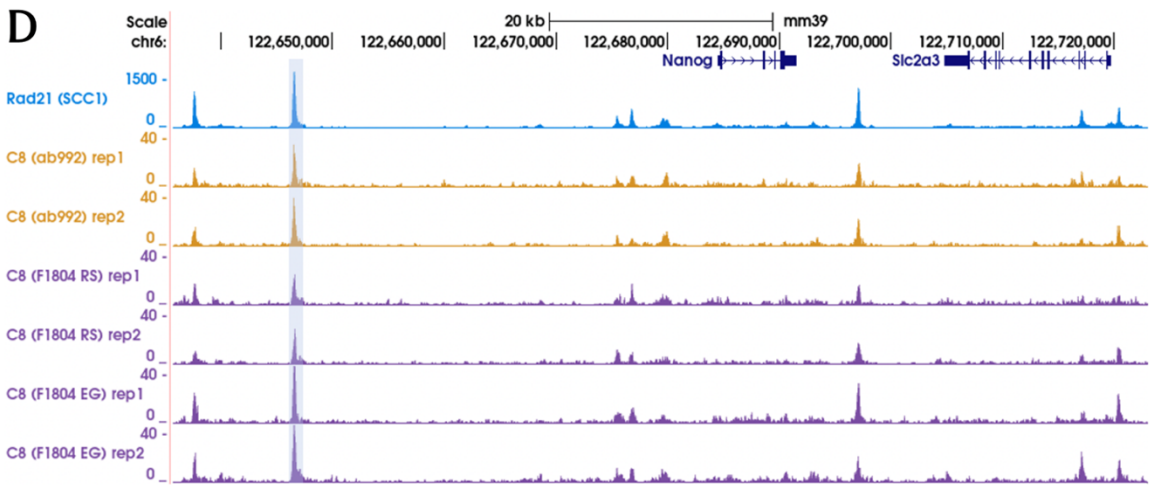
The aim of this chapter is to make a genome-wide assessment of RAD21 occupancy; therefore, it is important that the antibody and method of choice performs well across all loci on a genome-wide scale. To address this, I indexed the remaining DNA from this set of ChIP experiments to create libraries for Next-Generation Sequencing. The ChIP-seq tracks for the region containing Nanog, with the RAD21 loci used in the ChIP-qPCR experiment highlighted in blue, are shown in Figure 3.10D-E. By visual inspection at individual loci, it is again difficult to draw any conclusion about the effectiveness of each antibody on a genome-wide scale. Therefore, using a deeply sequenced RAD21 mES WT ChIP-seq data track, I used Lanceotron[180] to identify RAD21 peaks and then used this as a RAD21 reference peak track. Using deepTools[181] 'computeMatrix' and 'plotHeatmap' functionalities, I plotted the genome wide coverage across these reference RAD21 peaks in the ChIP-seq tracks from the different antibodies (Figure 3.11 and Figure 3.12). In doing so, it highlighted the stark differences between the antibodies and ChIP methods that were not immediately obvious from inspecting the bigWig tracks. From this analysis it seems that the magnetic bead method was suboptimal in comparison to both the RAD21 antibody (ab992) and both the FLAG antibodies, given the lower overall genome-wide signal. After assessing all the data from the ChIP-qPCR and ChIP-seq experiments, I selected the FLAG-M2 antibody (F1804) for use in future experiments.

A

| Antibody/beads | Supplier | Cat.no. | Lot no. (if known) | Plot colour |
|---|---|---|---|---|
| Monoclonal ANTI-FLAG® M2 (Anti-DDDDK) | Merck | F1804 | SLCF4933 (EG)<br>SLCC6485 (RS) | RGB |
| ECS (DYKDDDDK) Tag Polyclonal Antibody | Bethyl | A190-102A | 10 | RGB |
| Anti-RAD21 | abcam | ab992 | GR3310168-13 | RGB |
| Anti-FLAG® M2 Magnetic Beads (Anti-DDDDK) | ThermoFisher | M8823 | NA | RGB |



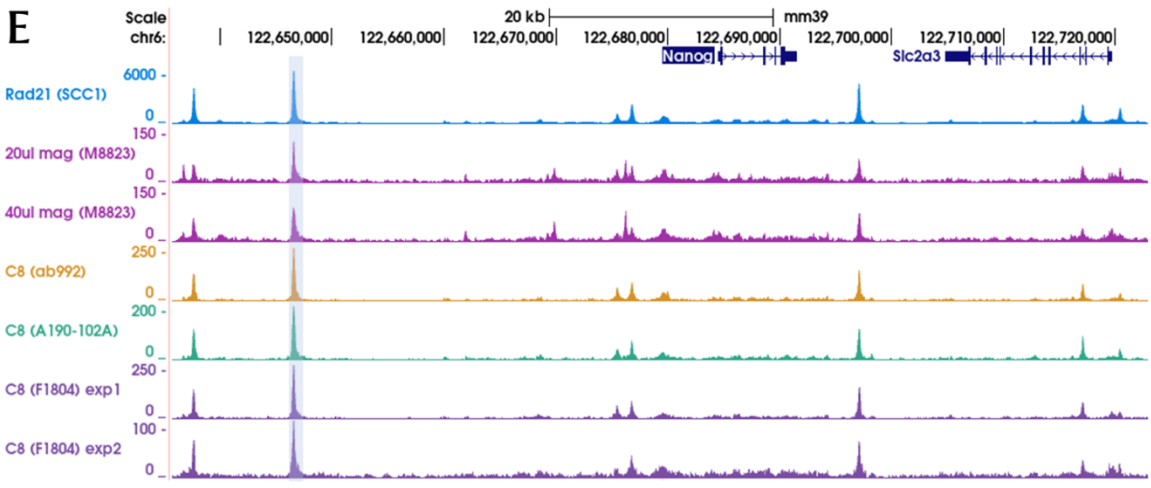



**Figure 3.10: Comparison of different antibodies for chromatin immunoprecipitation of RAD21$^{FLAG}$.** **a** | Table showing the different antibodies tested, including lot numbers if known and the colour used when plotting data. **b-c** | ChIP-qPCR data. Enrichment was calculated using the ΔΔCt (ChIP versus input controls) method and fold enrichment of a RAD21 peak over an intergenic background control was calculated. Either two or three technical replicates were performed per ChIP (R1, R2, R3). Three qPCR technical replicates were performed per ChIP replicate, error bars indicate +/- standard error (SE). **b** | Comparison of two different FLAG antibodies with RAD21 antibody using the regular ChIP-seq protocol, and the FLAG-M2 magnetic beads in the modified protocol. **c** | Different volumes of FLAG-M2 magnetic beads (M8823) used in the FLAG-M2 magnetic bead protocol compared with the standard ChIP protocol using the FLAG-M2 antibody (F1804). **d-e** | ChIP-seq antibody optimization. Region shown is chr6:122,635,703-122,723,045 (mm39). Peaks highlighted in blue (chr6:122,646,563-122,646,667 (mm39)) indicate the region amplified by the ChIP-qPCR experiment shown B and C. **d** | Comparison of ChIP experiments using native RAD21 antibody (ab992) with two different lots (EG/RS) of FLAG-M2 antibody. Blue track is the native RAD21 antibody used in a WT cell line for comparison. Orange and purple tracks are data from the RAD21$^{FLAG}$ cell line (clone #C8). **e** | Comparison of ChIP experiments using magnetic FLAG-M2 beads (pink) with two FLAG antibodies (purple/turquoise) and the native RAD21 antibody in WT cells (blue) and in the RAD21$^{FLAG}$ cell line (clone #C8) (orange).

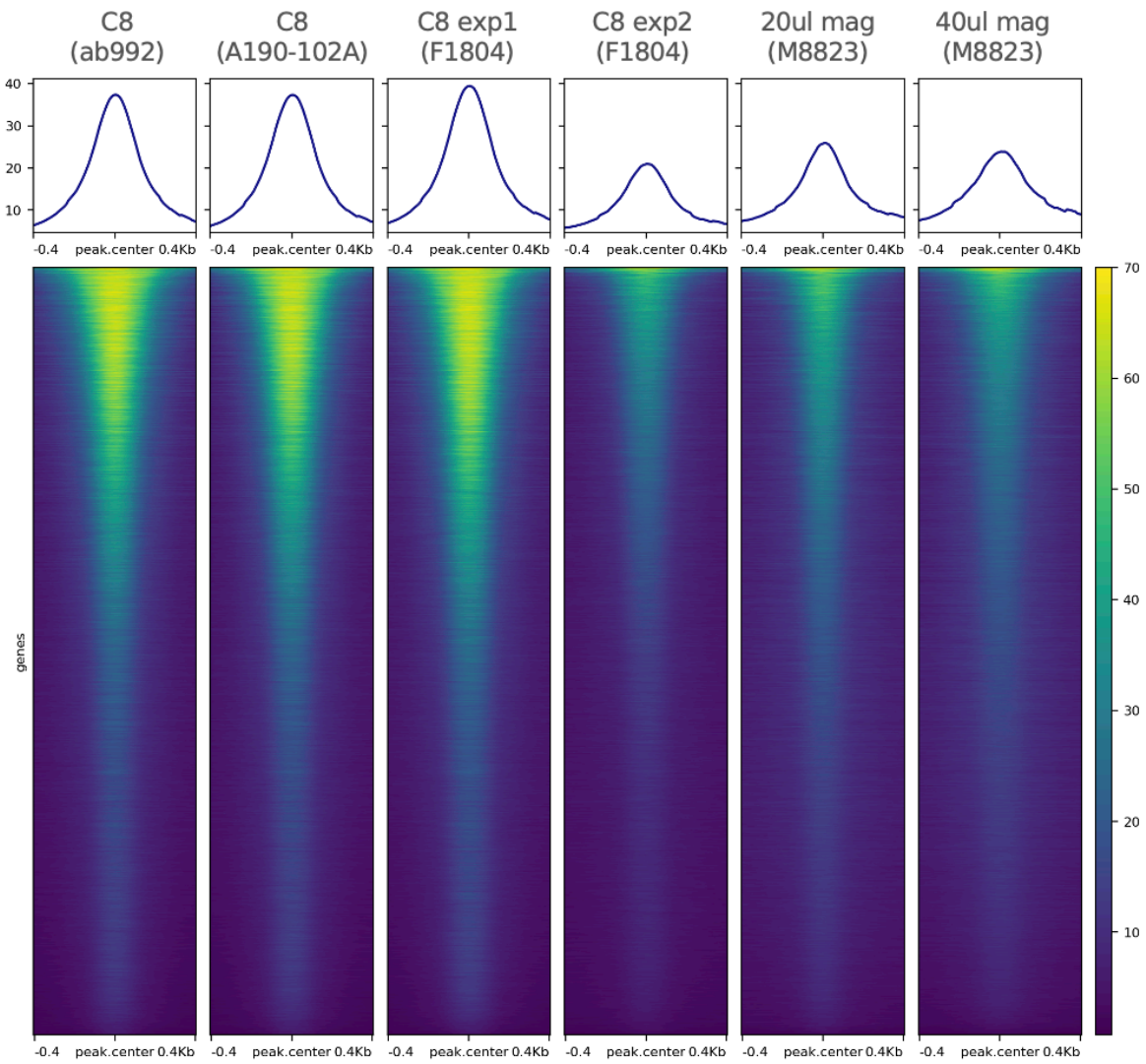



**Figure 3.11: ChIP-seq comparing different antibodies for RAD21-FLAG ChIP.**
Genome-wide analysis of ChIP-seq data. A deeply sequenced WT mES E14 RAD21 ChIP-seq track aligned to mm39 was peak called using Lanceotron[180] and used a reference BED file. Using deeptools (reference), matrices were computed for coverage at a +/- 400bp window around the centre of RAD21 ChIP-seq reference peaks in mES E14 cells. Heatmaps show genome-wide coverage across the reference RAD21 sites using different antibodies for the ChIP experiments performed in the dTAG-RAD21 cell line.

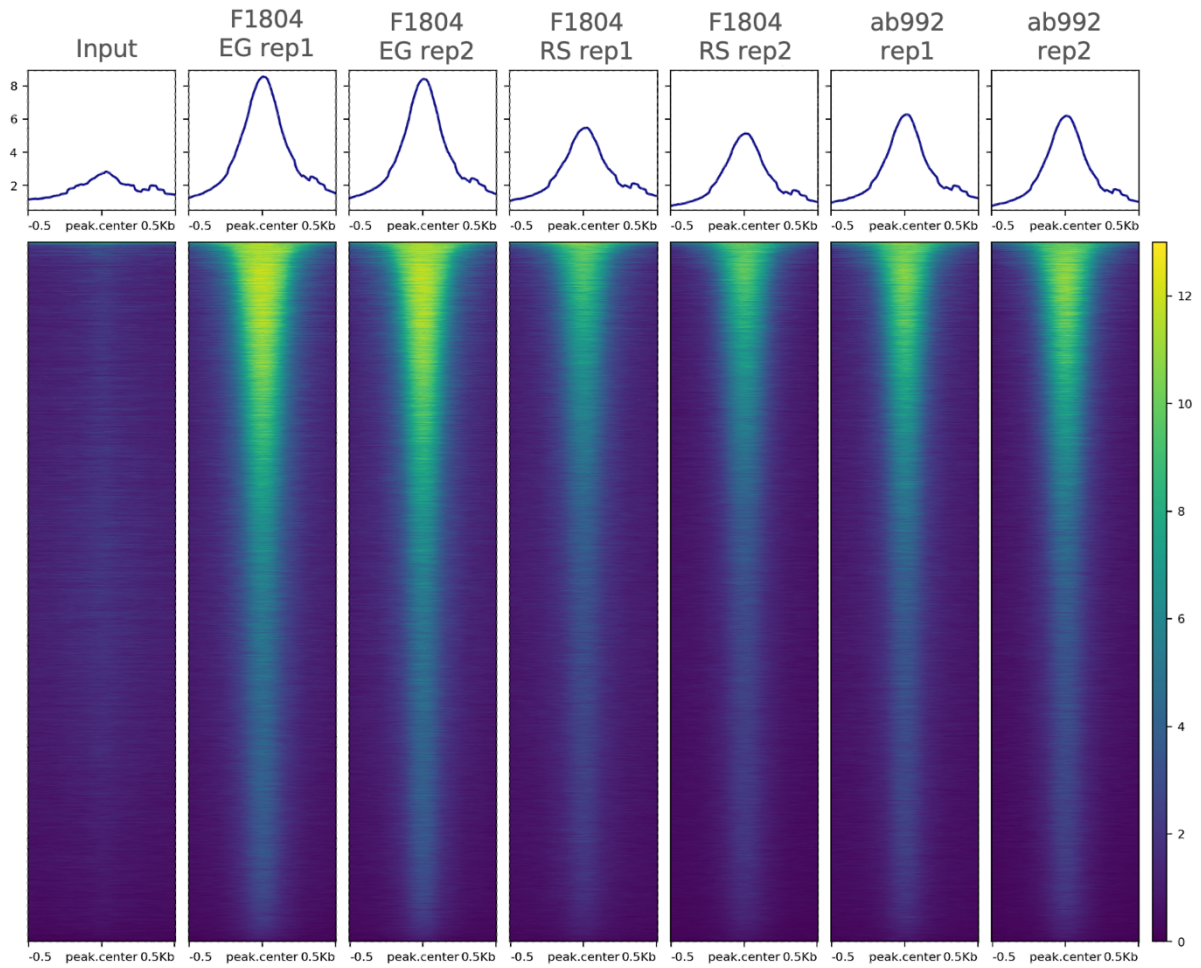



**Figure 3.12: ChIP-seq comparing different antibodies for RAD21-FLAG ChIP.**
Genome-wide analysis of ChIP-seq data. A deeply sequenced WT mES E14 RAD21 ChIP-seq track aligned to mm39 was peak called using Lanceotron[180] and used a reference BED file. Using deeptools (reference), matrices were computed for coverage at a +/- 500bp window around the centre of RAD21 ChIP-seq reference peaks in mES E14 cells. Heatmaps show genome-wide coverage across the reference RAD21 sites using different antibodies for the ChIP experiments performed in the dTAG-RAD21 cell line.

## Quantifying RAD21$^{FLAG}$ pre- and post-degradation

After selecting a suitable antibody for RAD21$^{FLAG}$ ChIP-seq, the next hurdle was to determine whether the degradation time-points that I tested and validated by western blot can also be captured using the ChIP-seq method. One of my main apprehensions was that given the increased sensitivity of ChIP-seq over western blot for detecting chromatin bound protein, ChIP-seq may still detect bound protein even when by western it appears to be fully degraded. Therefore, the first pass in optimising the ChIP-seq protocol was to test for the presence of RAD21$^{FLAG}$ before and after a timepoint by which degradation is seen to be complete by western blot. I chose to incubate the cells for 4 hours at 500 nM dTAG-13, as these were the conditions that by western blot showed complete degradation of RAD21 and this way I could make a direct comparison between the sensitivity of the two methods and determine if the results correlate. I ran three ChIP-seq replicates for uninduced (0 hours) and 4-hour dTAG-13 treatment, and then plotted the results using deepTools as before. Figure 3.13 shows the ChIP-seq tracks at three different loci, including the Nanog loci previously shown. There is a significant amount of background signal seen in these ChIP-seq tracks, making it difficult to distinguish by eye whether there is any difference between them. However, when these are plotted as a genome heatmap (Figure 3.14), a clear difference can be seen between the pre and post degradation samples indicating that the degradation is effective in reducing the amount of chromatin bound RAD21$^{FLAG}$. An important observation from these data is that RAD21$^{FLAG}$ signal is still seen at 4 hours suggesting that this is not a sufficient amount of time to fully deplete the pool of RAD21$^{FLAG}$ as was thought to be the case by western blot.

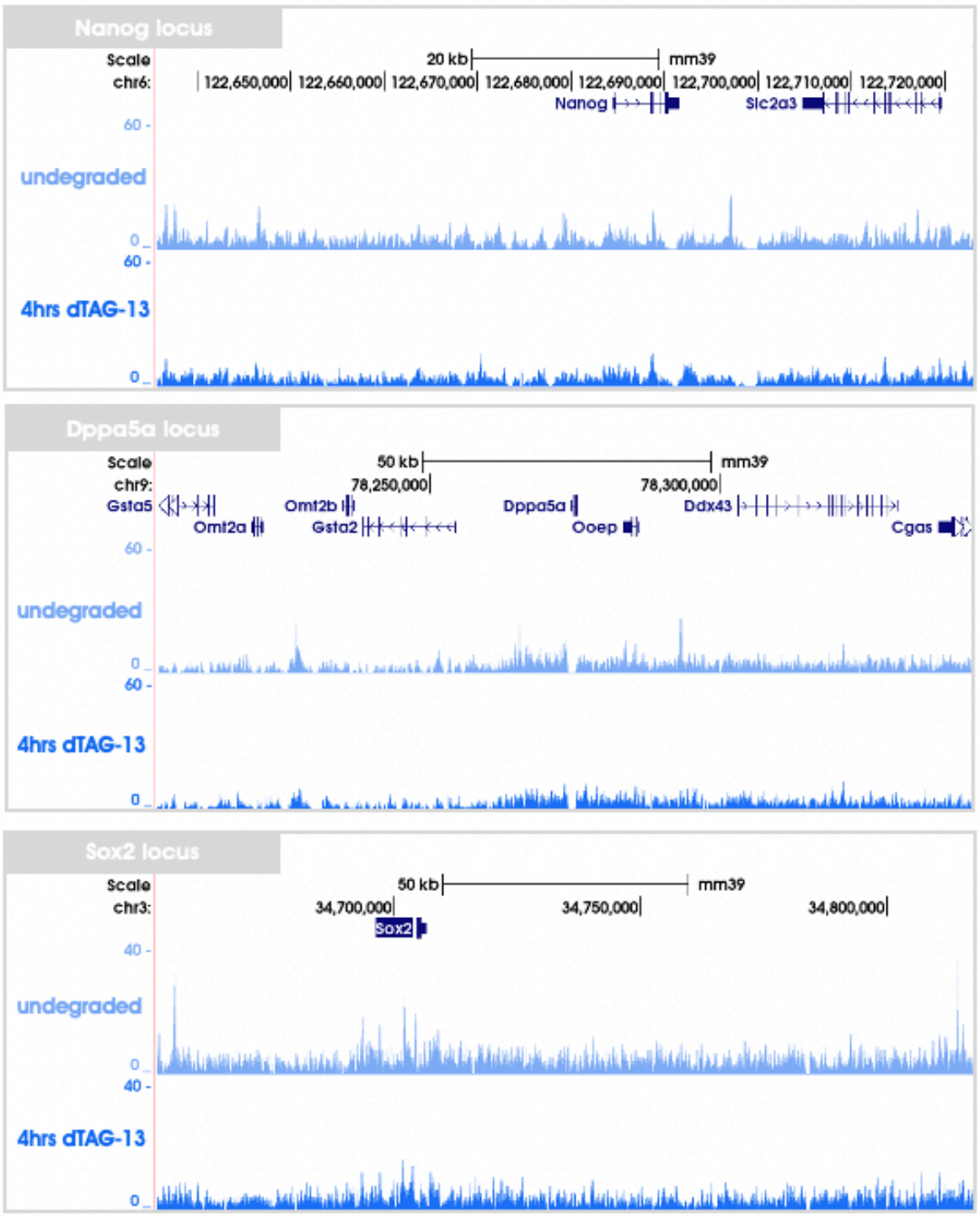



**Figure 3.13: RAD21$^{FLAG}$ ChIP-seq pre and post dTAG-13 treatment at three active loci in mouse embryonic stem cells.**

Using the dTAG RAD21$^{FLAG}$ cell line, RAD21 was degraded for 4 hours using dTAG-13 and ChIP-seq performed in these treated cells and an undegraded control. FLAG antibody (F1804) used for chromatin immunoprecipitation. Each track represents three merged replicates. **a** | Nanog locus, chr6:122,635,703-122,723,045 (mm39) **b** | Dppa5a locus, chr9:78,202,939-78,344,188 (mm39) **c** | Sox2, chr3:34,651,835-34,817,734 (mm39).

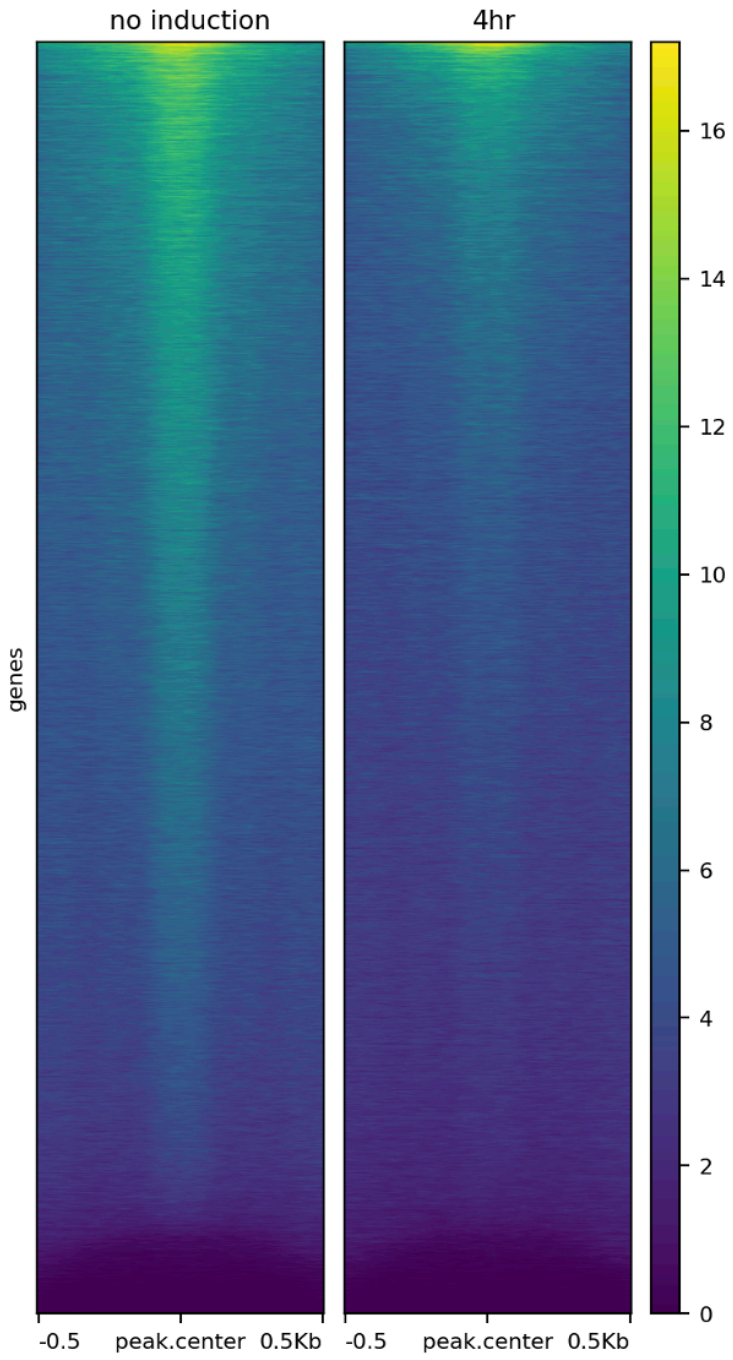



**Figure 3.14: RAD21 is degraded over a 4hr period.**

Using deepTools[181], matrices were computed for coverage at a +/- 500bp window around the centre of $RAD21^{FLAG}$ ChIP-seq reference peaks in mES E14 cells. Heatmaps show a genome-wide summary of $RAD21^{FLAG}$ ChIP-seq peaks before and after a 4-hour incubation with dTAG-13, demonstrating there is a significant but incomplete degradation of chromatin-associated $RAD21^{FLAG}$ over the 4-hour period.

### 3.2.5 Monitoring RAD21$^{FLAG}$ degradation over a 4-hour time-course using Calibrated ChIP-seq

Although RAD21$^{FLAG}$ was incompletely degraded genome-wide using 4-hour treatment of dTAG-13, the degree of degradation was sufficient to further divide this timeframe and begin to investigate the patterns of degradation. I harvested cells prior to treatment (0-hours), and at three time-points following the addition of dTAG-13: 1-hour, 2-hours and 4hours, and fixed for ChIP. From the ChIP-seq data I aimed to answer the following questions:

- Which sites most rapidly lose cohesin?
- Which sites (if any) retain cohesin?
- Are there differences in degradation rates?
- Do these patterns correlate with any other genomic features such as *cis*-regulatory elements or TAD boundaries?

These questions can be addressed by considering the rate of change of RAD21 occupancy and therefore do not require RAD21$^{FLAG}$ to be completely degraded, which from the previous data I did not expect that by 4-hours it would be. To make a direct comparison between the ChIP-seq data from each time-point I calibrated the ChIP-samples to a 1% spike-in of human chromatin using HEK293 cells (details in Materials and Methods).

Figure 3.15A-B shows calibrated ChIP-seq data at two loci for each of the 4 timepoints used in this experiment. Between 0 and 2 hours, depletion of RAD21$^{FLAG}$ is clearly visible by eye, between the 2 and 4-hour timepoints however there is little difference. This is further confirmed when plotting a genome-wide heatmap using deepTools[181], as shown in Figure 3.16. Plotting the degradation in this way shows that in fact there is a marginal increase in RAD21$^{FLAG}$ signal at 4-hours, rather than a continuation of the reduction as expected.

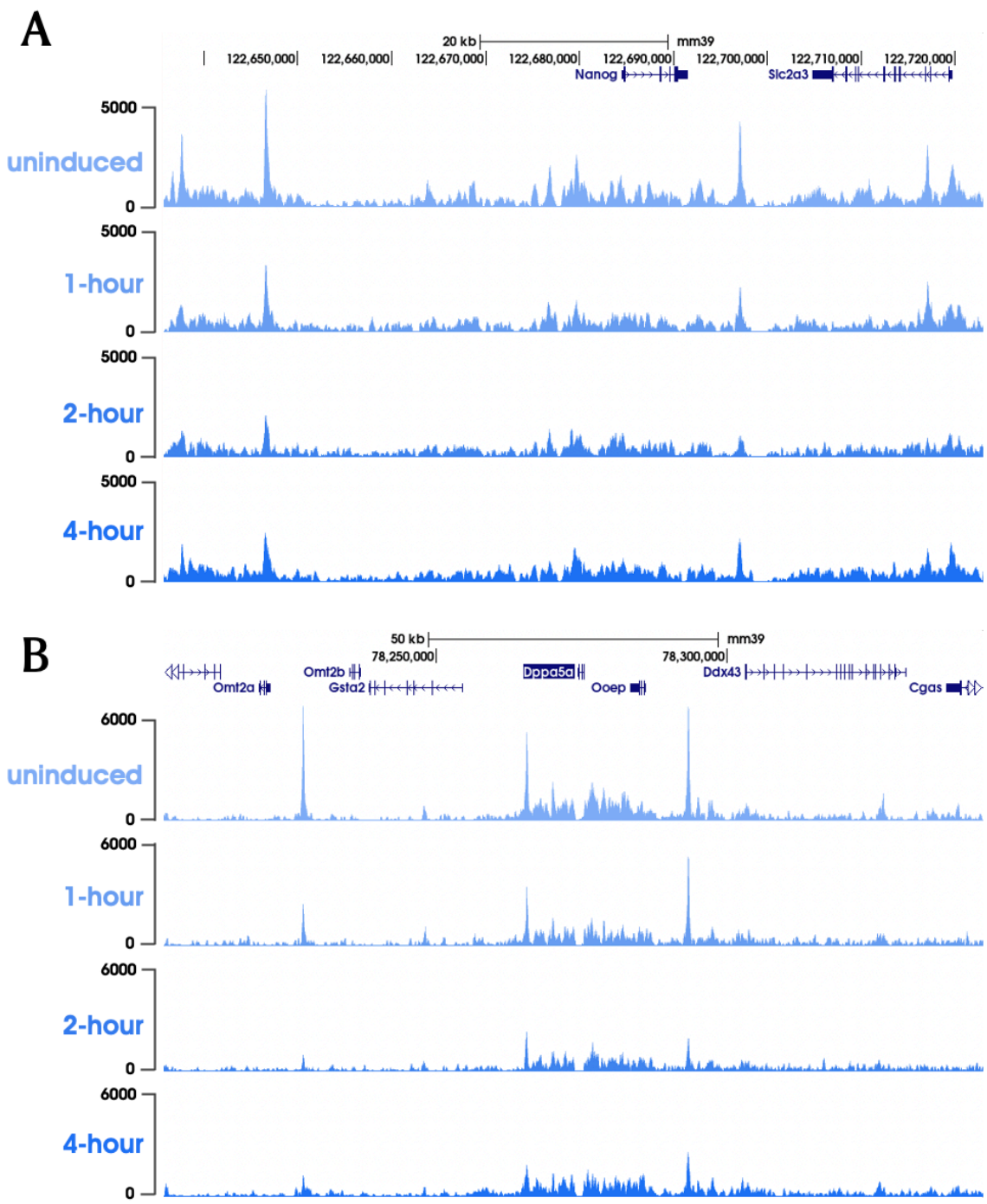



**Figure 3.15: RAD21 is degraded over a 4-hour period.**

Using the dTAG RAD21$^{FLAG}$ cell line, RAD21 was degraded over a 4-hour time-course using dTAG-13 and cells were fixed for ChIP at 0 (uninduced), 1, 2 and 4 hours. Each sample was calibrated using a 1% spike-in of HEK293 chromatin. FLAG antibody (F1804) used for chromatin immunoprecipitation. **a** | Nanog locus, chr6:122,635,703-122,723,045 (mm39). **b** | Dppa5a locus, chr9:78,202,939-78,344,188 (mm39).

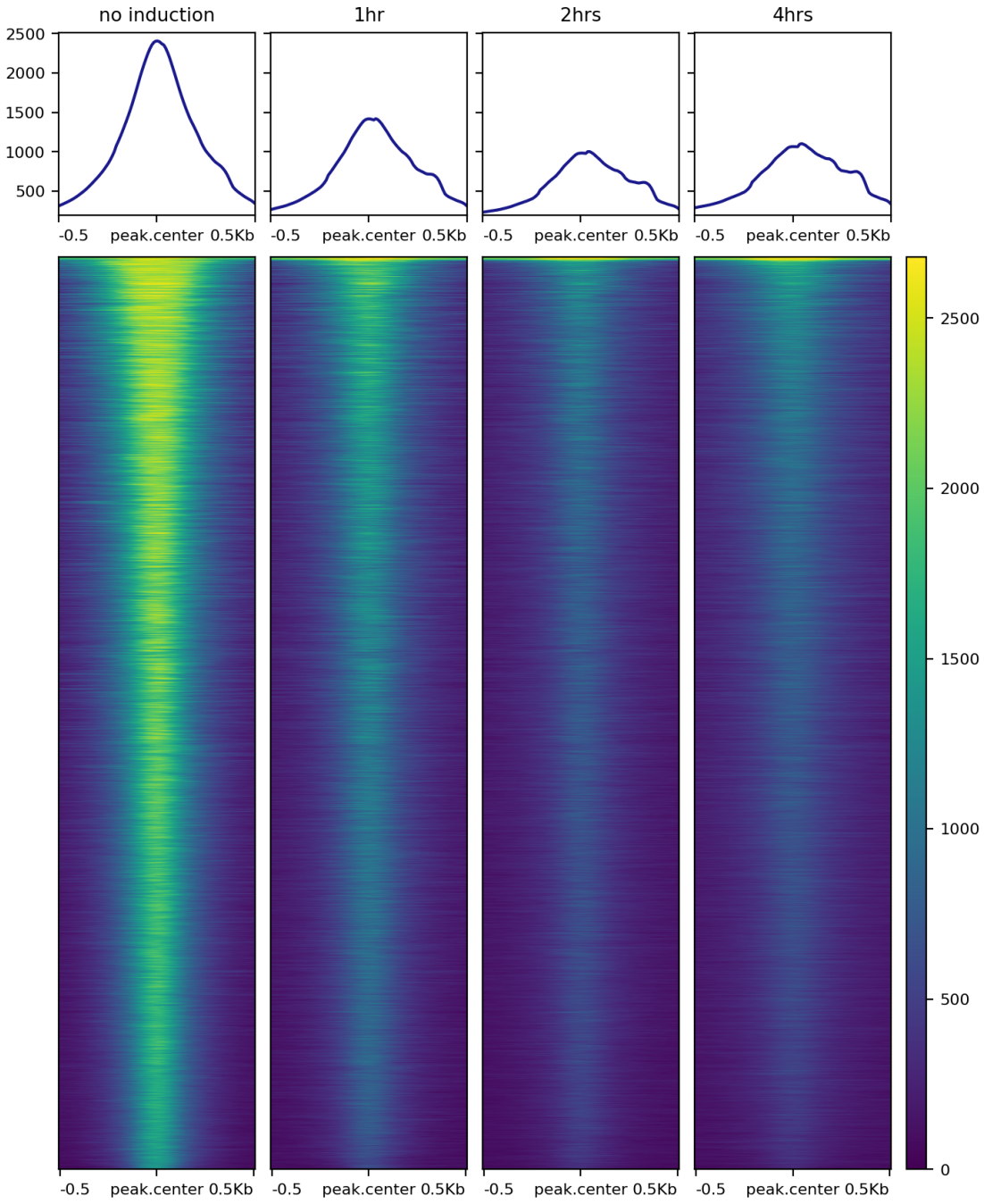



**Figure 3.16: Heatmap of genome-wide RAD21$^{FLAG}$ signal during a 4-hour degradation time-course.** Regions plotted were obtained by peak-calling the uninduced track using Lanceotron[180]. Deeptools[181] was used to plot an average peak profile for each time-point (top panel) and a genome-wide heatmap of coverage +/- 500 bp from the centre of the RAD21$^{FLAG}$ sites at each of the timepoints (bottom panel).

### 3.2.6 How does the distribution of RAD21$^{FLAG}$ change throughout the degradation time-course?

The first analysis I wanted to do was to assess the distribution of peaks across the timepoints. To do this I used Lanceotron to peakcall the ChIP-seq data tracks from each timepoint. Given that RAD21 ChIP-seq signal can be spread across the genome due to its ability to translocate along the chromatin, it will not necessarily form punctate peaks. As a result there may be regions that are occupied by transiting RAD21, that would ordinarily be filtered out using the standard peak score threshold of ≥0.5, therefore a peak score ≥0.1 was chosen as a lower threshold to filter out background noise but capture transiting RAD21 signal. Figure 3.17A shows a bar chart of the total number of peaks called, with those scoring ≥ 0.1 shown in purple. The total number of peaks decreases from 0 to 2 hours and then slightly increases at 4 hours. At 0 hours the proportion of peaks scoring ≥ 0.1 was 21.7%, this proportion decreased for the 1-, and 2-hour timepoints. However, between 0 and 1 hours there was an increase in peaks scoring < 0.1 suggesting that cohesin may be depleted first from the sites with higher cohesin occupancy and therefore highest peak scores. At 4 hours the number of peaks scoring < 0.1 differed from 2-hours by 836 peaks, whereas the number of peaks scoring ≥ 0.1 increased by 2667.

### 3.2.7 Can RAD21$^{FLAG}$ peaks be categorised based on the degradation trajectory?

Next, I wanted to investigate the pattern of degradation across the timepoints. Using the peak calls from all four timepoints filtered by peak score ≥ 0.1, I applied the bedTools 'multiinter' function to combine all peaks into a union BED file and then annotated each peak by the timepoint it was called in. Using the output from this, I was then able to classify the peaks by speed of degradation:

- Fast: peak only present at 0 hours
- Medium: peak present at 0 and 1 hours

- Slow: peak present at 0, 1 and 2 hours
- Persistent: peak present across all timepoints

Figure 3.17B shows a bar chart of the number of peaks belonging to each category. As expected from all the degradation data shown so far, the initial rate of degradation is fast. Therefore, it is not surprising that there are the most peaks in the 'fast' degrading category. My hypothesis introduced at the start of this chapter (referring back to Figure 3.1) was, if active enhancers recruit cohesin, these would be the first peaks to degrade, whereas CTCF insulating elements would retain cohesin for the longest as these are the sights of cohesin accumulation. Therefore, my next aim was to determine if there was a correlation between the classes of *cis*-regulatory elements (CREs), i.e., enhancers, promoters or CTCF sites, that intersect RAD21 peaks and the speed of degradation. To do this I manually classified the RAD21 peaks based on their intersection with ChIP-seq data from histone modifications characteristic of enhancers and promoters, and CTCF (Figure 3.18A). Out of the total peaks, 13,482 intersected with a CRE, and were used in the percentage analysis shown in Figure 3.18B. Firstly, when considering all RAD21 peaks intersecting CREs (labelled 'All'): just over 90% of these peaks intersected with pure CTCF sites (i.e., CTCF sites that were devoid of H3K27Ac, H3K4me1 and H3K4me3), approximately 8% of peaks intersected with enhancer-CTCF sites, and the remaining peaks were at enhancers, promoters and promoter-CTCF sites. This category 'All', can be considered as the baseline as this is taking into account any site where cohesin can reside, I was therefore interested in any deviation from this baseline distribution. The fast-degrading peaks contained a greater proportion of enhancer-CTCF peaks compared to baseline, whilst maintaining a similar proportion of distinct enhancers and CTCF peaks. Interestingly, the proportion of promoter peaks decreased. The 'medium' class showed a shift in which there was a greater proportion of CTCF peaks, which was maintained between the 'medium', 'slow' and 'persistent' classes. Finally, the 'slow' and 'persistent' classes were completely lacking any peaks intersecting promoter-related marks and only a

miniscule fraction of pure enhancer peaks, the majority (~92 %) coincided with pure CTCF sites, and the remainder predominantly intersected enhancer-CTCF sites. Overall, this analysis suggests that promoters are the sites that cohesin occupies most dynamically, closely followed by enhancers. CTCF sites however are able to maintain cohesin binding, suggesting that cohesin is more stably bound at these sites, or accumulates there prior to unloading. Whilst this analysis provided some interesting results, there were some obvious limitations: firstly, and as already mentioned, the classification method for enhancers and promoters is not straightforward. My choice of stringent cut-offs for chromatin mark and CTCF ChIP-seq coverage does not reflect the ambiguity that is often associated with CRE classification, I will therefore have undoubtedly ruled out some CREs from my analysis. I also chose not to further sub-classify the CREs into active, primed, or inactive states; these states add another level of complexity and are not straight forward to define, I therefore did not want to risk further reduce the number of *cis*-regulatory elements in my analysis. Consequently, this can be seen as a preliminary analysis that can be further refined through the application of more advanced classification methods.

The result that CTCF sites preferentially retain cohesin was a finding I wanted to investigate further. It has been previously reported that there may be two pools of cohesin[137]: the dynamic pool that is dependent on WAPL, and the stable pool that is dependent on CTCF. Whilst my work cannot speak for the relationship of cohesin and WAPL, my results are suggestive of the existence of a stable CTCF-associated class of cohesin. If this is the case, it would beg the question why there is the requirement for a stable class of cohesin. My idea was that perhaps the persistent cohesin-CTCF sites play a role in the maintenance of the larger scale TAD structure. To test this hypothesis, I aligned the peak calls of the persistent peaks with Hi-C data in mESC using Juicebox[182] to determine if there was an alignment of TAD boundaries and persistent RAD21 peaks (Figure 3.19). The persistent peaks are distributed across the genome in a pattern that

mirrors the TAD structures, with the majority of persistent peaks residing at regions of increased contact frequencies. As a reference I also plotted the total RAD21 (blue) and CTCF (yellow) peaks in Figure 3.19 to demonstrate that this distribution is specific to this subset of persistent RAD21 peaks.

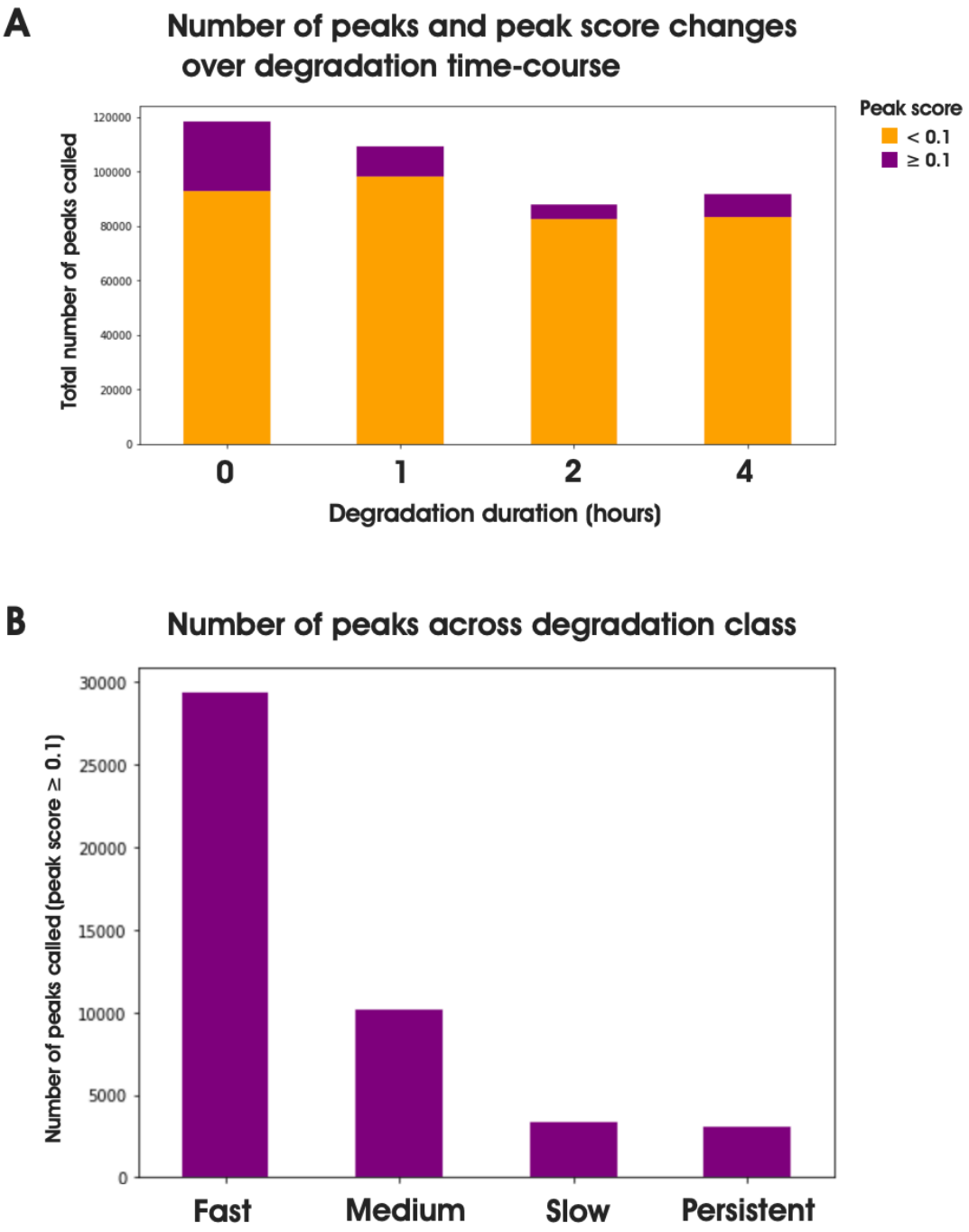



**Figure 3.17: Quantification of peaks during RAD21 degradation time-course.**

**a |** Number of peaks and peak-score changes over the 4-hour degradation time-course. ChIP-seq tracks were peak-called using Lanceotron and filtered by peak-score. Peaks scoring < 0.1 are shown in yellow, peaks scores of ≥ 0.1 are shown in purple. **b |** Bar plot showing the number of peaks per degradation class. Peaks were filtered for peak score ≥ 0.1, and peaks classified depending on if a peak was called at that timepoint (fast: 0 hours, medium: 0 and 1 hours, slow: 0, 1 and 2 hours, persistent: all).

**A** Classification of cis-regulatory elements

| Classification | H3K4me1 (enhancer) | H3K4me3 (promoter) | H3K27Ac (enhancer) | CTCF |
|---|---|---|---|---|
| CTCF | x | x | x | ✓ |
| Enhancer-CTCF | ✓ | x | ✓ | ✓ |
| Enhancer | ✓ | x | ✓ | x |
| Promoter | x | ✓ | x | x |
| Promoter-CTCF | x | ✓ | x | ✓ |

**B** Distribution of cis-regulatory elements across degradation classes.

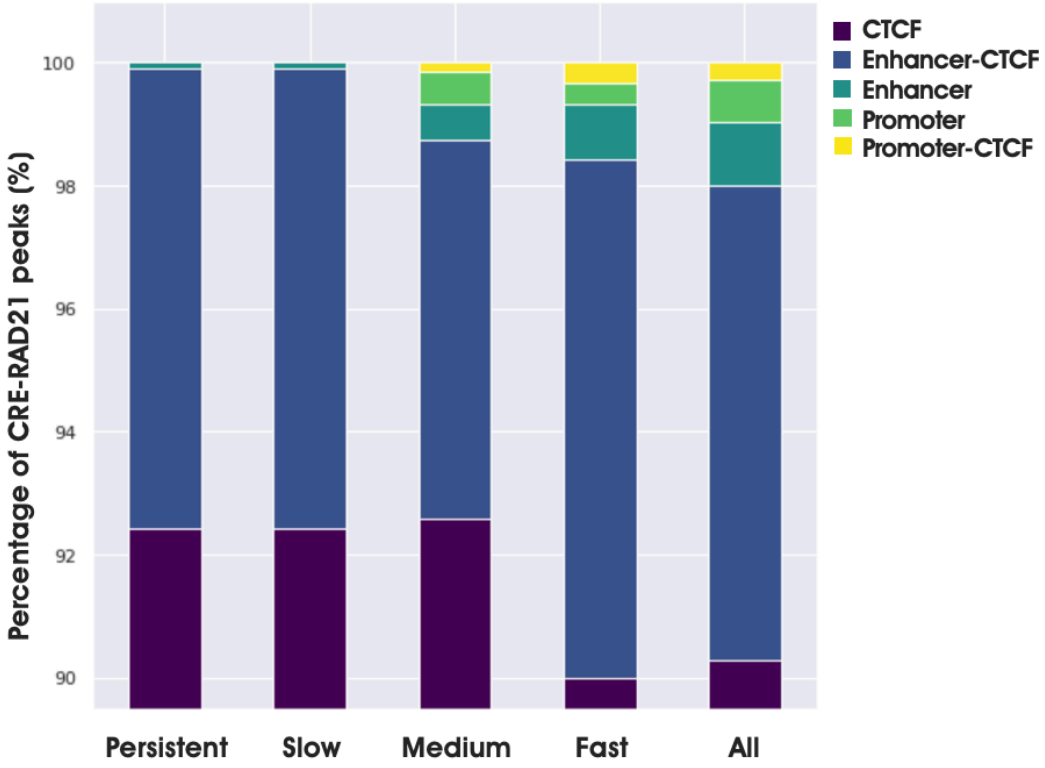
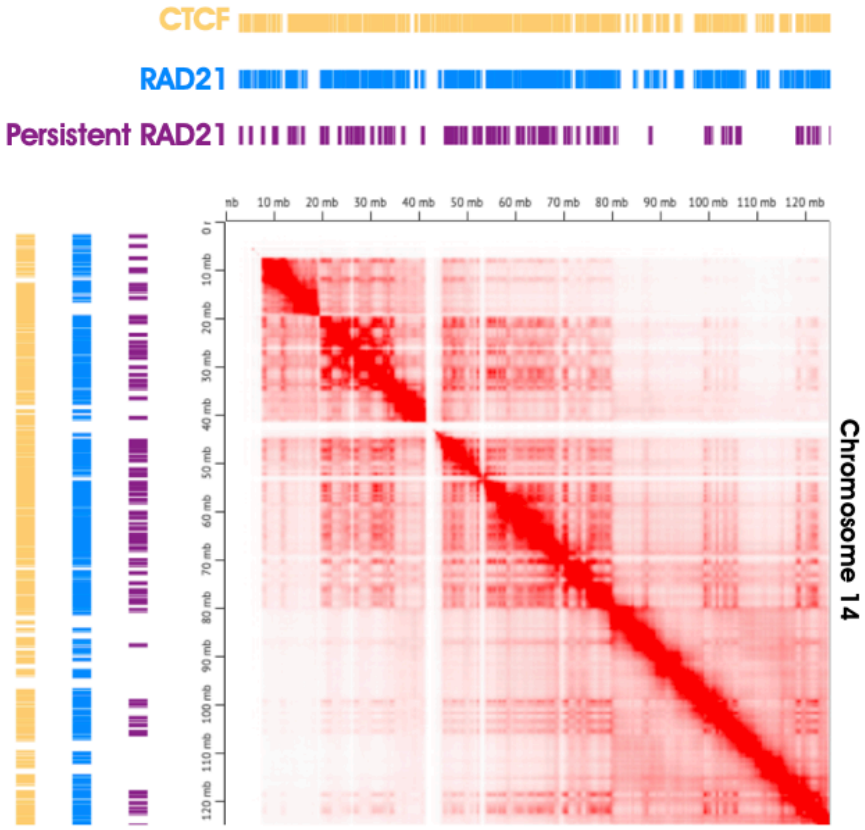



**Figure 3.18: Different classes of degradation dynamics, are enriched for peaks that intersect with different classes of *cis*-regulatory elements (CREs).**

**a** | Table showing the general classification of *cis*-regulatory elements used in this analysis. For histone-mark ChIP-seq (H3K4me1, H3K4me3 and H3K27Ac), an arbitrary cut-off of coverage < 10 was applied to classify peaks as not containing this mark. For CTCF ChIP-seq an arbitrary cut-off of coverage < 20 was applied to classify peaks as not containing this mark. Full details and ChIP-seq data sources can be found in Materials and Methods **b** | Once all peaks had been classified based their intersection with *cis*-regulatory elements, each class of degradation dynamics (persistent /slow/medium/fast/all) were plotted based on the percentage of peaks per class that fell into each CRE class.

**A** Intersection of Hi-C data with RAD21 and CTCF ChIP-seq peaks at chr14.




**B** Intersection of Hi-C data with RAD21 and CTCF ChIP-seq peaks at chr6.

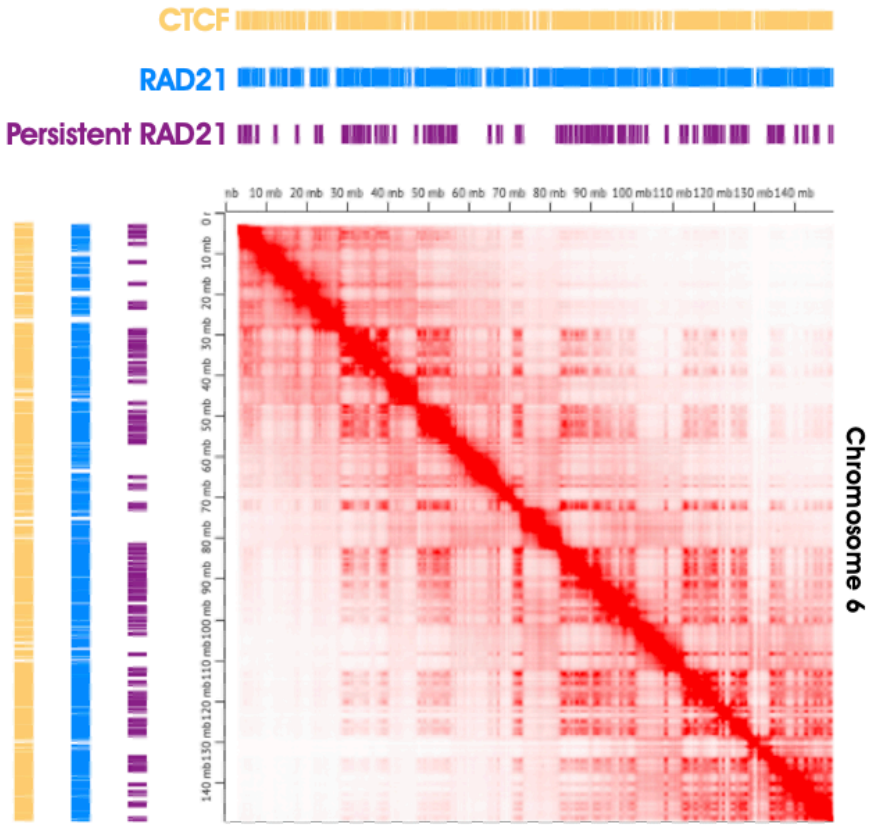


**Figure 3.19: Intersection of mESC Hi-C data (Bonev et al.) with RAD21 and CTCF ChIP-seq peaks, demonstrating the alignment of RAD21 persistent peaks with TAD structures.**

Hi-C data from Bonev *et al.*[183] generated in mESC was visualised using Juicebox[182] and aligned with BED files of peaks called in CTCF (yellow) and RAD21 (blue). Persistent RAD21 peaks are shown in purple. **a** | Region plotted: chromosome 14. **b** | Region plotted: chromosome 6.

## 3.3 Discussion

The aim of this chapter was to make use of a well-established degron technology and use this to gain an insight into the nature of the cohesin dynamics. Throughout the duration of the time taken to complete this thesis, there have been studies published which use the dTAG technology to degrade cohesin and study the consequences of the depletion on the 3D organisation of the genome[131,184]. To my knowledge, no studies thus far have used this technology and taken a time-course approach to probe the differential, and dynamic role of cohesin in the regulatory genome. This chapter has shown that the stability of cohesin binding differs across the genome, and most likely co-segregates with specific *cis*-regulatory elements. My findings agree with previous studies that suggest there are specific CTCF-associated cohesin binding sites that are less dynamic and more stable. Through the intersection with Hi-C data, my data suggests that these stable sites are linked with the overarching TAD structure.

One of the main limitations of the existing studies that use degron technologies to study the effect of specific proteins on the regulation of genome architecture, is the potential confounding effect of incomplete degradation, either through insufficient degradation of existing protein, or the appearance of newly synthesised protein. The assumption throughout this chapter was that if cells were maintained in media containing dTAG-13, any newly synthesised protein will be degraded in the same manner as non-chromatin bound protein. The results shown in the reaccumulation experiment (Figure 3.9), would even suggest that there is sufficient dTAG-13 maintained either within the cells, or accumulated within the hydrogel coating of the culture flasks, to prevent the reaccumulation of protein even after replacing the culture media. This would suggest that newly synthesised protein will also be degraded. By using a time-course approach I have

circumvented the issue of incomplete degradation: by assessing the changes between time points, complete degradation is not required as the main focus is to assess the rate of change in cohesin occupancy across the genome throughout the degradation process, rather than making inferences on genomics changes before and after degradation. Accordingly, peaks of cohesin that persist throughout the degradation are those that have remained associated with the chromatin throughout the time course, rather than those that have been reloaded with newly synthesised cohesin. If newly synthesised cohesin was able to reload onto the chromatin, rather than being immediately degraded, it is unlikely that a change in cohesin occupancy would have been observed throughout the time course.

To gain an understanding of how cohesin works to regulate 3D genome organisation and whether it is responsible for the instruction of domain formation, it would be informative to know which sites, if not a random process, is cohesin first loaded onto the chromatin. If this could be determined, it would answer the question of how the folding of the chromatin is instructed and potentially provide a mechanism for tissue-specific gene regulation coordinated by loop-extrusion. The current understanding in the field is that there is a correlation between patterns in 3D genome architecture and gene transcription, and that cohesin mediated loop extrusion is one mechanism by which tissue specific enhancers and promoters can be encouraged to interact. It is currently unknown whether the recruitment of cohesin is a cause or a consequence of domain activation, such as that shown in Chapter 2. In the next chapter I aim to develop a method capable of identifying the precise genomic location of cohesin loading, in an attempt to begin to answer this question.

## 3.4 Acknowledgements

Philip Hublitz (WIMM, Genome Engineering Services) designed and produced the constructs for degron tagging RAD21.

# 4 Genome-wide methods to study dynamics using RAPID-release

## 4.1 Introduction

The work presented in previous chapters has provided evidence for the link between active regulatory elements and the directed recruitment of cohesin and the establishment of domain structure. Using a cohesin degron model I have shown that, on a genome-wide scale, sites enriched with cohesin behave differently with respect to cohesin retention. In this chapter I aimed to determine the precise sites at which cohesin is first recruited, in order to disentangle the recruitment and migration processes. To do so I set out to develop a novel genomic approach capable of tracking cohesin recruitment on a genome-wide scale over a time-course until an equilibrium point is reached.

The current understanding in the field of cohesin dynamics is that upon binding the chromatin, cohesin is able to translocate along the chromatin[95,185-187] until it reaches a boundary element such as CTCF. These boundary elements can therefore become sites of cohesin accumulation, given that this is not counteracted by cohesin unloading. The method used in this chapter is designed to capture snapshots throughout this process, to identify where cohesin is loaded, and the subsequent path of translocation until the

system reaches equilibrium. This is achieved through the combination of an induction method, to flood a uniquely identifiable form of cohesin into the nucleus, with ChIP-seq at set intervals throughout the induction process, building on the work in the previous chapter. Provided that cohesin translocation is not too fast to capture, my hypothesis is that that peaks would first appear at sites of cohesin recruitment, which over time signal would spread between the CTCF sites where cohesin accumulates until unloaded (see Figure 4.1). To determine the specific types of genomic element that directly recruit cohesin, ChIP-seq data will be peak called, peaks classified based on their dynamic behaviour and then intersected with chromatin annotations.

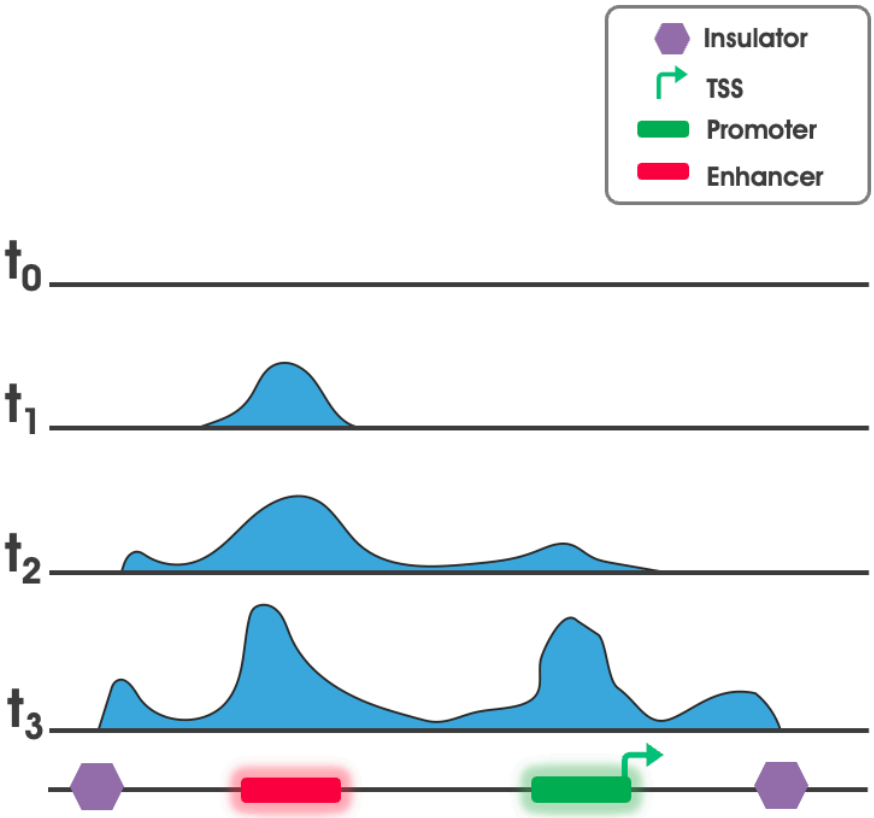



**Figure 4.1: Prediction of the calibrated ChIP-seq time-course for induction of RAD21 if cohesin is loaded at the enhancer.**

Figure shows a locus containing an active enhancer (magenta rectangle) and an active promoter (green rectangle) situated between two insulator sites (purple hexagons). The hypothetical RAD21 ChIP-seq signal across the timepoints $t_0$-$t_4$ is shown in blue. Prior to induction the RAD21 will be tethered to the mitochondria, therefore at $t_0$ there will be no RAD21 bound to the chromatin. Signal

will be first observed at the locations where RAD21 is initially recruited, which I hypothesise to be at active enhancers. I then expect that over the duration of the time-course, RAD21 signal will spread out until it reaches an insulator element which acts as boundary, preventing further translocation of RAD21. The final time-point represents equilibrium, that is the ChIP-seq signal that would be seen in a regular ChIP-seq experiment.

### 4.1.1 RAPID-release: a system to control cohesin influx

There are a number of pre-existing methods to study protein localisation by imaging e.g. FRAP and FLIP, however they have limited applicability in rapid processes given the maturation times of fluorescent processes often exceeds the rate of protein dynamics[188]. In a similar way, methods such as SNAP-tag, would not be feasible for studying rapid dynamic processes given the time needed for the quenching, pulsing and labelling steps. In order to investigate cohesin a system is required that can be rapidly induced, and the progress precisely monitored. The inducible system selected for use in this chapter is called Rapamycin Activated Protease through Induced Dimerization (RAPID)-release, a system that was originally developed by the Bowman lab[189] to investigate the nuclear import of histone proteins H3.1 and H4 by microscopy (Figure 4.2). By genetically engineering a cell line to express the protein of interest (POI) tagged with the RAPID-release machinery, the induction is a simple, single step process. RAPID-release relies on the well characterised Rapamycin induced dimerization of FK506 binding protein (FKBP) with the FKBP12-Rapamycin-binding (FRB) domain of mTOR. The novelty of RAPID-release is the FRB domain fusion with OMP25 which tethers the tagged POI to the outer mitochondrial membrane. Thus, the protein is held in an inactive state in the cytoplasm, whilst the FRB fusion protein is untethered and localised in the cytoplasm. Upon addition of Rapamycin, the FRB and FKBP domains dimerise, resulting in the activation of the auto-inhibited TVMV protease consequently cleaving the POI from the mitochondrial tether. The subsequent nuclear localisation of the POI, driven by an intrinsic nuclear localisation signal (NLS) is rapid (in the order of minutes), making it an ideal system to study dynamic

processes that occur over a short timeframe. In the original design, the two halves of the system are individually fluorescently labelled: EGFP$^{TVMVx2}$-FRB-OMP25 and mCherry-FKBP12-TVMV-AI. This dual colour system enables the induction process to be easily tracked by live-cell imaging.

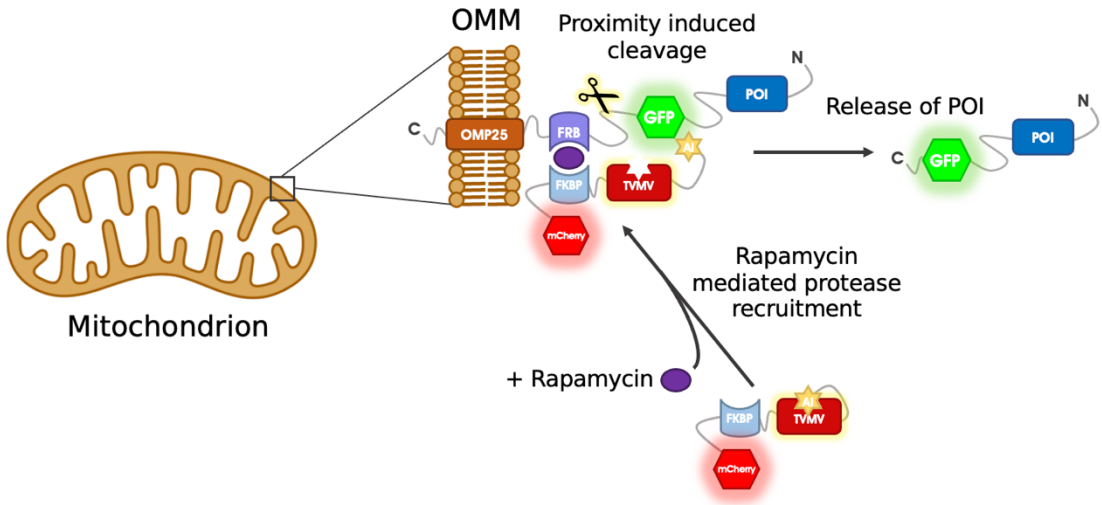



**Figure 4.2: Schematic representation of the RAPID-release technique.**
The addition of Rapamycin induces the dimerization of the FRB and FKBP domains, triggering the release of the autoinhibitory domain of TVMV and subsequent cleavage of the EGFP labelled POI from the mitochondrial tether. Figure adapted from the original publication[189].

### 4.1.2 Distinguishing between endogenous and engineered RAD21

Cohesin is an essential protein with roles in sister chromatid cohesion and in regulating 3D genome architecture; loss of function mutations within the various subunits of the cohesin complex are embryonic lethal[135,190]. Tagging the endogenous copy of RAD21 with the RAPID-release machinery would cause a redistribution of RAD21 to the mitochondria and depletion from the nucleus. Given that cohesin is predominantly found in the nucleus where it performs its biological function, this could be as detrimental to the cells as a complete knock-out. To circumvent this, cells could be cultured under Rapamycin so that the tether is continually released. However, in doing so this may initiate some unknown and undesired off-target effect of prolonged Rapamycin exposure that could negatively impact regular cell function. I therefore decided that the most sensible approach to avoid these issues was to genetically engineer an additional copy of the RAD21 protein tagged with the RAPID-release machinery into the genome of the cell line.

Whilst for microscopy only the fluorescently tagged copy of RAD21 will be visualised, for the ChIP-seq experiment it would be impossible to distinguish between the endogenous and the engineered protein if an anti-RAD21 antibody was used for immunoprecipitation. Therefore, I decided to use the same tagging strategy as described in Chapter 2 and tag the RAPID-release RAD21 with a twin StrepTag (TST) in order to distinguish this copy from the endogenous RAD21 in the downstream ChIP-seq experiments. The influx of 'new' cohesin will be the $RAD21^{TST}$ version, which will be compete with the endogenous form for loading onto the chromatin.

# 4.2 Results

### 4.2.1 RAPID-release in HeLa

#### Design of RAPID-release system for human cell lines

The molecular machinery for RAPID-release can be divided into two independently expressed vectors: OMP25-FRB-EGFP-POI (mitochondrially tethered) and mCherry-FKBP-TVMV (cytoplasmic). This gives the option to use one plasmid per half (2x single vector system), or to combine them into a single plasmid (1x dual vector system). To control stoichiometry, I chose to use the dual vector system for this project. I was kindly gifted the original RAPID-release plasmids by the Bowman lab (plasmid maps provided in Supplementary Materials). Philip Hublitz (WIMM, Genome Engineering Services) used InFusion technology to swap the original Histone3.1 cDNA for human RAD21 cDNA tagged with a TST (Figure 4.3). In the original system the EGFP was situated on the C-terminal end of the POI, however it was decided to move the EGFP to the N-terminal side of RAD21 to allow for C-terminal tagging of RAD21 with the TST as we already knew that WT RAD21$^{TST}$ could be homozygously C-terminally tagged without loss of function. In parallel to this an 'empty' RAPID-release vector (which will now be referred to as 'EV') was generated which contained all the RAPID-release components excluding the RAD21$^{TST}$. I decided to retain the fluorescent markers in the design, as although they are superfluous for ChIP-seq, they are invaluable to the clone screening process (both by microscopy and by FACS) and provide an easy readout when optimising the induction process by live cell imaging (Figure 4.4).

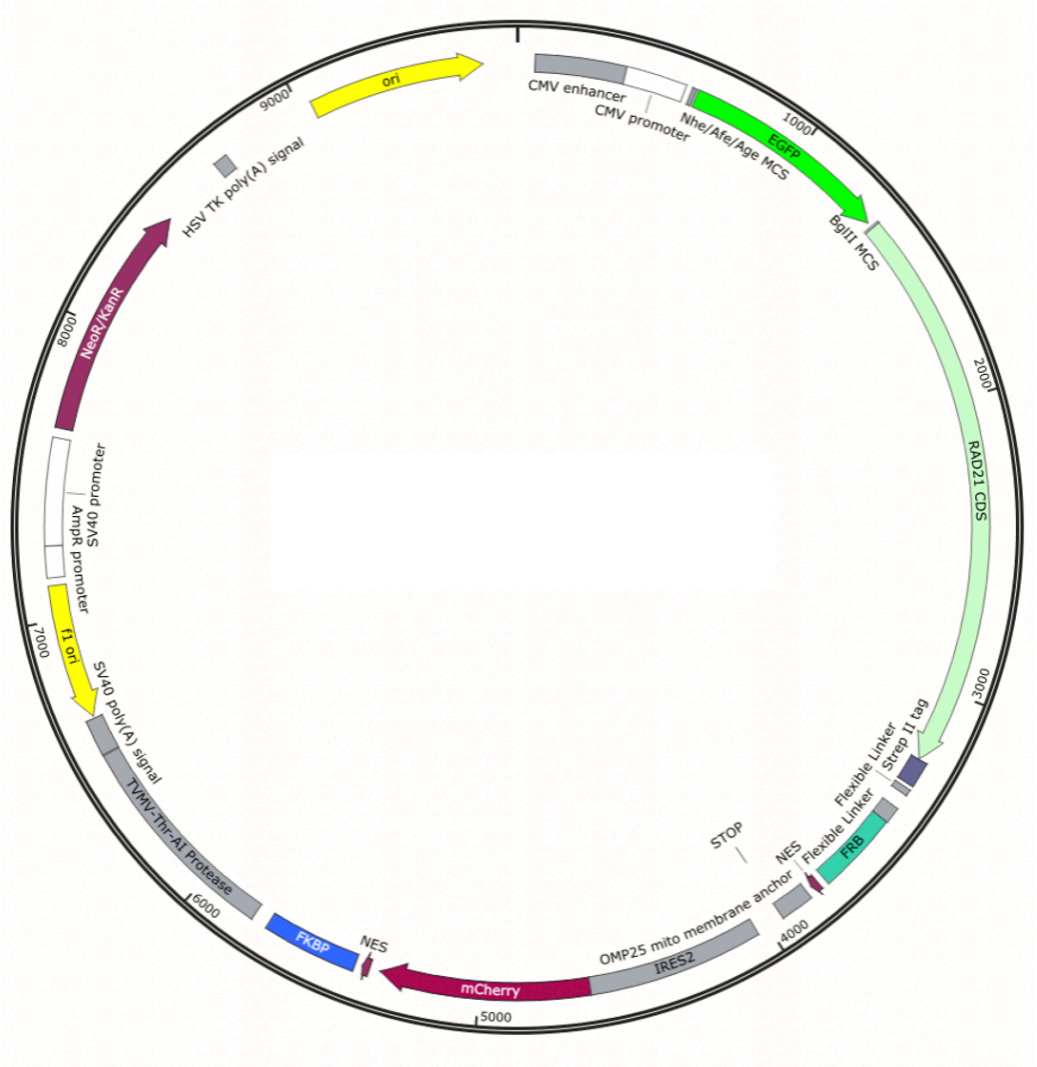



**Figure 4.3: Plasmid map of the RAD21$^{TST}$ RAPID-release dual expression vector.**
Total vector size is 9743bp. Vector contains EGFP- RAD21$^{TST}$-FRB-OMP25 and mCherry-FKBP-TVMV driven off of a strong CMV promoter. Neomycin/Kanamycin (NeoR/KanR) resistance cassette for G418/kanamycin selection. Additional nuclear export signals (NES) are included in both halves of the cassette to ensure cytoplasmic localisation prior to induction. Following induction, the NES will be cleaved, and nuclear import will be driven by the inherent RAD21 nuclear localisation signal (NLS).

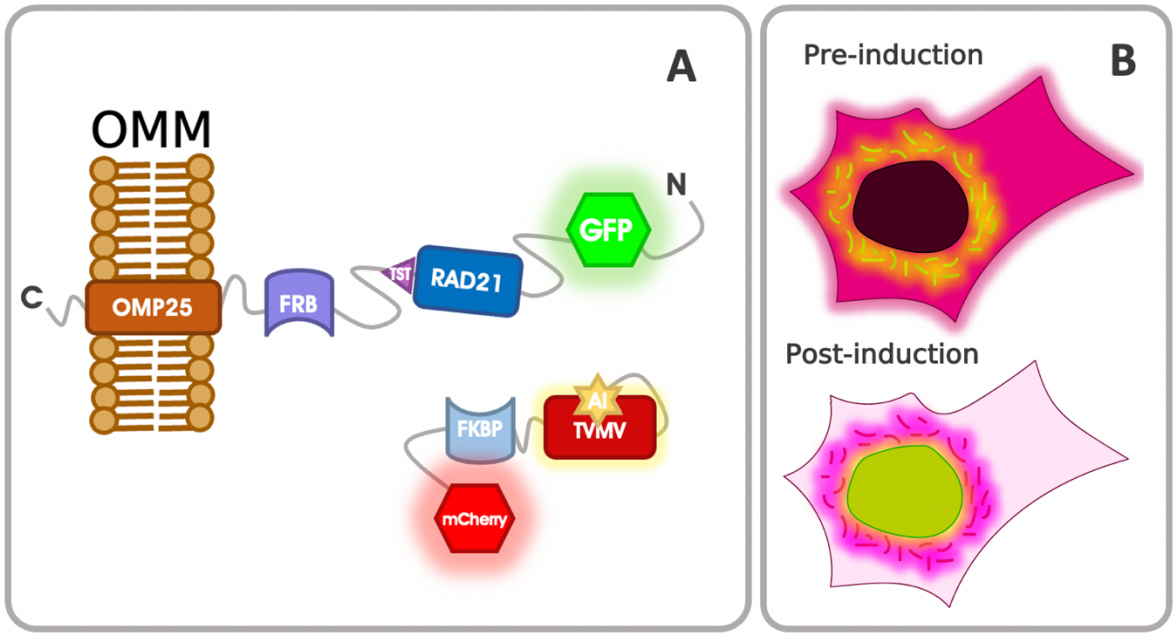



**Figure 4.4: Summary of the RAD21TST RAPID-release design, and the fluorescence signals pre and post induction.**

**a** | The system is made up two halves: first is the EGFP- RAD21$^{TST}$-FRB-OMP25 which tethered to the outer mitochondrial membrane (OMM) prior to induction, second is the mCherry-FKBP-TVMV-AI which freely diffuses in the cytoplasm. **b** | The expected fluorescence pattern before and after induction with Rapamycin. Prior to induction the mitochondria fluoresce green due to the tethered GFP-tagged fusion protein, and the cytoplasm in red (shown in magenta) due to the freely diffusing mCherry tagged fusion protein. Following the addition of Rapamycin, the GFP-RAD21$^{TST}$ is cleaved and moves into the nucleus causing it to fluoresce green, and the mCherry-FKBP becomes mitochondrially tethered causing them to change from green to red.

## Optimisation of live cell imaging

Transfection optimisations were carried out using HeLa cells in which different lipofection reagents were tested and plasmid DNA concentration was titrated in order to determine the optimal transfection conditions (see Materials and Methods for full details). Live-cell imaging was used to qualitatively assess cell viability, cell morphology and to confirm that cells were expressing mCherry and EGFP in the correct cellular compartments. JetPRIME was selected as the transfection reagent that resulted in the highest transfection efficiency and least cellular toxicity. Three vectors were transfected in parallel to create three models: (1) the original single vector OMP-FRB-EGFP$^{TVMVx2}$-H3.1 as a positive control, (2) the new modified RAD21$^{TST}$ RAPID-release dual vector, (3) the empty dual vector (EV) as an additional control. By live cell imaging, the cells transfected with the H3.1, RAD21 RAPID-release vectors, and the EV appeared identical prior to induction as was expected (see Figure 4.5 and Figure 4.6) with respect to EGFP localisation, and mCherry for vectors (2) and (3).

Once the transfection protocol had been optimised, a preliminary induction experiment was performed to test the RAPID-release mechanism worked as expected. This was once again performed using all three vectors (H3.1, RAD21$^{TST}$ and EV), this time all in dual format. The H3.1 and RAD21 were expected to appear identical with respect to the observed fluorescence pattern, with the EV as the negative control for nuclear EGFP expression given the lack of NLS-containing protein. Prior to induction, an initial 5 images were acquired, Rapamycin was administered at the concentration used in the original publication (200nM)[189] and subsequent images were acquired every 5 seconds for up to 30 minutes or until no further change in fluorescence localisation was observed. Upon the addition of Rapamycin, the change in fluorescence localisation was rapid and followed the pattern as expected (Figure 4.7). In the cells expressing RAD21$^{TST}$ and H3.1 RAPID-release: the mitochondria switched from green to red as the FRB dimerised with the mCherry-

tagged FKBP domain to initiate the cleavage of RAD21$^{TST}$-EGFP, which then migrated into the nucleus, directed by the NLS, resulting in an increase in green fluorescence in the nuclear compartment. The EV expressing cells do not contain a protein with an NLS fused to EGFP, therefore upon the addition of Rapamycin the EGFP is not directed to the nucleus and thus is left to freely diffuse around the cytoplasm. The difference in these two mechanisms was clearly distinguishable by live cell imaging (see Figure 4.7).

## Quantifying the progress of RAPID-release using image analysis

I analysed the imaging data from these preliminary imaging experiments using Fiji[191], measuring the nuclear and mitochondrial EGFP and mCherry signals throughout the duration of the imaging experiment (Figure 4.8). For all three systems, the nuclear mCherry signal remained low and relatively unvaried throughout the experiment, whilst the nuclear EGFP signal increased until it surpasses the mitochondrial EGFP signal at a time of between ~7-9 minutes post induction. Interestingly, the mitochondrial mCherry signal did not drastically increase after induction as I expected. However, for this analysis I did not segment the images to precisely select mitochondrial signal and therefore the signal measured is most likely diluted by background cytoplasmic signal. Overall, the RAPID-release system proved to be as rapid and efficient as was reported in the original paper, shown by the striking switch in EGFP and mCherry fluorescence in the different cellular compartments (link to movies in Materials and Methods Chapter). Although, in both the H3.1 and RAD21$^{TST}$ RAPID-release expressing cells there was a degree of residual green fluorescence visible in the nucleus indicating that a proportion of expressed fluorescent protein had escaped being tethered to the mitochondria, and was instead leaking into the nucleus.

### Random transient transfections are inconsistent

The aim of this initial set of experiments, in which HeLa cells were transiently transfected with RAPID-release, was to validate that the system was functional and could be induced with Rapamycin. Throughout these experiments it was clear that there was a high degree of variability in protein expression levels across cells in the same experiment (as seen in Figure 4.5), which could be due to a number of reasons. Firstly, the variability in the number of copies of the plasmid DNA transfected into the cell can affect expression levels. Whilst this can be optimised to some degree based on the ratio of DNA to transfection reagent and cell numbers, without using a targeted genome engineering approach such as the CRISPR-cas9 system it is difficult to precisely control copy number. Secondly, if the plasmid becomes stably integrated into the genome, the non-directed approach results in random insertions within the host genome. Given the differences in activity and silencing effects at different genomic loci, the location of plasmid integration can significantly affect expression levels[192] and cause a high level of heterogeneity across transfected cells. Whilst the transient transfections were sufficient to confirm the system was functioning as expected, the percentage of viable cells with visible, sustained and consistent EGFP and mCherry expression across cells was at best 10-20% (results not shown). To obtain enough cells for ChIP-seq (for which at least $5x10^6$ cells per replicate are required), the protocol would require significant optimisations to increase the percentage of RAPID-release expressing cells and reduce the heterogeneity of expression levels within the cell population. To address this issue, I decided that the most efficient approach would be to generate a clonal line with RAPID-release randomly, but stably integrated into the genome.

### Plasmid linearization for increased transfection efficiency

For all transfections up to this point I used the plasmids in their circular form. Whilst it is well reported that uptake of plasmid by the cell is higher with circular plasmids[193], linear DNA is known to be more recombinogenic and therefore stable integration into the host chromosome is more likely[194,195]. Linearising the plasmid enables the break point within this circular plasmid to be controlled and ensures that the insertion sequence is more likely kept intact, maximizing the chance of the cells integrating the full target sequence. In doing so the population of false positives is also minimised: cells that express the selection marker are more likely to have an intact copy of the full plasmid sequence, rather than having integrated a short section that includes the selection marker. I therefore decided to linearise the RAD21$^{TST}$ RAPID-release plasmid using restriction enzyme digestion (AseI) prior to transfection in order to generate the stable lines.

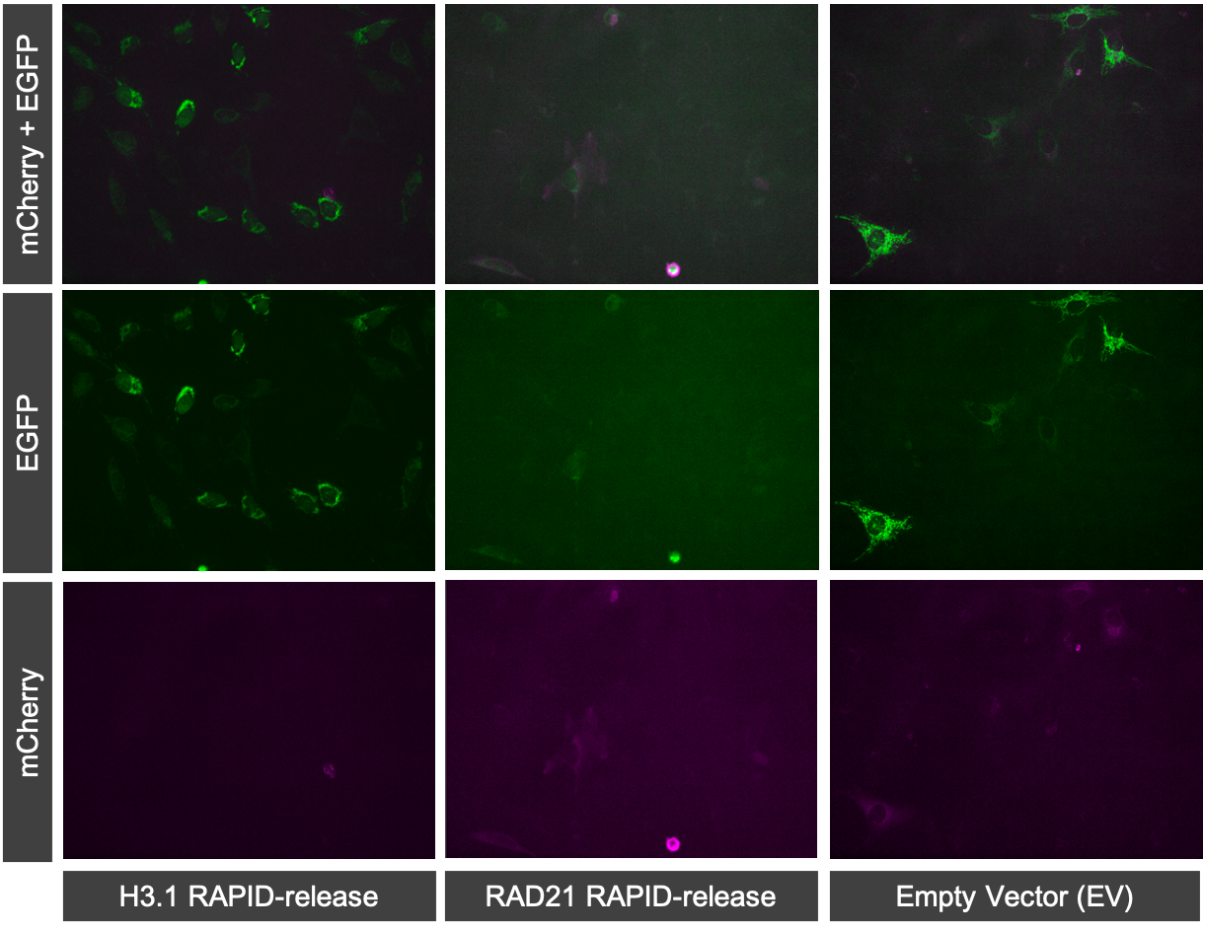



**Figure 4.5: Comparison of cells transfected with circular H3.1, RAD21$^{TST}$ and empty vector (EV) RAPID-release plasmids by live cell imaging (20x magnification) in HeLa cells.**
Images were acquired 24 hours post-transfection. EGFP is represented in green and mCherry in magenta. Top row shows the two-colour channels merged, below are the split channel views. Figure shows the heterogeneity in expression levels across the transiently transfected cells.

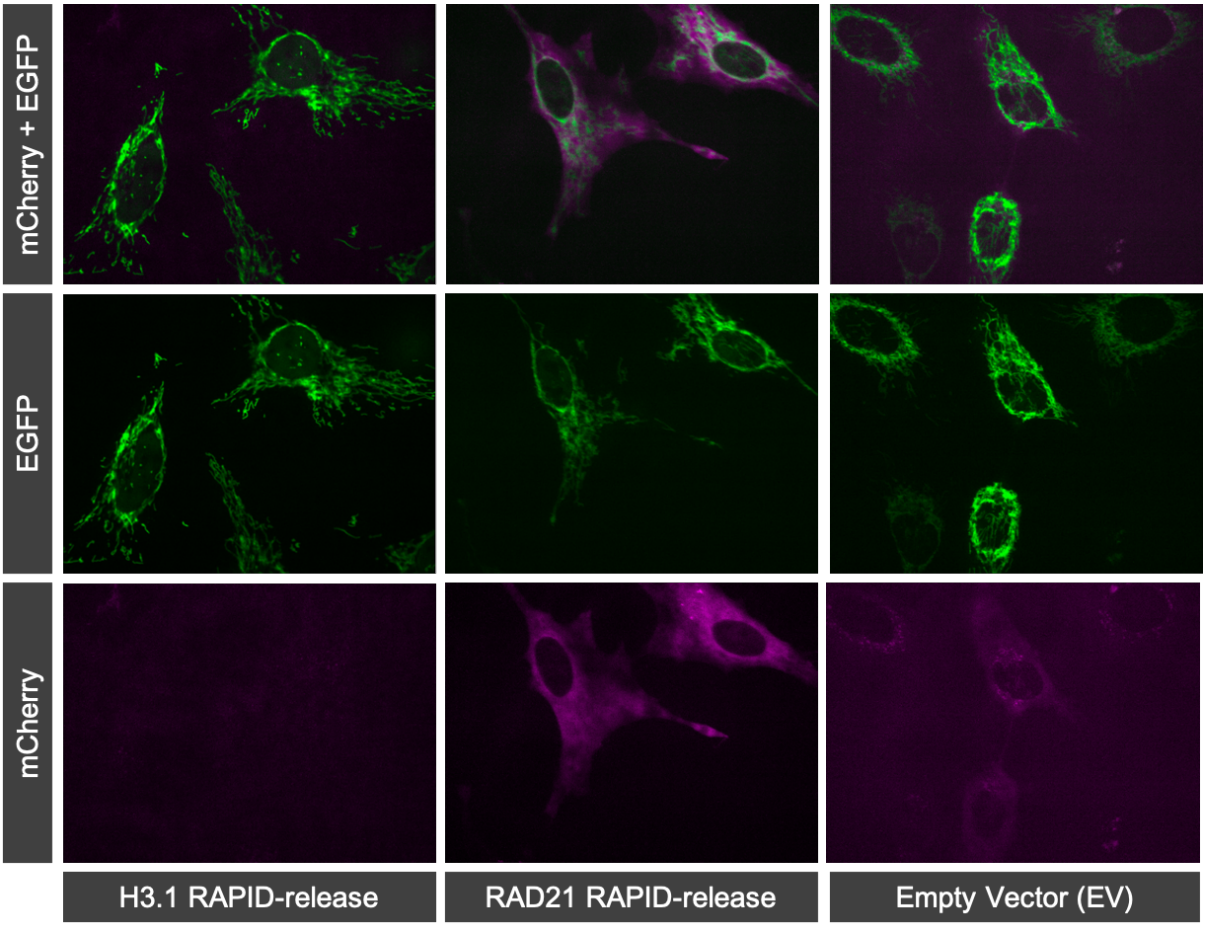



**Figure 4.6: Comparison of cells transfected with circular H3.1, RAD21$^{TST}$ and empty vector (EV) RAPID-release plasmids by live cell imaging (63x magnification) in HeLa cells.**
Images were acquired 24 hours post-transfection. EGFP is represented in green and mCherry in magenta. Top row shows the two-colour channels merged, below are the split channel views. Figure shows the heterogeneity in expression levels across the transiently transfected cells.

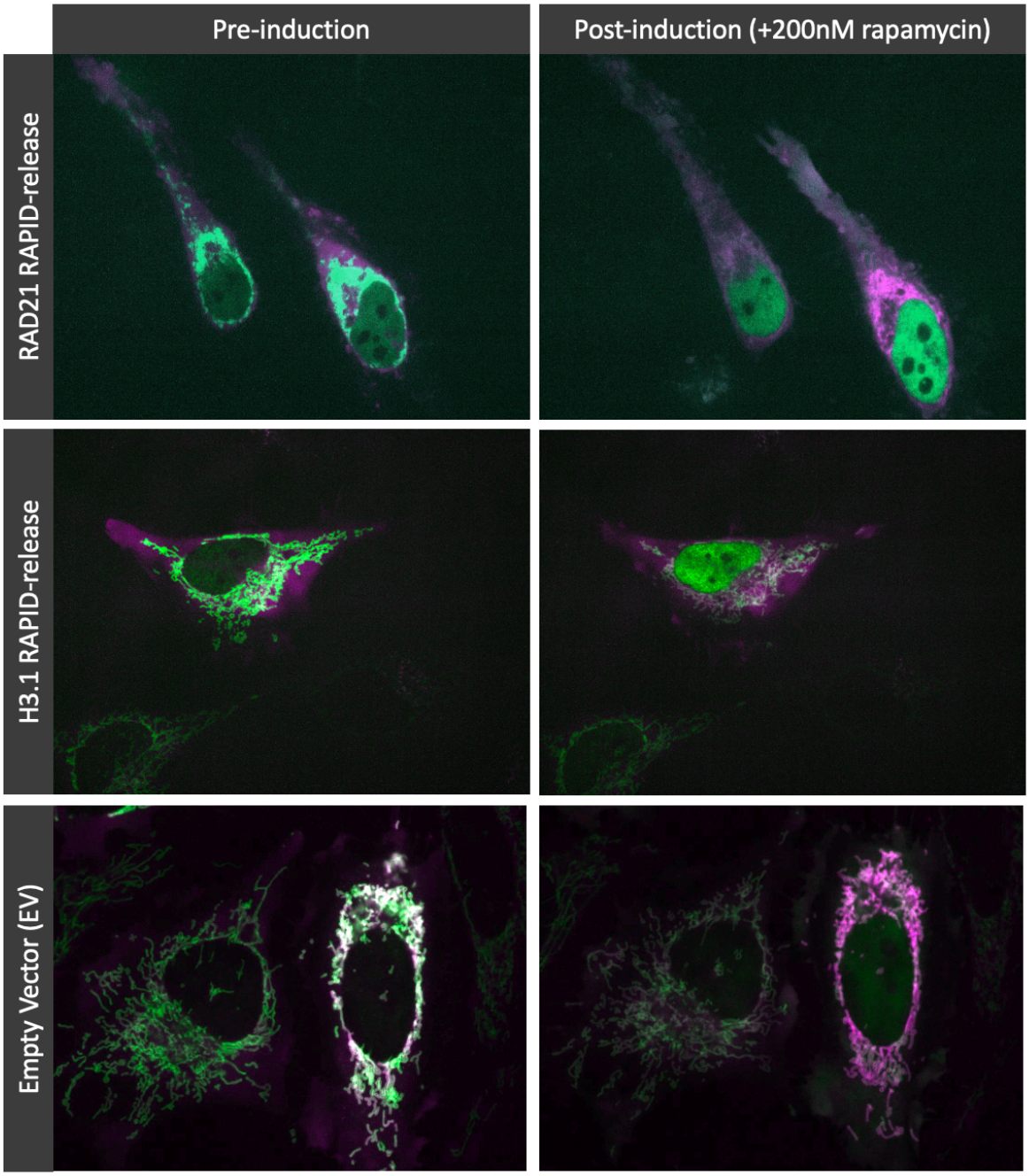



**Figure 4.7: Live cell imaging of HeLa cells transiently expressing RAPID-release (H3.1, RAD21$^{TST}$ and empty vector (EV)) before and after the addition of Rapamycin (63x magnification).**
For full movies see link in Materials and Methods section. Images were acquired 24 hours post-transfection. EGFP is represented in green and mCherry in magenta. Images are snapshots from the first and last frame of the induction time-course: the first image was acquired prior to the addition of Rapamycin; the second image is the last snapshot from the movie in which the cells have been induced with 200 nM Rapamycin and left to equilibrate for approximately 10 minutes.

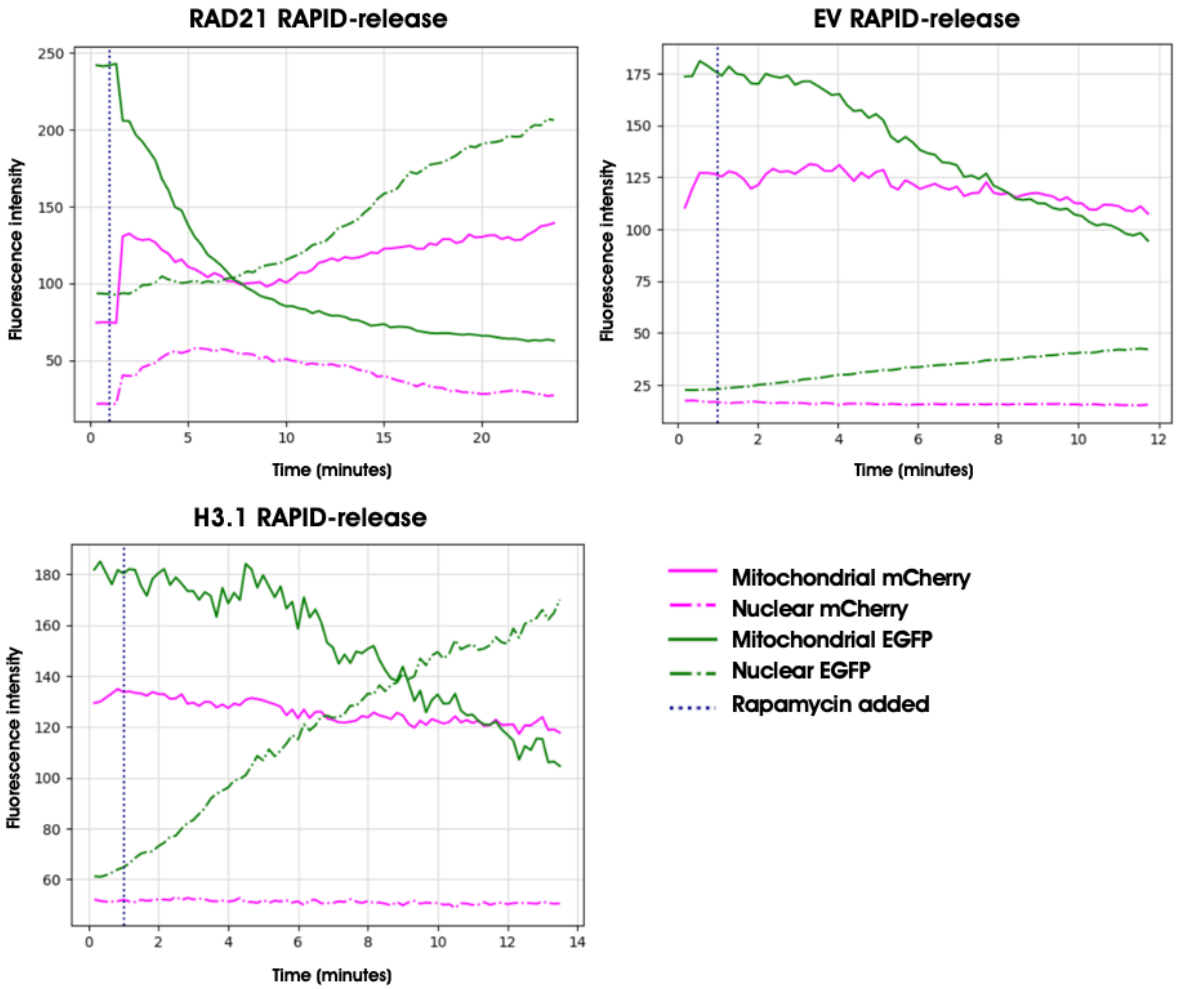



**Figure 4.8: Image analysis showing the change in fluorescence profiles of the nucleus and mitochondria throughout the RAPID-release experiment.**

Graphs show the change in mCherry and EGFP fluorescence in the nucleus and the mitochondria of the three different models (RAD21$^{TST}$, EV and H3.1). Vertical blue dotted line indicates the addition of Rapamycin, magenta lines represent mCherry and green represent EGFP. No segmentation was performed to separate the mitochondria from the cytoplasm so 'mitochondrial' fluorescence readings will also take into account any cytoplasmic fluorescence. Analysis was performed using Fiji[191], n=1.

### 4.2.2 Optimising ChIP-seq conditions for time-course experiment

Using live cell imaging I showed that the RAPID-release process was extremely quick; taking ~10-20 minutes to achieve equilibrium starting from the point of Rapamycin addition, with the initial changes being most rapid. For microscopy this order of time does not affect the ability to acquire data, ChIP-seq on the other hand has a number of rate-limiting steps that could prevent the dynamic changes being captured. For example, the regular ChIP-seq protocol used in the lab involves double fixing the cells: first with Disuccinimidyl Glutarate (DSG) for long-range cross-linking for 30 minutes, then with formaldehyde for short-range cross-linking for another 30 minutes. Including the rinsing and centrifugation steps, the full fixation process can take up to 1.5 hours, which is clearly not suitable to capture changes in protein-chromatin interaction that change over the course of seconds and minutes. Traditionally, ChIP-seq is used to measure the population average of the binding patterns of chromatin-associated proteins at equilibrium. It therefore did not matter how long the fixation took, and the timings were chosen to find a reasonable balance between maximal fixation and ease of experimental procedure. For the purposes of the experiments in this chapter however, this method must be modified so that proteins can be fixed to chromatin as instantaneously as possible: once the cells are harvested at a selected time-point, any additional time taken to fix the cells will be blurring the time points. In order to achieve the required temporal resolution to resolve cohesin binding dynamics, it was necessary to optimise faster fixation methods.

#### Increasing the temporal resolution by optimising the fixation protocol

As already mentioned, the ChIP protocol used routinely in the lab involved two sequential 30-minute fixation steps. I predicted that this was in excess of the amount of time required to fix RAD21 to the chromatin and could therefore be significantly reduced. I therefore tested different timings and combinations of fixatives to determine which was

the most efficient (Figure 4.9). A consecutive double fixation of 30 minutes DSG followed by 30 minutes of 1% formaldehyde (Figure 4.9, "30mins DSG, 1% FA") gave the greatest fold enrichment, which was significantly greater than a single fixation of 1% formaldehyde for 1hr, or by using the same fixatives for only 10 minutes each (which gave the lowest fold enrichment). Using a 2% formaldehyde fixative gives similar fold enrichment to 1% for 30 minutes and 1 hr, but with the advantage of having a 3x and 6x shorter fixation time respectively. Using 1% formaldehyde for 10 minutes gives a significantly lower fold enrichment than using 2% formaldehyde for the same amount of time, suggesting that the higher concentration is required for effective fixation over a shorter time. Interestingly, by combining DSG and formaldehyde in a single fixation step (Figure 4.9, purple bar) the fold enrichment is significantly worse and inconsistent (shown by large error bars). This result aligns with previous studies showing that DSG and formaldehyde biochemically interfere with one another and result in reduced fixation[196]. From these results, I selected the 2% formaldehyde for 10 minutes as the fixation method for all further ChIP experiments as this gave the highest fold enrichment for a short fixation time which will make time-course experiments with multiple timepoints over a period of minutes experimentally feasible.

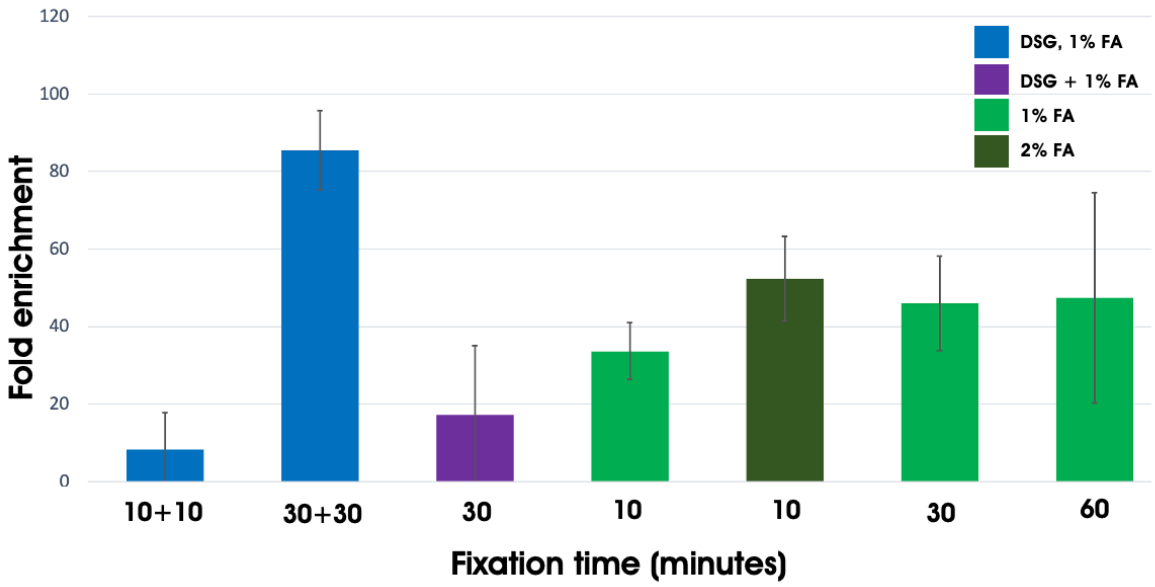



**Figure 4.9: Assessing the efficiency of different fixation conditions using ChIP-qPCR.**
Results show fold enrichment of RAD21$^{TST}$ at a site in the Nanog locus (chr6:122,646,563-122,646,667) compared to an intergenic background control (chr6:122652386-122652480). Error bars shows show +/- standard deviation from the mean across the three replicates. All experiments were performed in the mESC RAD21$^{TST}$ cell line using an anti-Strep-tag antibody for immunoprecipitation. Disuccinimidyl Glutarate (DSG), formaldehyde (FA). Bar colours indicate fixative: blue indicates sequential fixation with DSG followed by FA, purple indicates combined DSG and FA treatment, green bars indicate FA fixation. See Materials and Methods for detailed description of the experiment.

## Methods to quench the fixation

In the current ChIP protocol used in the lab, no quencher is added to stop the fixation reaction. I thought it would be useful to investigate the efficiency of quenchers to halt the fixation process to further increase the temporal resolution of the ChIP-seq method. The most widely used quencher for this purpose is glycine, however there are a few publications that suggest that this does not efficiently quench the reaction and subsequently result in over fixation. Tris has been tested as an alternative to glycine and showed convincing results for quenching the fixation reaction[197,198]. I performed a comparison of the fixation quenchers, glycine and tris: mES E14 cells were fixed with 2% formaldehyde for 1 minute, then quencher was added (either 125 mM glycine or 1.5 M tris).

At timepoints of 1 minute, 5 minutes and 10 minutes after the addition of the quencher, the cells were centrifuged, and pellets rinsed with PBS to remove the fixative-quencher solution. I expected that an efficient quencher would rapidly quench the reaction and therefore the fold enrichment assayed by qPCR at 1, 5 and 10 minutes after its addition will be very similar. An inefficient quencher would either fail to quench, or take more time to completely quench the reaction and therefore fixation will continue after the addition of the quencher and will be seen by an increasing fold enrichment over time. ChIP-qPCR was performed following this test and the results are shown in Figure 4.10. Tris is a significantly more efficient quencher than glycine, with the fold enrichment remaining similar after its addition. Glycine on the other hand does not effectively quench the fixation reaction and therefore the fold enrichment increases at each time point.

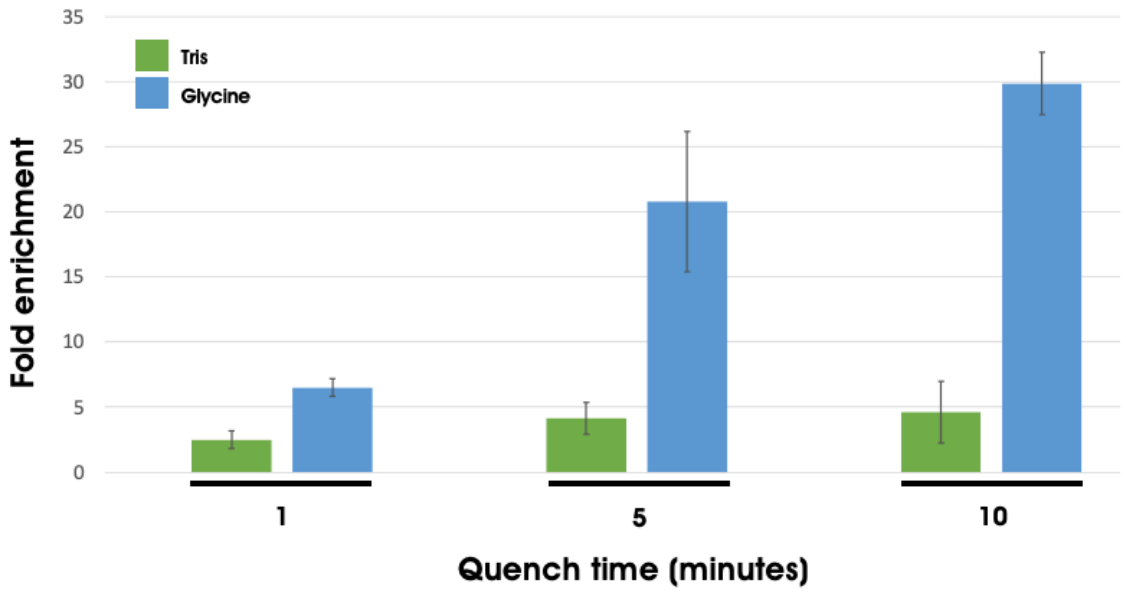



**Figure 4.10: Comparison of the quenching efficiency of tris and glycine using ChIP-qPCR.** Results show fold enrichment of RAD2 at a site in the Nanog locus (chr6:122,646,563-122,646,667) compared to an intergenic background control (chr6:122652386-122652480). Three replicates per condition were performed in WT mESC using an anti-RAD21 antibody (ab992, abcam). Tris (green), Glycine (blue). Error bars show +/- standard deviation from the mean across the three ChIP replicates and three qPCR replicates. See Materials and Methods for detailed description of the experiment.

### 4.2.3 RAPID-release in HEK293

#### Generating a stable line using HEK293 cells

Given the difficulty in obtaining consistent and uniform transfections in HeLa cells, I decided to use a HEK293 cell line to generate the stable lines given the well reported propensity for high efficiency transfection in these cells[199]. Using an in-house HEK293 cell line kindly gifted by Ryan Beveridge (WIMM, Genome Engineering Services) and a pre-optimised protocol using PEI-pro as a transfection reagent, a much higher transfection efficiency and homogeneous expression of RAPID-release machinery (visualised by EGFP and mCherry expression) was achieved compared to the previous HeLa cell transfections (see Figure 4.11). To generate clonal lines, media containing 600 µg/mL of G148 was added 48 hours after transfection and cells were cultured in these conditions for 10 days to select for clones that have integrated the intact plasmid containing the antibiotic resistance gene. See Materials and Methods for kill curve and full stable line protocol. The previous experiments in HeLa cells exhibited a degree of EGFP leaking into the nucleus prior to induction, I therefore decided to include a sorting step into the protocol for stable line generation following the antibiotic selection, using EGFP and mCherry fluorescence levels as a proxy for protein expression. I expected that the leakage issue was due to a saturation of available mitochondrial binding sites, thereby sorting cells based on expression levels I could find a threshold at which the mitochondria are not saturated and consequently will not show leakage of fluorescently labelled $RAD21^{TST}$ into the nucleus. Using fluorescence activated cell sorts (FACS), single cells were isolated into 96 well plates based on low, medium, and high levels of EGFP and mCherry (see Figure 4.12). Clonal populations of low, medium, and high expressing cells were cultured and expanded for further investigation.

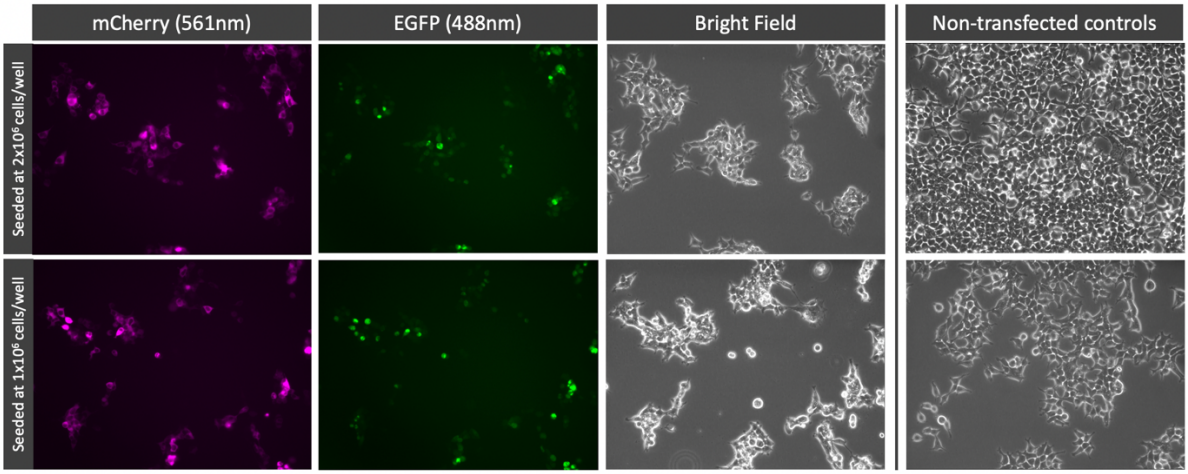



**Figure 4.11: Live cell imaging of HEK293 cells transiently expressing RAD21 RAPID-release 48hrs post transfection (10x magnification).**

Cells were seeded at two different densities: $1x10^6$ and $2x10^6$ cells per well of a 6 well plate. There is a significant difference in the number of cells that have survived post transfection, as can be seen by the difference in cell density between the brightfield images of transfected versus non-transfected control. However, of the cells that survived, the transfection efficiency is high.

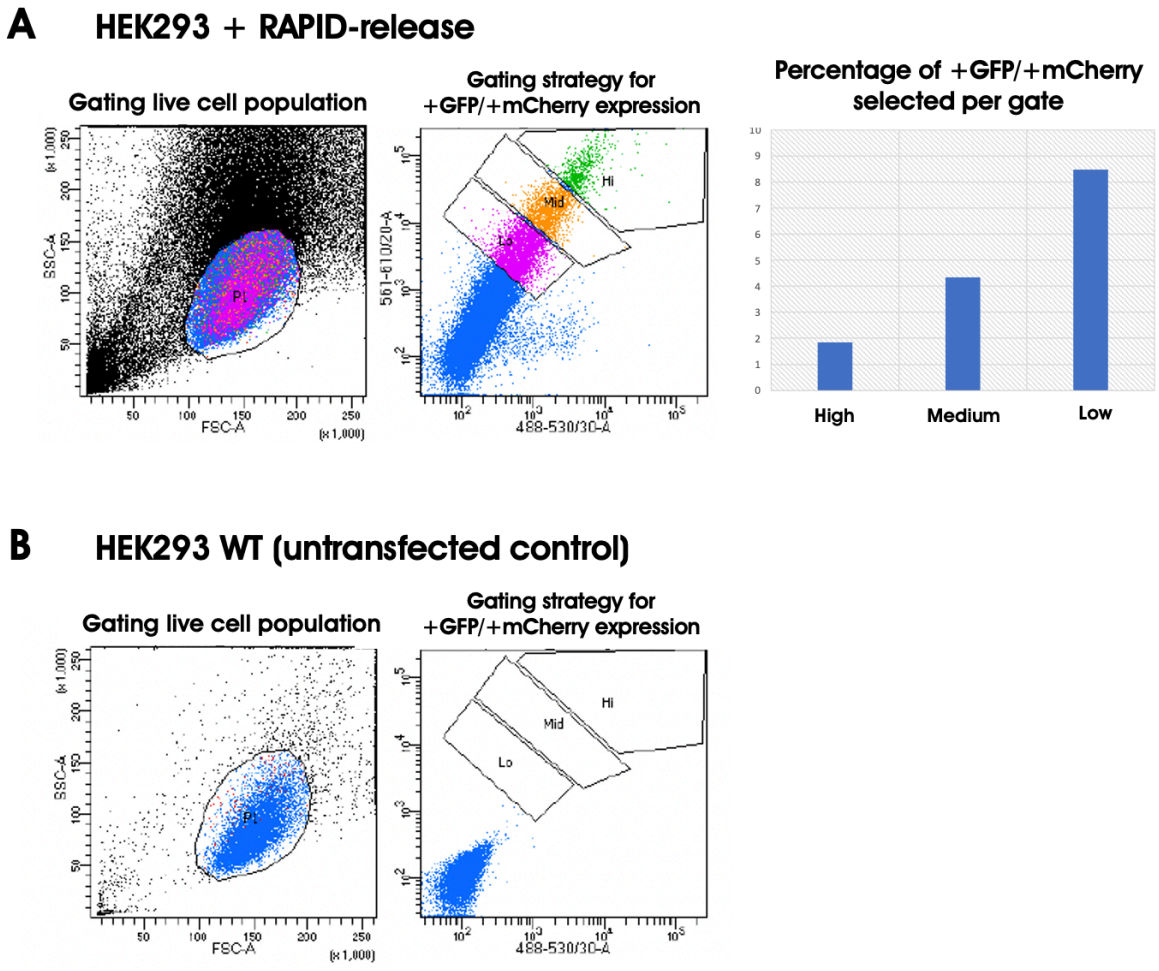



**Figure 4.12: FACS gating strategy for sorting populations of HEK293 cells based on expression levels of mCherry and EGFP.**

**a** | FACS plots showing fluorescence signals in RAD21$^{TST}$ RAPID-release HEK293. Left-hand plot shows the gating for selection of live cells. Right-hand plot shows the gating strategy within the live cell population to select cells expression low, medium, and high levels of EGFP (488-530/30-A) and mCherry (561-610/20-A). Bar chart indicates the percentage of cells selected per gate. **b** | FACS plots showing fluorescence signals in WT HEK293 cells as a negative control. The same gating was applied for selection of live cells as in (a). The right-hand plot shows that by using this same gating, there were no cells exhibiting EGFP or mCherry expression.

### Live cell imaging of HEK293 stable lines

After expanding the clonal lines of high, medium, and low expression, surviving clones were taken for imaging. Using clonal lines vastly improved the issue with expression heterogeneity seen in the transiently transfected HeLa cells. The stably transfected clonal cell line exhibited homogenous levels of fluorescent proteins, making it easier to find and focus on an individual cell or group of cells with adequate expression levels of the fluorescent marker for imaging. Despite having FACS sorted the cells based on the intensity of EGFP and mCherry, by live-cell imaging there was no visibly observable difference between the fluorescence levels across the different categories of clones (see Figure 4.13). Additionally, from this live-cell imaging there still seemed to be high levels of EGFP visible in the nucleus prior to induction. This suggests that the approach of using fluorescence levels as a proxy for protein expression would require further optimisation, particularly with regards to the sorting thresholds given that even the 'low' level of expression appears to be saturated with EGFP.

The objective of this Chapter is to adapt the RAPID-release method to track cohesin loading and dynamics at the chromatin level. Imaging methods are a useful way to validate the system but cannot provide resolution of protein localisation beyond the level of organelles; it is unknown whether the EGFP fluorescence exhibited in the nucleus is reflective of what is seen at the chromatin level. I therefore decided to further investigate protein levels at the organelle and chromatin level in the FACS sorted HEK293 clonal lines, before devising a new strategy to generate new RAPID-release cell lines with decreased expression levels.

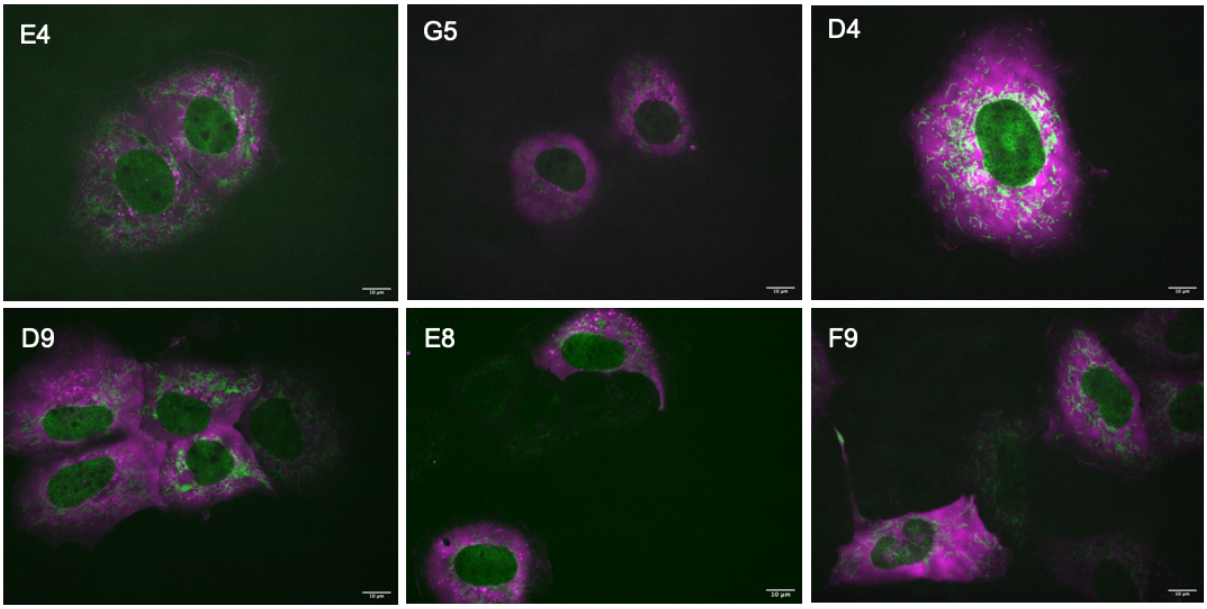



**Figure 4.13: HEK293 RAD21$^{TST}$ RAPID-release stable lines without induction (63x magnification).** EGFP is represented in green, mCherry in magenta. Labels indicate clone name, with associated numbering corresponding to expression level: 1-4 high, 5-8 medium, 9-12 low. EGFP is represented in green, mCherry in magenta. 10µm scale bar included for reference.

## Subcellular fractionation and western blot

The derived HEK293 clonal lines were expanded, and protein was extracted by whole cell lysis, and using a subcellular fractionation to isolate protein from the following cellular compartments: cytoplasmic, membrane (including mitochondria and ER), nuclear soluble, chromatin bound and cytoskeletal (see Materials and Methods Chapter for details). The distribution of protein across the fractions was then assessed using western blot. As in Chapter 3, I chose to use fluorescent western blot, as this gave the ability to multiplex the detection of RAD21 and RAD21$^{TST}$ in order to distinguish between the two forms of RAD21 in a single blot without the need to strip and re-probe between antibodies.

Although this was the same process as used in Chapter 3, I wanted to account for any cell-type variability between mESC and HEK293 cells. Therefore, I tested the subcellular fractionation and western blot protocol in WT HEK293 cells to ensure that the methods

and fraction controls were appropriate for this cell type. Figure 4.14 shows the results of the subcellular fractionation in WT HEK293 cells, demonstrating that the same fraction control antibodies were suitable in this cell type. Additionally, given the reliance of this technique on mitochondrial tethering, I tested two mitochondria-specific antibodies (anti-TOMM20 and anti-VDAC) for use as a membrane fraction control, but these showed no signal in any fractions, and I decided to proceed without this control (Supplementary Materials, Figure S12). I multiplexed the detection of RAD21 with StrepTag to ensure that there was no off-target detection of the TST in WT cells. Figure 4.14 shows that RAD21 is present in both nuclear fractions as expected in the WT cells.

After establishing the expected protein distribution across fractions in WT cells, the clone F9 (low expression) was used to test whether this protein distribution differed in cells stably expressing RAD21$^{TST}$ RAPID-release. For the system to be working correctly, there should be only WT RAD21 detected in the nucleus and no protein detected by anti-StrepTag in either of the nuclear fractions. In the membrane fraction I would expect protein to be detected by the multiplexed RAD21/StrepTag antibodies due to the presence of RAD21$^{TST}$ tethered to the OMM. However, given the leakage seen in the RAD21$^{TST}$ RAPID-release cells by microscopy, I expected that RAD21$^{TST}$ may also be detected in the nucleus in these cells.

Figure 4.14B shows the result of the subcellular fractionation in the F9 RAPID-release clone. The Alpha-tubulin and Histone-2B controls appear almost identical to the WT experiment, suggesting that the fractionation was clean. By microscopy, RAD21$^{TST}$-EGFP signal was strong and clearly localised to the mitochondria, it was therefore surprising that there was signal from the RAD21/StrepTag in the membrane fraction which should contain these organelles. This complete lack of signal in the membrane fraction is suggestive of a false negative result due to an unsuccessful isolation of membrane bound

proteins, however this cannot be confirmed due to a lack of a working positive control for this fraction. Aside from the membrane bound fraction, the RAD21 signal (green) was as expected due to the leakiness of $RAD21^{TST}$: the strongest signal came from the nuclear fractions (which contained both native and leaked $RAD21^{TST}$), with weaker bands shown in the cytoplasmic and whole-cell lysate fractions which should originate from the $RAD21^{TST}$. Interestingly, StrepTag antibody (red) only detected proteins in the nuclear soluble fraction. RAD21 is tagged with the TST (i.e., $RAD21^{TST}$) in the RAPID-release cells, thus any leakage of $RAD21^{TST}$ detected by anti-RAD21 should be mirrored by the detection of StrepTag. As a result, I would have expected there to be StrepTag bands seen in both the cytoplasmic and whole cell lysate fractions which was not the case. I also expected the ratio of soluble and chromatin bound $RAD21^{TST}$ that had leaked into the nucleus to be the same as WT, instead none was detected in the chromatin fraction. If this was the case, it could be that the tagging of RAD21 with TST and GFP was somehow inhibiting the ability of RAD21 to associate with the chromatin. Detection of $RAD21^{TST}$ in the RAPID-release cells is a crucial requirement in order for this model to be used to investigate the dynamics of cohesin loading. Before any further optimisations of the fractionation or western blot protocols, I decided to carry out a ChIP-seq experiment to determine whether $RAD21^{TST}$ could be detected on the chromatin.

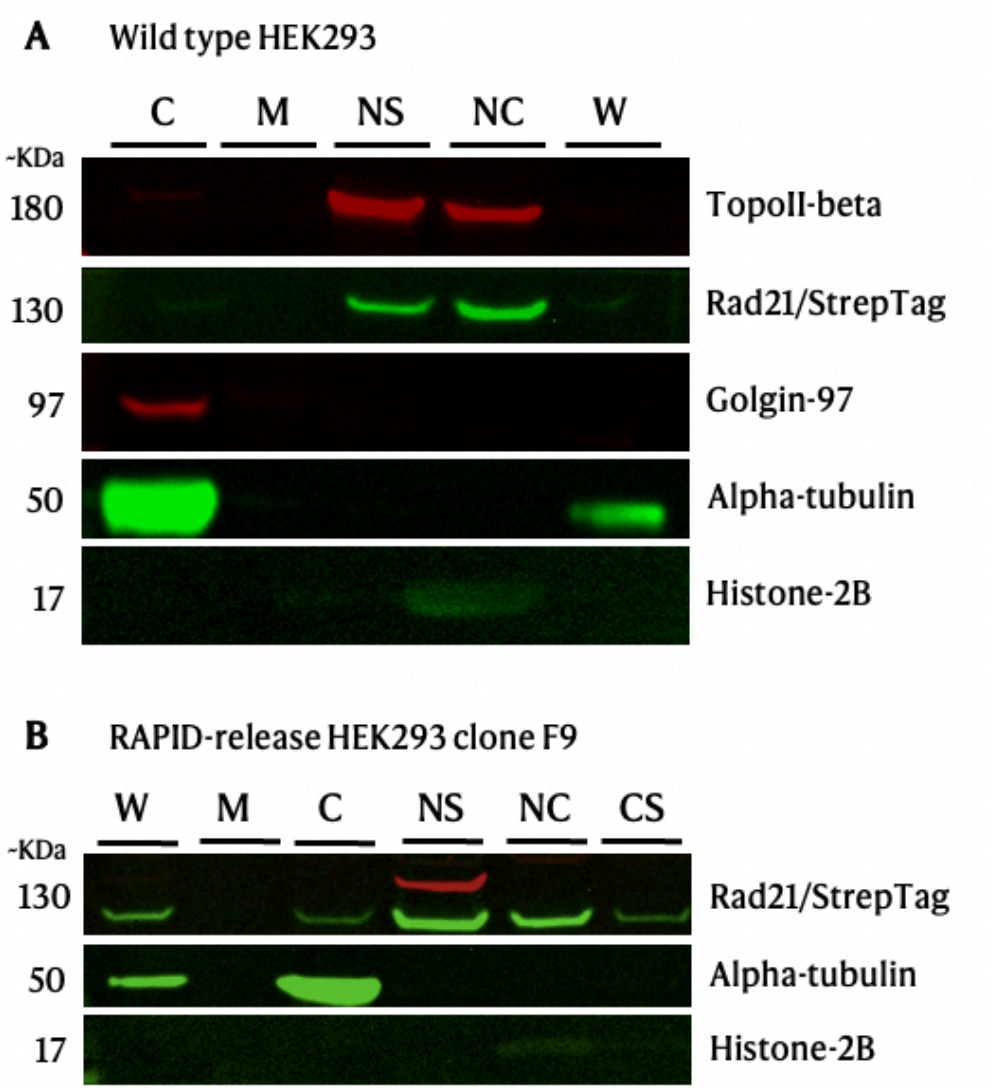



**Figure 4.14: Subcellular fractionation of HEK293 cells.**

**a** |Fluorescent western blot showing subcellular fractionation in WT HEK293 cells. Cytoplasm (C), nuclear soluble (NS), nuclear chromatin (NC) and whole cell lysate (W) fractions are shown. RAD21 (green, 800CW) and StrepTag (red, 680RD) were multiplexed for dual detection. **b** | Fluorescent western blot showing subcellular fractionation in RAPID-release clone F9 HEK293 cells. Whole cell lysate (W), membrane (M), cytoplasm (C), nuclear soluble (NS), nuclear chromatin (NC), cytoskeletal (CS). RAD21 (green, 800CW) and StrepTag (red, 680RD) were multiplexed for dual detection.

### Can RAD21$^{TST}$ be detected by ChIP-seq?

Western blot analysis of the RAPID-release clone F9 showed no detectable RAD21$^{TST}$ in the nuclear chromatin fraction. I therefore wanted to investigate if there truly was no RAD21$^{TST}$ bound to the chromatin, or if this was a false negative result of the western blot experiment. The main aim for this initial ChIP-seq experiment was to determine whether chromatin bound RAD21$^{TST}$ could be detected, I decided to test this before and after the addition of Rapamycin. I harvested cells at two timepoints: (1) prior to induction, at which point there should theoretically be no RAD21$^{TST}$ bound to the chromatin as it should be tethered to the mitochondria; (2) 30 minutes after the addition of Rapamycin, by which time the system has had time to fully equilibrate following the induction and the RAD21$^{TST}$ ChIP-seq profile should look the same as a WT RAD21 profile. Based on the live cell imaging prediction in which the nuclear EGFP signal surpassed the mitochondrial signal around 10 minutes post induction, therefore I chose a timepoint of 30 minutes post induction to leave sufficient time for cohesin to equilibrate in the nucleus and load onto the chromatin. As a negative control, I used a WT HEK293 line: by using an anti-StrepTag antibody for the immunoprecipitation this should give blank, input-like profiles. To allow quantitative comparisons across the timepoints, the ChIP-seq experiment was calibrated by including a 4% spike-in of RAD21$^{TST}$ mES E14 cells to the HEK293 cells prior to fixation (details in Materials and Methods chapter).

Figure 4.15 shows the data from the preliminary ChIP-seq experiment. From this data it is clear to see that the system was saturated with RAD21$^{TST}$ causing it to leak into the nuclear compartment where it binds the chromatin. Given the imaging work data already generated in these clones, this result was not surprising and confirms the need to find a way of reducing the RAD21$^{TST}$ expression to more physiological levels. However, these data does confirms two factors that will be important in future experiments: (1) it is possible to immuno-precipitate the RAPID-release RAD21$^{TST}$ using an antibody against the StrepTag,

and this is specific to the RAD21$^{TST}$ protein, (2) the peaks of RAD21$^{TST}$ binding in the RAPID-release HEK293 cell line align with RAD21 binding peaks in WT cells, confirming that tagging RAD21 with the RAPID-release machinery does not visibly disrupt or alter the normal binding pattern of the protein. The next approach was to attempt to considerably reduce the levels of RAPID-release RAD21, to a level within the order of endogenous RAD21 expression.

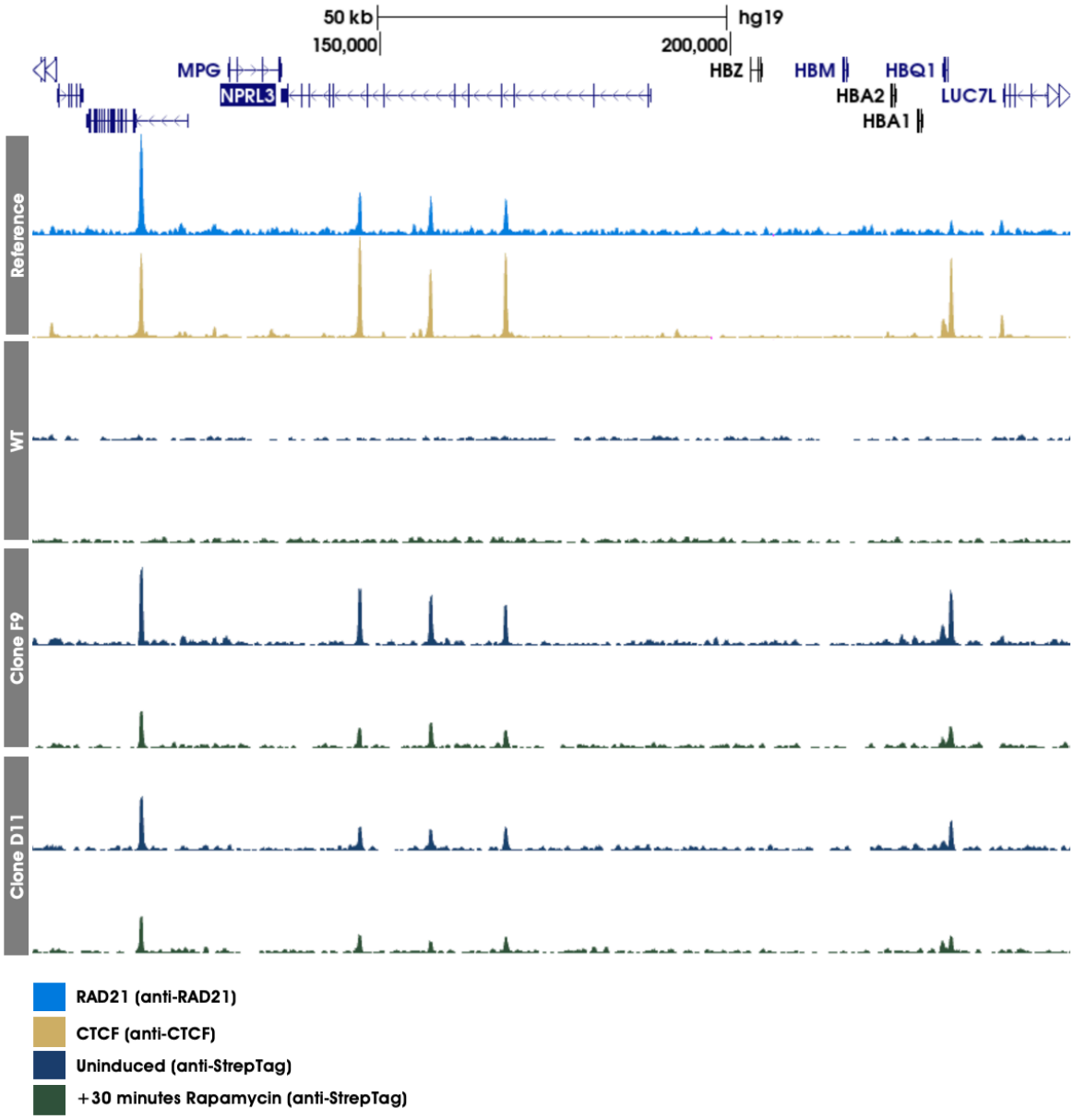



**Figure 4.15: Comparison of RAD21TST ChIP-seq in RAPID-release HEK293 cells F9 and D11 before and after the addition of Rapamycin.**

Two RAPID-release HEK293 clones were selected for comparison, along with an unedited WT line. Cells were fixed prior to the addition of Rapamycin (uninduced, dark blue), and 30 minutes after the addition of Rapamycin (+30 minutes, dark green). ChIP-seq was performed, using an anti-StrepTag antibody to selectively precipitate the RAPID-release RAD21$^{TST}$. A track showing WT HEK293 RAD21 ChIP-seq using an anti-RAD21 antibody is shown in light blue as a comparison. CTCF ChIP-seq is shown in gold as a reference. Samples were calibrated using a 4% spike in of RAD21$^{TST}$ mES E14 cells. Region shown is chr16:100,205-248,932 (hg19). Data sources for reference tracks available in Table 6.16.

### 4.2.4 A new strategy for RAPID-release in mESCs

Given all the data generated thus far, it is clear that a more controlled method was required to engineer a cell line to express the RAPID-release system so as not to overwhelm the cells with $RAD21^{TST}$. Originally HEK293 cells were selected for use in these experiments as they are particularly amenable to transfection, however this was likely to be one of the factors responsible for the high copy number problem which was not initially anticipated. I also speculated that the CMV promoter used to drive expression of the RAPID-release machinery, being a well reported strong promoter[200], could be a factor responsible for the extreme over expression causing the $RAD21^{TST}$ saturation in the nucleus. I therefore decided to try using a mouse embryonic stem cell line which has been well studied in the lab and for which the transfection strategy has been extensively optimised. The additional benefit of using this cell line was the plethora of genomic data (ATAC-seq, RNA-seq, ChIP-seq, Hi-C etc) that we have either generated as a lab or is publicly available for this cell line, which can be made use of in the downstream analysis. There have been a number of studies demonstrating that the CMV promoter is silenced in mESCs [201,202], therefore the original CMV promoter was exchanged for a minimal EF1α by AseI/AfeI PCR-cloning by Philip Hublitz (WIMM, Genome Engineering Services). The new construct containing the EF1α promoter was then transfected into the cells, and the same strategy was used to create stable lines with random insertion as was used for the HEK293 stable line generation (see Materials and Methods for full details).

I selected two positive mES clones to quantify the copy number of integrated RAPID-release DNA compared to the F9 HEK293 clone. To quantifiably compare the expression levels between the mESC and HEK293 clones, RNA was extracted from one HEK293 clone (F9) and two mESC clones (E5/E6) and PCR was performed. Figure 4.16 shows the DDCt analysis in which fold change of two components of the RAPID-release machinery (FRB and FKBP), were normalised against a housekeeping gene (GAPDH) and the log(fold-

change) across clones was plotted. The fold-change difference can be used as a proxy for copy number, as it is expected that clones that have a higher number of copies of the plasmid DNA integrated into their genome will have higher expression levels of RAPID-release. As expected, the HEK293 clone has significantly higher levels of FRB and FKBP suggesting this clone has integrated more copies of the plasmid DNA than the mESC clones. Although the fold change in the mESC clones was considerably lower than in the HEK293 clone, the fold change suggested that they still exhibit supra-physiological levels of RAD21 which would most likely be beyond a 1:1 ratio of RAD21:RAD21$^{TST}$. Consequently, I decided to adopt a more controlled mode of genome engineering using a CRISPR-cas9 mediated homology directed repair (HDR) strategy.

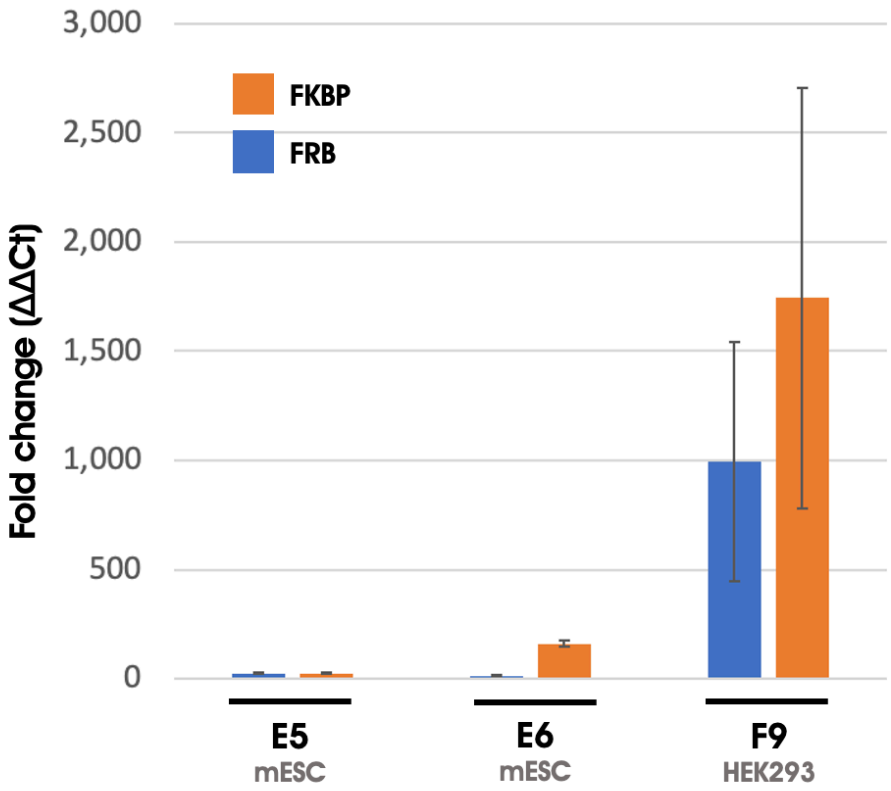



**Figure 4.16: DDCt RT-qPCR analysis to investigate copy number variation in mESC RAPID-release clones compared to HEK293 clone.**

qPCR primers were designed against FRB and FKBP (each representing a half of the RAPID-release machinery). GAPDH was used as a housekeeping gene to normalise cDNA input. Three qPCR technical replicates were performed per ChIP replicate, error bars indicate +/- standard error (SE).

### Using CRISPR-cas9 to insert a single copy of the RAPID-release -release machinery

For this alternative strategy, CRISPR-Cas9 mediated homology-directed repair (HDR) was used to introduce the RAPID-release machinery into a neutral region on chromosome X in mES cells. This region (chrX:11238556-11240251) was extensively characterised by Caz Harrold [147] and shown to be neutral (see Chapter 2), therefore any inserted genes are theoretically amenable to expression and their insertion will not interfere with any intrinsic cellular processes. The additional benefit of the chromosome X region is that the mES cell line used in this thesis is male: correctly edited cells will therefore only carry one copy of RAPID-release RAD21$^{TST}$ which should help to achieve as near as possible to endogenous level of expression of RAD21$^{TST}$. A number of modifications were required in order to redesign the RAPID-release vector for CRISPR-cas9 engineering. Firstly, the 6.3 kb region containing the RAPID-release machinery was excised from the original plasmid using AseI and MluI restriction endonucleases (REs) and cloned into a donor backbone between two ~500 bp homology arms that have been previously validated for the target site[147], Figure 4.17. It was decided to include LoxP sites in the design so that RAD21 can be easily exchanged for different factors for future experiments. The sgRNA used was one previously shown to have the highest targeting efficiency to this region [147]. This sgRNA was cloned into pX458 [203], a plasmid encoding Cas9 from Streptococcus pyogenes as well as enhanced green fluorescent protein (EGFP) to enable FACS sorting of transfected clones. All cloning carried out in this project was done by Philip Hublitz (WIMM, Genome Engineering Services).

Given that both the donor and the guide plasmids contain fluorescent markers, in order to determine the degree of fluorescence exhibited by cells transfected with each counterpart and whether this effect was additive, three separate transfections were carried out and taken for FACS 48 hours post transfection with wild-type cells used as a negative control (see Figure 4.18). Slightly unexpectedly, the cells transfected with the

donor only showed no EGFP, or mCherry signal (mCherry data not shown). Cells transfected with the guide only exhibited a similar level of EGFP expression as the cells transfected with both the donor and the guide. Consequently, it was assumed that cells that were co-transfected with the donor and guide plasmid that were EGFP+ would have been subject to HDR, I therefore used FACS to isolate EGFP+ single cells for clonal expansion. Surviving clones were replica plated prior to PCR screening. After extensive screening by PCR, only WT sequences were amplified indicating that there were no successfully edited clones. Figure 4.19 shows the screening strategy and an example of the screening from the 3' end in which only WT bands were identified. Using this CRISPR-cas9 HDR strategy I was unable to successfully edit the RAPID-release transgene into mES cells. Due to time constraints, I did not pursue this model any further, in the discussion I will touch upon a number of avenues that could be explored in future work.

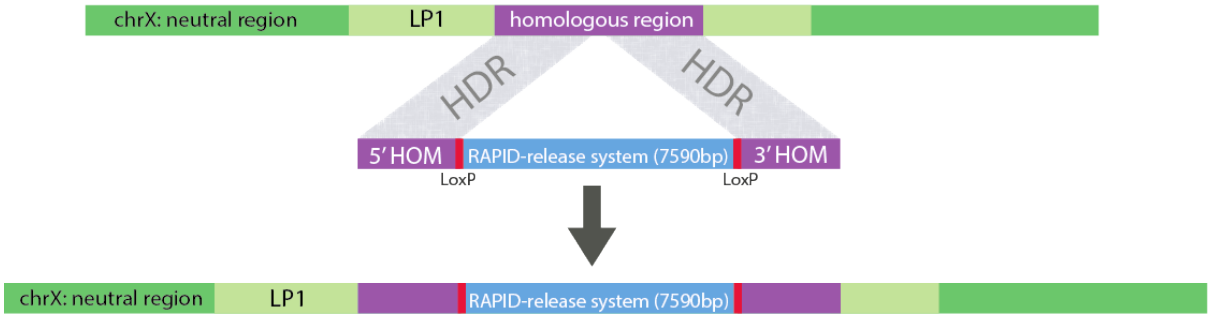



**Figure 4.17: the CRISPR-cas9 homology directed repair (HDR) strategy for the insertion of the RAPID-release machinery into a landing pad that lies within a neutral region of chromosome X.** The RAPID-release machinery is inserted into a donor backbone that contains ~500bp arms (5'HOM and 3'HOM) that are complementary to a homologous region within the landing pad (chrX:11238556-11240251). By CRISPR-cas9 editing the RAPID-release machinery is inserted by homology directed repair. This new design also contains LoxP sites (shown in pink) which allow for easy exchange of cargo for future experiments, bypassing the requirement for further engineering.

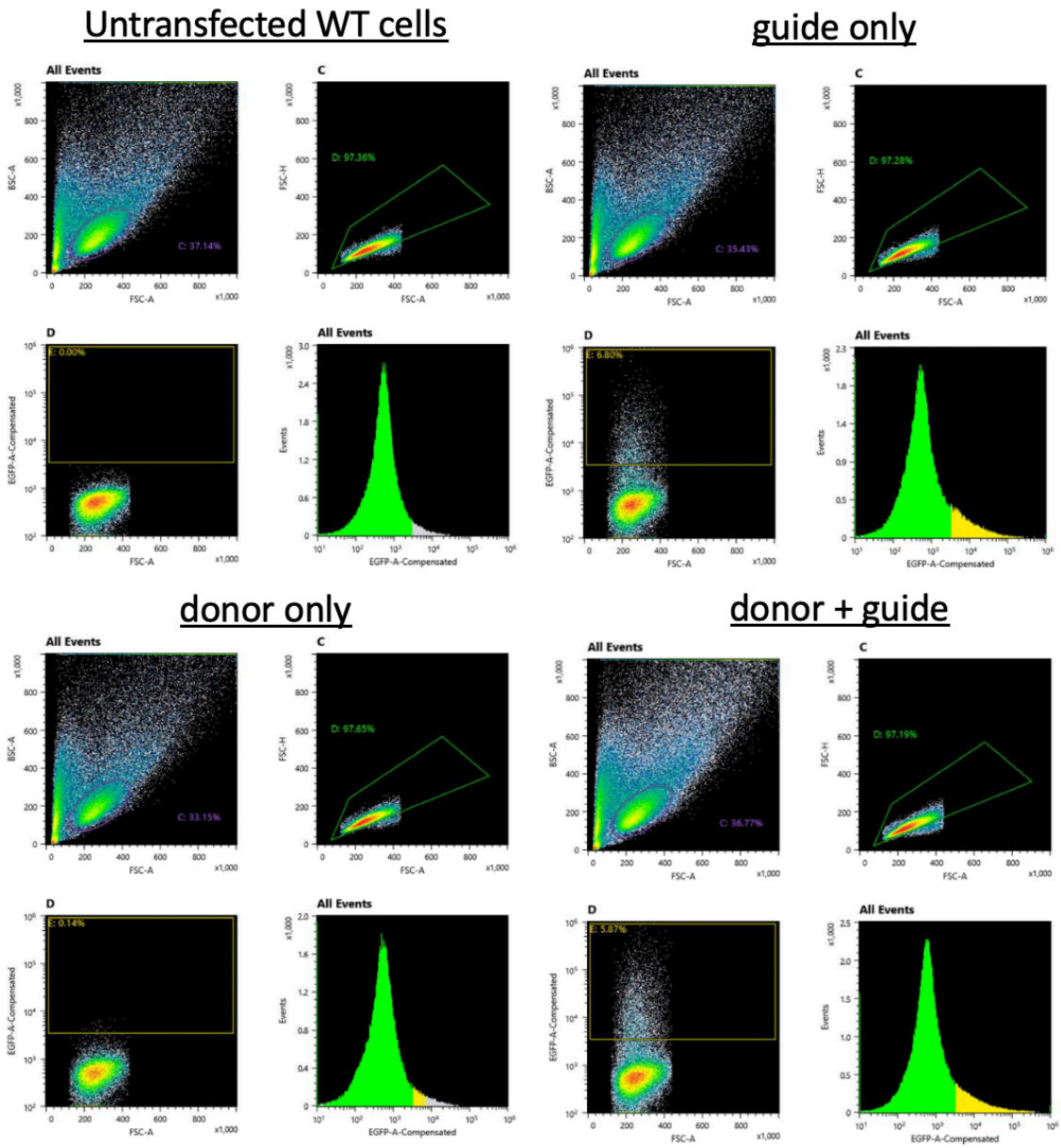



**Figure 4.18: Plots showing the fluorescence activated cell sorting (FACS) gating strategy for the CRISPR-cas9 RAPID-release mES E14 cells 24hrs post transfection.**

From top-left clockwise: the first panel shows WT untransfected cells for which no EGFP signal is shown (E: 0.14%), this is the negative control for gating. The next panel shows cells transfected with the guide only, the pX458 guide contains EGFP and this acts as the positive control for gating (E: 6.80%). Cells transfected with the donor only do not exhibit any EGFP signal (E: 0.14%). Finally, cells transfected with the donor and the guide show similar levels of EGFP signal to the guide only cells (E: 5.87%).

| Primer set | Forward primer | Reverse primer | Size (bp) WT | Size (bp) Edited | Notes |
|---|---|---|---|---|---|
| External | chrX 5'EXT FW | chrX 3' EXT RV1 | 1161 | 7732 | WT band only seen if unedited (edited band too large for PCR) |
| Unedited 5HOM 1 | chrX 5'EXT FW | Between HOM arms RV | 610 | - | Only present in WT |
| Unedited 5HOM 2 | 5HOM FW2 | Between HOM arms RV | 161 | - | Only present in WT |
| Unedited 3HOM 1 | 5HOM FW1 | chrX 3' EXT RV3 | 584 | - | Only present in WT |
| Edited 5HOM | chrX 5' EXT FW | JH_023 EF1a RV | - | 649 | Position dependent |
| Edited 3HOM 1 | JH_023 3' INT FW1 | chrX 3' EXT RV1 | - | 637 | Position dependent |
| Edited 3HOM 2 | JH_023 3' INT FW2 | chrX 3' EXT RV2 | - | 624 | Position dependent |
| Edited 3HOM 3 | JH_023 3' INT FW3 | chrX 3' EXT RV3 | - | 649 | Position dependent |
| Edited 3HOM 4 | JH_023 3' INT FW4 | chrX 3' EXT RV4 | - | 642 | Position dependent |

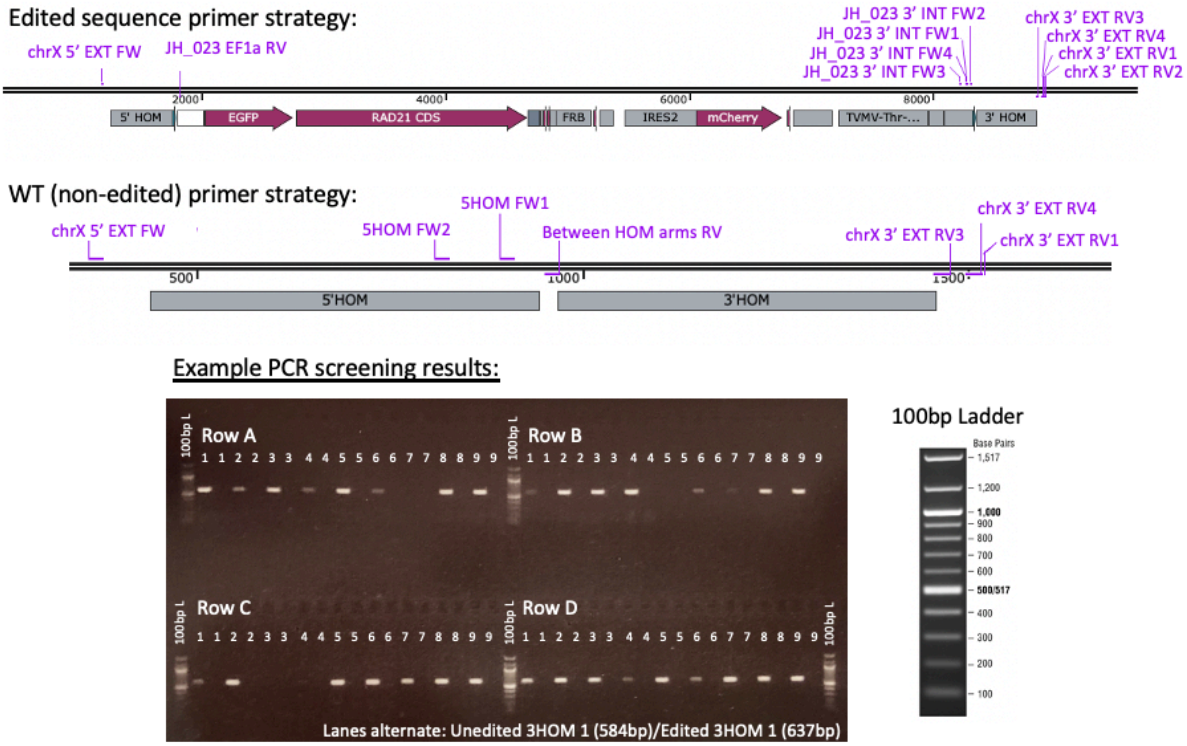



**Figure 4.19: Colony screening by PCR strategy for RAD21 RAPID-release (EF1a) in mESCs.**
Table shows the pairs of screening primers used for screening, their binding region and the expected band size. The two diagrams below show where these primers bind in the edited versus WT (non-edited) sequences. The agarose gel image is an example of one of the PCR screens using 'Unedited 3HOM 1' and 'Edited 3HOM 1', showing that none of the clones contained the edited sequence.

## 4.3 Discussion

In Chapter 2 I showed that when an enhancer element is taken out of its native context and isolated in a neutral region devoid of other regulatory elements, it is able to direct the formation of a tissue-specific regulatory domain structure similar to that seen in its original context. Having shown that a single active element can recruit cohesin at an isolated locus, in this chapter I wanted to scale up to investigate cohesin recruitment to active elements in their native locations on a genome-wide scale. To do so I set out to develop a novel induction method that can be tracked by ChIP-seq.

The RAPID-release technology that was modified for use in this chapter was appealing given its proven rapid and controllable sequestration of the target protein, and the ability to trace the protein localisation by microscopy and ChIP-seq. I was able to engineer both HeLa, HEK293 and mES cell lines to express RAD21 RAPID-release and demonstrated that the system could be rapidly induced by the addition of Rapamycin at a similar rate to the original system containing H3.1. To accurately assess and precisely monitor naïve cohesin recruitment, a true ground state was required in which no $RAD21^{TST}$ was present in the nucleus. All models generated in this chapter exhibited extremely high levels of the RAPID-release machinery, presumably overloading the system, and causing $RAD21^{TST}$ to be detectable in the nucleus prior to induction by microscopy, western blot, and ChIP-seq. Numerous approaches were taken to attempt to dial down the expression including the use of different cell lines, substituting promoters, and using targeted genome engineering strategies. Unfortunately, none of these approaches were able to generate a cell line that expressed the RAPID-release machinery at a level suitable to proceed with the planned downstream experiments. Given the time restraints of this thesis, I could not pursue this method any further. I will however, highlight and discuss a few factors that I believe

contributed to the difficulties in obtaining a successful model and how they could be resolved in the future.

### 4.3.1 Mitochondria counts

Mitochondria numbers may vary between cell types and within clonal populations of the same cell line, for example in a HeLa cell line mitochondria count can vary between 383 and 882 per-cell [204]. For the RAPID-release model to work effectively, all the RAD21$^{TST}$ should be tethered to the OMM prior to the addition of Rapamycin. One potential cause of the leakage of RAD21$^{TST}$ into the nucleus could be the saturation of the mitochondrial membranes with tethered proteins. In this chapter I did not quantify the mitochondrial count in the cells, working on the assumption that this would not be an issue if using the same HeLa cell line as was used in the original publication[189].

### 4.3.2 Nuclear export signal (NES) versus nuclear localisation signal (NLS)

A vital part of the RAPID-release design was the fine-tuning and balance of the transgenic NES and endogenous RAD21 NLS. For this method to work, an artificial NES was added to EGFP- RAD21$^{TST}$-FRB-OMP25 to override the endogenous RAD21 NLS and ensure it is not directed to the nucleus before it is tethered to the mitochondria[189,205]. When Rapamycin is added, the protein is cleaved to leave the NES behind and allow the endogenous RAD21 NLS to dominate. Evidence of RAD21$^{TST}$ residing in the nucleus prior to induction could suggest that the artificial NES is not overriding the endogenous NLS. There is a database of NES sequences (NESdb)[206] which could be used to re-design a stronger NES into the system that would override the NLS. This could be done in combination with modifying the endogenous NLS to a weaker version, although this would require a more challenging editing strategy.

### 4.3.3 Size of RAPID-release machinery

One of the major limiting factors of engineering RAPID-release into the genome of any model cell line is the large size of the plasmid DNA: the machinery alone is 6.3 kb in size without taking into account vector backbone or the requirement of homology arms for CRISPR-cas9 editing. It is well known that the vector size has a significant impact on editing efficiency [193] and is most likely the major contributing factor to the lack of successful clones obtained using CRISPR-cas9 mediated editing of the mES cells in this Chapter. A potential way to circumvent this issue would be to use a 2x single vector system: 1x vector (EGFP- RAD21$^{TST}$-FRB-OMP25) inserted into the neutral region on chromosome X by CRISPR-cas9 mediated HDR; 1x vector (mCherry-FKBP12-TVMV-AI) randomly but stably integrated using lipofection. Using this single vector strategy reduces the plasmid DNA size by half, thereby increasing the chance of successful edits. The limiting factor for RAPID-release would be the copy number of the RAD21$^{TST}$ transgene, and it therefore should not matter how many copies of the mCherry-containing transgene were present.

The RAPID-release method used in this chapter came with numerous challenges, which despite a number of iterations of the original design made it unsuitable for use in investigating the dynamics of cohesin loading. However, the techniques that I developed in this chapter have been directly transferable for use in Chapter 3 in which I developed an orthologous inducible system for investigating cohesin dynamics.

## 4.4 Data availability

Movies from RAPID-release movies are available on YouTube,

scan QR code for access or visit: https://bit.ly/3iDuJ88

## 4.5 Acknowledgments

The original RAPID-release plasmids were kindly gifted by Andrew Bowman (University of Warwick), Alonso Pardal from the Bowman lab (University of Warwick) advised on experimental set-up for this system. Philip Hublitz (WIMM, Genome Engineering Services) cloned and validated all constructs used in this chapter. Nigel Roberts (WIMM, Hughes lab) produced and validated the RAD21$^{TST}$ WT cell lines. I worked together with Jill Brown (WIMM, Buckle lab) to optimise the imaging set-up in this chapter. The Milne lab (WIMM) provided the ChIP-seq protocol which I adapted for use throughout this thesis. Nadya Fursova (University of Oxford, Klose lab) provided the original scripts and advice for the analysis of calibrated ChIP-seq data.

# 5 Discussion

## Dynamic gene regulation is evolutionarily important

Throughout evolution, the complexity of organisms vastly increased from single-cell, to complex, multicellular, multi-tissue organisms. For example, the simple unicellular organisms, *Saccharomyces cerevisiae* (yeast) has a genome size of 12 Mb, which contains 6,241 genes[207]. *Homo sapiens* (human) in comparison have a 3,400 Mb genome, containing ~20,000 protein-coding genes[207,208]. The nucleus of almost every cell in an organism contains the same DNA, therefore as the organism becomes more complex, there becomes a need to regulate which genes are expressed in which cell type, which we now know is governed to some degree by enhancer-promoter contacts. As the genome gets larger, often the density of genes decreases, which usually correlates with an increasing distance between genes and their regulatory elements. Topologically associated domains (TADs) are a way by which complex organisms can control the regulation of gene transcription and enhancer-promoter contacts, by delimiting the genome into "neighbourhoods" which contain specific sets of regulatory elements. Within these TADs, enhancers and promoters are segregated from the rest of the genome, which not only prevents non-desirable regulatory interactions from outside the domain, it also increases the frequency with which elements within the TAD are able to interact. It has also been suggested that TADs provide protection against mutations, by which a change in the DNA sequence is insulated from the whole genome context and only impacts the inter-TAD environment. Additionally, it has been shown that genomic rearrangements

that disrupt TAD structure can generate ectopic enhancer-promoter contacts and misexpression, which in some cases, can be disease causing[209].

There may be different mechanisms for the formation of TADs that do not rely on the presence of CTCF, which is commonly considered to demarcate the TAD boundaries. For example: CTCF has not been found in yeast, plants, or *C. elegans*, although there is evidence of topologically associated domains in these organisms[210-212]. This suggests that domain formation is not reliant on CTCF. Instead, it could be that the presence of CTCF allows the formation of dynamic domains which could facilitate complex gene regulatory patterns. This provides evidence for the existence of alternative mechanisms for topological organisation and confirms the evolutionary importance for the division of the genome into topological domains.

We have seen that the formation of tissue-specific sub-TADs may arise gradually during differentiation. The formation of sub-TADs is also thought to be mediated by cohesin-driven loop extrusion. The fundamental concept underlying this thesis is that an understanding of cohesin dynamics, particularly loading, would provide an insight into how tissue-specific domain formation is regulated. A major question in the field is: where is cohesin loaded onto the chromatin? A comprehensive understanding of the mechanism of cohesin loading would help to understand it's role in orchestrating tissue-specific looping and enhancer-promoter interactions (Figure 5.1). The most widely adopted hypothesis is that cohesin is actively loaded by NIPBL-MAU2 at active regulatory elements, although at which element, is up for debate: some think that cohesin is loaded at the promoter[82], others at the enhancers[28,75,76]. Yet others have proposed that cohesin is loaded randomly across the genome[213]. The rest of this chapter will be dedicated to the discussion of these various hypotheses and the current evidence to support them. I will

also discuss a few caveats of the methods presented in this thesis, and the future perspectives.

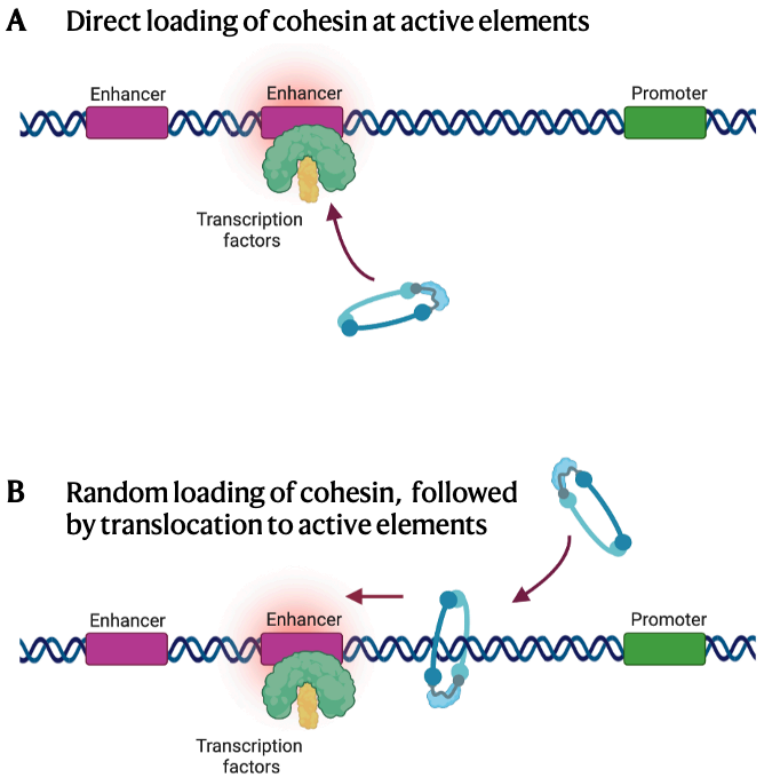



**Figure 5.1: Is cohesin loading directed or random?**
Figure shows two potential situations, which are most likely indistinguishable by regular ChIP-seq methods. **a** | Cohesin is loaded directly at active regulatory elements. **b** | Cohesin is randomly loaded along the chromatin and translocates until it reaches an active element, at which it stalls. Figure created using BioRender.

## Cohesin loading is directed to specific elements

Using the R2 enhancer model in Chapter 2, I showed that upon activation of the erythroid-specific enhancer, cohesin was recruited to this locus, and confirmed the formation of the 3D domain structure using the high-resolution Tiled-C approach. At first thought, this would appear to provide evidence for cohesin loading at enhancers: the insertion of the 325bp erythroid enhancer sequence was the only factor that distinguished it from wild-type, given that there is no evidence of cohesin occupancy at the WT locus, it would seem

that the enhancer element is responsible for recruiting cohesin. However, this observation in complicated by the evidence that the R2 enhancer element, when inserted into the neutral chromosome X region, exhibited promoter-like histone marks in contrast to the enhancer-like marks seen when in its native position on chromosome 11. This suggests that there is an element of context dependency to whether an element acts as an enhancer or promoter, which is not purely reliant on the DNA sequence and transcription factor binding motifs present in the element as is typically thought[214]. In a similar vein, a recent pre-print by Blayney *et al.* [215], in which they dissected the alpha-globin super-enhancer to investigate the component enhancer functions, found that the strength of the capacity of individual enhancers to rescue alpha-globin expression was to some extent determined by its position in the locus, and not exclusively based on DNA sequence. From their findings, they suggested that certain enhancers within the super-enhancer cluster may act as "facilitators", which play a role in increasing the interaction frequency between the enhancers and the promoter, rather than acting as classical enhancers. In the case of the R2 enhancer model presented here, the insertion and activation of R2 appeared to activate downstream enhancer-like elements, to create a distribution of regulatory elements similar to those seen at the alpha-globin locus. It could be that throughout the genome there are primed enhancer-like elements, that become activated when in the vicinity of a highly active regulatory element, irrespective of whether this is a promoter or enhancer, and are able to recruit cohesin, thereby acting as facilitators to set up a regulatory domain. With respect to the R2 model, this could be seen as evidence supportive of directed recruitment of cohesin to active elements. However, cohesin in this case is not directed to a specific sub-class of regulatory elements and can instead be loaded at both enhancers and promoters depending on the context.

In the recent publication by Rinzema *et al.*[80], they state that cohesin recruitment to enhancers is necessary for the activation of distant, but not proximal genes. In this study

they used CRISPR interference (CRISPRi) to deplete two cohesin components: SMC1A and RAD21 in cell lines containing a GFP-reporter with downstream enhancers at varying downstream positions. They reported that following the depletion of cohesin subunits there was a loss in enhancer-induced chromatin contacts. I would argue that although this is evidence that cohesin plays a role in mediating these contacts, it does not necessarily mean that cohesin is directly recruited to these elements.

## Cohesin is loaded in a non-directed manner, and translocates to specific regulatory elements

In opposition of the previous section, is the idea that cohesin is loaded onto chromatin in a non-directed manner. Whilst the proposition of directed cohesin loading seems an attractive one, it is of course speculative. The detection of cohesin at active regulatory elements does not necessarily mean that cohesin was loaded there. It has been shown that there is a direct correlation between the frequency of cohesin mediated interactions and the levels of transcription factors bound at enhancers[213]. In the study by Lyu, Rowley and Corces[213] they suggest that the large assembly of proteins at enhancers act as a physical obstacle which slows down cohesin mediated loop extrusion, leading to an increase in enhancer-promoter interactions and an increased transcription bursting frequency. Similarly, the recent research by Dequeker *et al.*[216] showed that cohesin often stalls as a result of a multi-protein complex bound to chromatin, in this case the minichromosome maintenance (MCM) complex. This suggests that cohesin can load at random sites across the genome and simply progress until it comes across a protein-complex obstacle, at which point it stalls and inadvertently facilitates tissue specific transcription.

In support of this theory of non-directed cohesin loading, a recent paper published by the Losada group[177] suggests that cohesin may not be loaded at specific sites. A widely

accepted model of cohesin loading, and the one presented in this thesis, is that of active loading of cohesin onto chromatin mediated by the NIPBL-MAU2 complex. In this study by the Losada group, the focus was to investigate the different NIPBL requirements of cohesin-STAG1 versus cohesin-STAG2. Their findings suggested that NIPBL is not required for the initial association of cohesin with chromatin but is required for loop extrusion, through its ability to activate the ATPase domain of cohesin and therefore initiate extrusion. If NIPBL is not involved in actively loading cohesin to specific sites, this would explain the inconsistencies seen in the NIPBL ChIP-seq data: the continuous extrusion of chromatin by the cohesin-NIPBL-MAU2 complex would mean that NIPBL ChIP-seq signal would be enriched at the locations at which the complex stalls, such as sites densely populated with transcription factors or boundary elements.

### A new hypothesis

An attractive mechanism that would explain the emerging evidence discussed thus far, would be that cohesin is loaded on chromatin at 'random' locations. By random, I mean to say not at specific regulatory elements or genomic sites; however, such regions of chromatin would have to allow loading of cohesin. Therefore, it could be that cohesin preferentially loads at regions of open-chromatin (as shown in ATAC-seq data). Pioneer transcription factors are responsible for opening up chromatin, to facilitate further protein binding. It is known that regions of open chromatin are present at active regions of the genome, the concentration of cohesin loading at these regions would reduce the amount of non-productive loading. As we saw in the work by Oudelaar and Beagrie[56], ATAC-seq signal increases throughout cellular differentiation; if cohesin is loaded at these sites it might explain how a sub-TAD develops throughout differentiation. Once on the chromatin, NIPBL can associate with cohesin to activate the ATPase domain and facilitate translocation towards the active elements. By loop extrusion, cohesin reels the chromatin through the ring and eventually bring enhancers and promoters into close proximity so

that they can interact with one another. These active elements, already bound with transcription factors provide a temporary block to cohesin, which would explain the longer residence time of cohesin at these sites and why cohesin appears to be enriched at these sites when using ChIP-seq approaches. As previously mentioned, it has been recently proposed that there may be two pools of cohesin within a cell, the stable CTCF bound pool, and the dynamic pool which is offloaded by WAPL[137]. This hypothesis would also explain the pattern seen by the persistent cohesin sites residing at regions correlating with TAD structure.

Once this TAD structure is formed, through an enrichment of shared chromatin modifications within the domain, there is perhaps also a degree of phase separation to further isolate the TAD structure from outside interactions. This would also align with the recent findings by Crump *et al.* [23] which suggest that phase separation is most likely a consequence than a cause of enhancer-promoter contacts. This hypothesis could also explain how cohesin mediated interactions may be maintained. If cohesin molecules become restricted to a compartment of chromatin corresponding to a TAD or sub-TAD, the search time required to re-load cohesin onto this chromatin would be reduced. Such a model might also explain why even after effective depletion, some cohesin remains bound to chromatin. In this situation cohesin would spend less time disassociated from chromatin and it might be less exposed to proteasomal degradation (Figure 5.2).

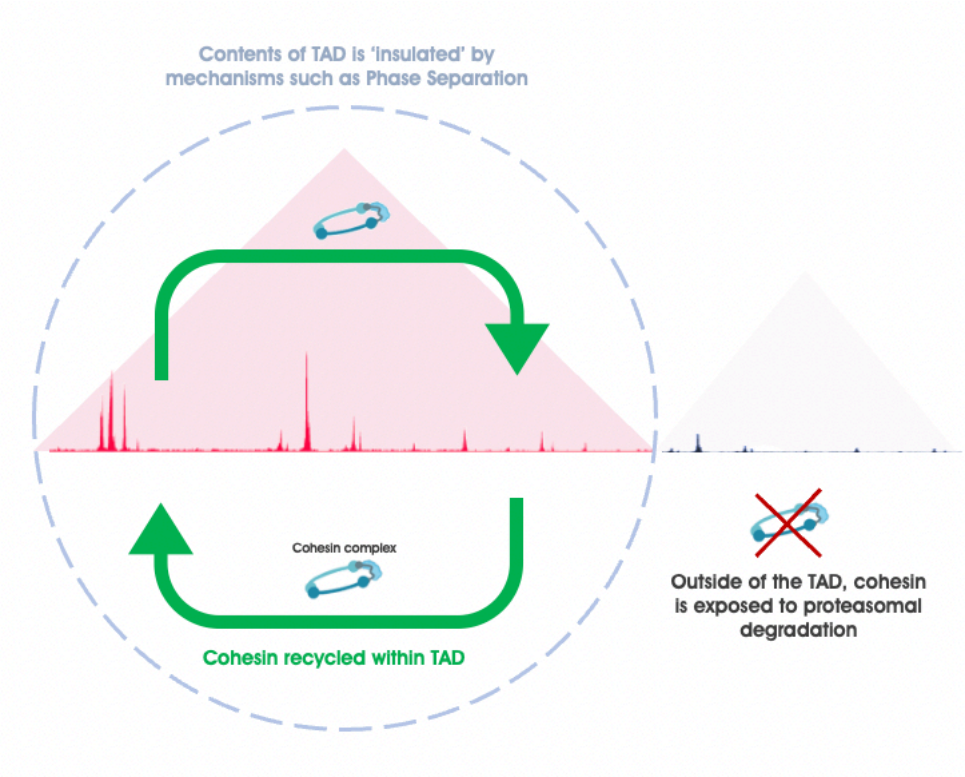



**Figure 5.2: A proposed mechanism for cohesin can be insulated from proteasomal degradation.** Schematic shows how a TAD could be insulated and proteins such as cohesin protected from the environment outside of the TAD structure. See main text for further description.

# The caveats of current methods for studying dynamic genome regulation

## Enhancers and promoters are difficult to classify

When I first started this thesis, I was testing the hypothesis that cohesin was recruited at active enhancers. Although Chapter 2 could not prove this to be the case, it certainly suggested that the presence of active enhancers was somehow linked. Using a degron model (Chapter 3) I obtained genome-wide profiles of cohesin degradation over a 4-hour time-course. Using this data, my idea was to classify the cohesin peaks based on trends over the time-course and co-cluster these with chromatin annotations to determine whether there was any difference in cohesin retention at putative enhancer and promoter elements, or CTCF sites. One of the main difficulties when carrying out this analysis, was

the arduous task of chromatin annotation. It is widely accepted in the field that enhancers are enriched for H3K4me1 and H3K27ac, and promoters are enriched for H3K4me3, however there is known to be an overlap between the two, making their classification less clear-cut. In addition, enhancers can be further classified as inactive, primed or poised, and the activation of these enhancers is linked to an increased in transcription factor binding[213].

## Cohesin dynamics is intertwined with the cell cycle

One of the main confounding effects when studying dynamics is the cell cycle. A fundamental aspect of cell biology that contributes significantly to cellular dynamics is the cell cycle, and this is particularly the case for cohesin. Throughout the cell cycle chromatin is dramatically reorganised: during interphase the chromatin is decompacted and organised into cell-type specific compartments and TADs[34,58,217], during metaphase chromosomes are subject to large-scale compaction, associated with a cessation of transcription and loss of certain boundaries[218-220] (Figure 5.3). Cell cycle analysis of the mESC E14 cell line used in this thesis revealed that approximately 80% of cells reside in S-phase at any given time (Jennifer Herrman, unpublished) and that the cell cycle in vitro takes up to 12 hours[221]. Therefore, it could be assumed that for the majority of studies in non-synchronised cells, results will be predominantly reflective of effects during S-phase. However, given that cohesin is essential for sister chromatid cohesin, its loss may result in mitotic arrest which could change the distribution of cells across the cell cycle stages. A recent study published by Liu and Dekker[222], showed that after 6 hours of cohesin degradation in HCT116 cells the majority of cells were arrested in mitosis. This suggests that if cells are subject to cohesin degradation for a duration that exceeds the average time cells spend in interphase it could lead to an enrichment of cells arrested in mitosis. Consequently, any measured effects on genome architecture will therefore be difficult to disentangle from those naturally occurring during metaphase. Although cell cycle effects

are not the focus of this thesis, given its significant impact on cohesin dynamics and 3D genome architecture it is a confounding effect that must be considered when designing experiments and interpreting results.

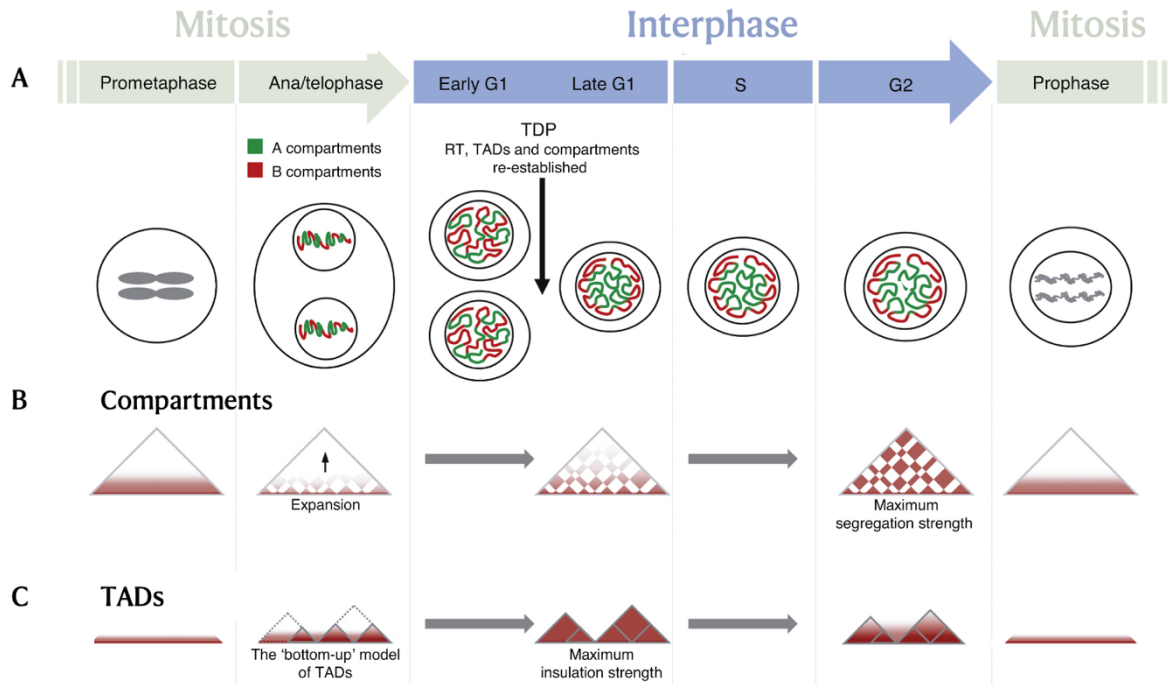



**Figure 5.3: Genome architecture changes throughout the cell cycle.**
Figure adapted from Miura and Hiratani[223]. **a** | Chromosomal compartmentalisation changes throughout the cell cycle. **b** | A/B compartments disappear in mitosis. **c** | TAD structures are dynamic throughout the cell cycle. Replication timing (RT), timing decision point (TDP).

To address this issue, cells can be synchronised either using chemical agents which halt the cell cycle at specific stages depending on the agent used; or by cell sorting strategies that can specifically isolate cells depending on their cell-cycle stage. The first strategy is known to produce undesirable side-effects, thus the potential benefits of having synchronous cells is counteracted by the potential confounding side-effects. The second method is thought to be a much cleaner approach. However, it can be difficult to obtain sufficient numbers of cells in each population required for the downstream experiments, and the sorting procedure can be time-consuming and may require additional rounds of genome engineering. Overall, I believe that attempting to synchronise cells by their cell-

cycle stage introduces additional and potentially unwanted background variation, as well as adding additional time-restraints to the experimental procedure. For experiments assaying highly dynamic processes, the best protocols are those that are quick and highly efficient in order to capture the highest temporal resolution. Whilst cell-cycle effects should be taken into consideration when designing experiments to investigate these processes, advances in single-cell techniques will provide a means of subclassifying cells and removing this confounding effect.

### A subset of cohesin may be resistant to depletion

One of the most notable findings shown in Chapter 3 was that some RAD21 peaks were resistant to depletion, and persisted throughout the course of the dTAG-13 treatment. This finding highlighted the issue that not all RAD21 sites are equal, and there may be subset s of sites that have unique properties. For this reason, I thought it was important to investigate the data from other depletion studies to determine whether this was a universal feature or whether it was an artifact of my study in particular. In Figure 5.4 I have collated some of the data from recent depletion studies. This demonstrates a degree of incomplete depletion in all of the studies, apart from Figure 5.4C in which only western blot data is shown. In my experiments, the western blot data did not reflect what was observed by ChIP-seq and I suspect that this was also the case for the WAPL depletion shown[144].

In a review by Schoenfelder and Fraser[224] they reference a number of studies, which I have presented in Figure 5.4, when stating that only a few protein coding genes display significant expression changes upon acute depletion of cohesin or CTCF. I would argue that without an understanding of which sites retain the protein following depletion, it is premature to draw any conclusions. It could be the case that the non-depleted sites are retained as they are the most essential for structural regulation, and therefore are the

most stably bound and resistant to depletion. In the study by Rao *et al.*[128], contact domains remain following the depletion of cohesin. They suggest that this could be an artefact of incomplete depletion (see Figure 5.4A) or that low-resolution Hi-C is not able to distinguish between loop- and compartment-domains. This aligns with my data, which suggests that the RAD21 sites resistant to depletion align with the TAD structure. Similarly, in the NIPBL depletion study by Haarhuis and colleagues[144], they showed that loop domains remained after the 'near-complete' depletion, but observed that these loops were abnormally small. This supports the idea that NIPBL is not necessarily involved in the loading of cohesin, but instead facilitates translocation of cohesin. The presence of loop domains suggests cohesin is still being loaded even in the absence of NIPBL. However, the smaller loops suggest that cohesin is not extruding cohesin as efficiently, which could indicate that a loss of NIPBL is linked to the translocation and extruding ability of cohesin.

In summary, whilst depletion studies are an excellent tool to investigate the roles of individual proteins in complex multi-protein regulatory systems, the completeness of the degradation should be considered when drawing conclusions. Additionally, validation of the depletion assay should reflect protein occupancy at the chromatin level as I have demonstrated in my work that western blots does not have the dynamic range to detect of low levels of chromatin bound protein.

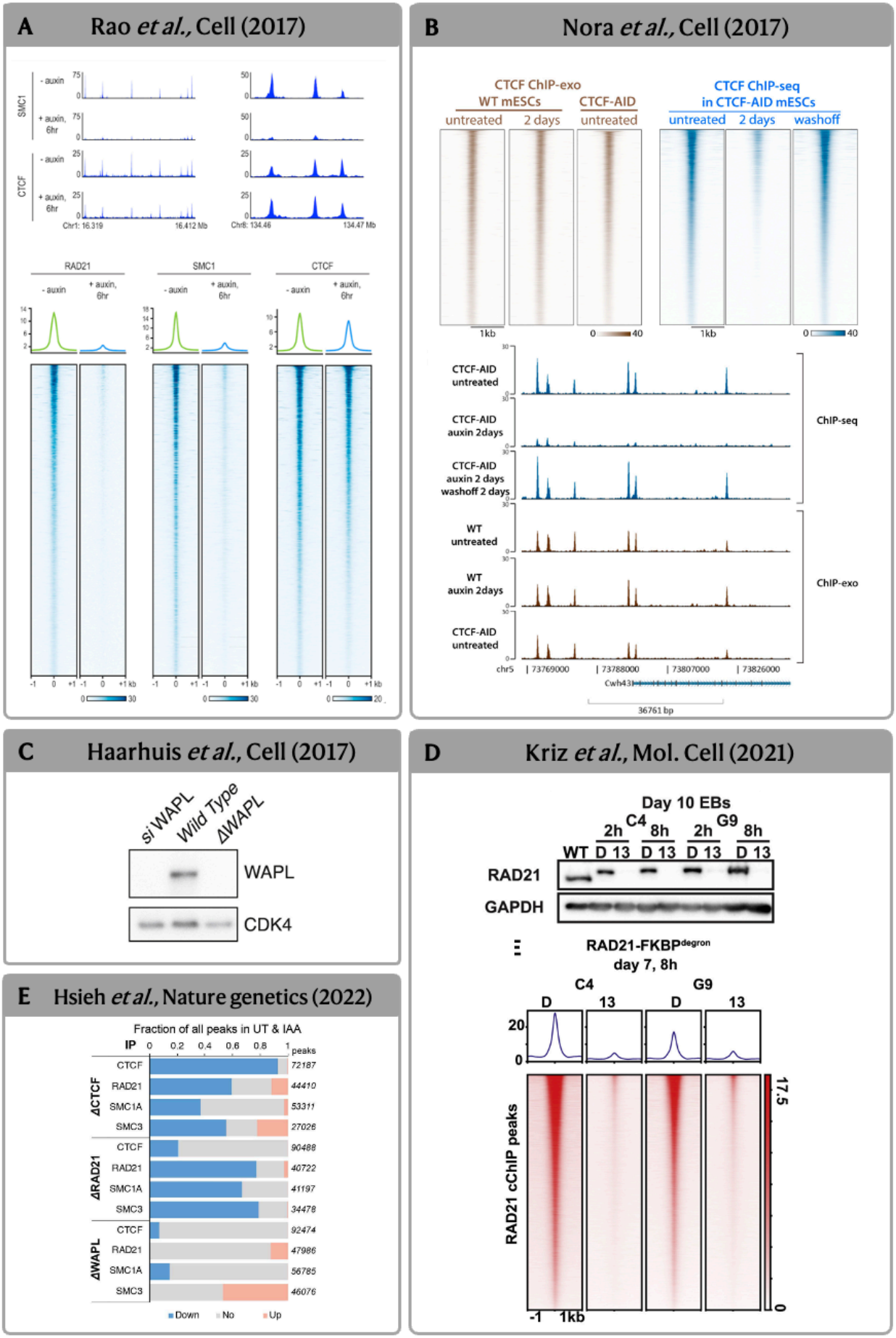

A
Rao et al., Cell (2017)
SMC1
CTCF
- auxin
+ auxin, 6hr
RAD21
SMC1
CTCF
B
Nora et al., Cell (2017)
CTCF ChIP-exo
WT mESCs
CTCF-AID
untreated
2 days
CTCF ChIP-seq
in CTCF-AID mESCs
washoff
CTCF-AID untreated
CTCF-AID auxin 2days
CTCF-AID auxin 2 days washoff 2 days
WT untreated
WT auxin 2days
ChIP-seq
ChIP-exo
C
Haarhuis et al., Cell (2017)
siWAPL
Wild Type
ΔWAPL
WAPL
CDK4
D
Kriz et al., Mol. Cell (2021)
Day 10 EBs
C4
G9
2h
8h
WT
D 13
RAD21
GAPDH
RAD21-FKBP degron
day 7, 8h
RAD21 cChIP peaks
E
Hsieh et al., Nature genetics (2022)
Fraction of all peaks in UT & IAA
IP
peaks
ΔCTCF
ΔRAD21
ΔWAPL
CTCF
RAD21
SMC1A
SMC3
Down
No
Up

**Figure 5.4:** All figures have been adapted from the original publications which are referenced in the figure legend. **a |** In the study published by Rao *et al.*[128] they used the auxin induced degradation (AID) method to deplete RAD21. After 6 hours of treatment, they achieved a near-complete depletion. **b |** Using the AID method for CTCF depletion, Nora *et al.*[122], also demonstrated a near-complete depletion. After 2 days of treatment CTCF could still be detected at low levels and showed smaller peaks in the ChIP-seq tracks. **c |** In the study by Haarhuis *et al.*[144], they only showed the completeness of WAPL depletion by western blot. **d |** Hsieh *et al.*[225], quantified the completeness of degradation in their study, showing the fraction of peaks which went down, up or didn't change in the untreated (UT) and treated (IAA) samples. **e |** Using dTAG-13 mediated depletion, Kriz *et al.*, showed that RAD21 was incompletely depleted following 8 hours of treatment.

# Future perspectives

## Alternative inducible systems

The RAPID-release method is an exciting avenue to explore further and despite the attempts thus far not coming to fruition, I believe there are still possible modifications that could be tried as discussed in Chapter 4. Nigel Roberts in our lab has also begun working on the development of an alternative anchor-away style technique, similar to RAPID-release, using a mutant Estrogen receptor ($ER^{T2}$) ligand-binding domain fusion protein that can be induced using tamoxifen. This type of system has been used in mouse models for inducible gene expression and is therefore much more tightly controlled, we expect this system to not have the same leakage issue we experienced with RAPID-release. It is also a much smaller construct to engineer (~2 kb compared to ~7.6 kb for RAPID-release), therefore increasing the potential editing efficiency.

## Repurposing the cohesin-dTAG model to investigate RNA Pol II loading

Another application for which the cohesin-dTAG model could be used, is to investigate the hypothesis of RNA polymerase transfer. Enhancer RNAs (eRNA) can be detected by Poly(A)- RNA sequencing and have been shown to correlate with RNA Pol II recruitment to the enhancer[226]. This evidence supports the argument that enhancers themselves recruit RNA Pol II which is then transferred to the promoter following direct enhancer-promoter interaction[227-229]. Whilst this theory is not widely accepted, it has been postulated to provide a mechanism by which genes can be maintained in a potentiated state without the risk of premature gene activation[226], and further emphasises the potential importance of DNA looping in transcriptional regulation. Assuming that cohesin is required for enhancer-promoter looping, performing RNA Pol II ChIP-seq in a cell line in which cohesin has been depleted and comparing this to an un-depleted state will show if there is a change in the distribution of RNA Pol II. If it was indeed the case that RNA Pol II was transferred

from enhancers to the promoters by this mechanism, I would expect that following the depletion of RAD21, less RNA Pol II would be visible at the promoters by ChIP-seq.

### Single-cell techniques are required to truly understand dynamic systems

Advancements in single-cell technologies, such as single cell Hi-C (scHi-C) have enabled researchers to investigate genome architecture at an unprecedented cellular resolution[230], and further elucidated the dynamic nature of genome organisation. Although presently, scHi-C has a poor genomic resolution, given the sparsity of the data points, it has emphasised the heterogeneity that is present across individual cells. For example, at the single-cell level TAD structures at individual loci can differ[231]; it is only when single-cell data is combined that the TAD structures observed in bulk Hi-C data is revealed. Whilst it is perhaps not surprising that combining the single-cell data into a pseudo-bulk dataset recreates the original TAD structure, it does highlight the fact that this data represents population average contact frequencies and domains in which contacts are most frequently constrained. In individual cells, these domain boundaries may be somewhat permeable, giving rise to the differences seen in the scHi-C data[231,232]. This is may reflect the stochastic and rapid nature of CTCF binding[110], which acts to delimit TAD boundaries and block loop extrusion. A momentary lack of CTCF binding at TAD boundaries could allow cohesin translocation to bypass the boundary and facilitate genome interactions that are less frequently observed.

Overall, in this thesis I have developed methods able to begin to bridge the gap between imaging techniques that assay dynamic processes, and the genomic techniques able to provide genome-wide resolution. Through further developments of these techniques, coupled with advancements in single-cell technologies, I hope that an understanding of dynamic cohesin recruitment and how this facilitates tissue-specific genome organisation will be achieved in the not-too-distant future.

# 6 Materials and methods

This chapter provides an overview of the experimental methods used to generate the data presented in this thesis. Experiments that were performed by others, for which I have presented and used the data, I have referenced the original publication or thesis work where the full methodological details can be found.

## 6.1 Cell culture

### 6.1.1 General culture procedure

All cell culture media was filter-sterilised using a 0.22 µm filter unit (Corning: 431097). For routine passaging, cells were dissociated using 0.05 % trypsin-EDTA (Gibco: 25300054), quenched with pre-warmed media, pelleted at 1200 rpm for 5 minutes and seeded into culture flask containing pre-warmed media. Cells were cultured in a humidified incubator with 5% $CO_2$ at 37°C. For storage, cells were frozen down in freezing media containing standard culture media supplemented with 20% foetal bovine serum (FBS)(Gibco: A4766801) and 10% dimethyl sulfoxide (DMSO) (Sigma: D2650) and temporarily stored at -80°C before being moved to long-term storage in liquid nitrogen (~160°C).

### 6.1.2 Mouse embryonic stem cells

The ES-E14TG2a.IV mouse embryonic stem cell (mESC) line is used for all mESC work in this thesis. This cell line is a subclone of the male HPRT-deficient male ES-E14TG2a cell line, originally derived from the 129/Ola mouse strain [233,234]. This cell line was cultured feeder-free and was originally provided by Austin Smith (Wellcome/MRC Cambridge Stem Cell Institute) and maintained in-house by Jackie Sloane-Stanley (WIMM Transgenics Facility). Cells were expanded and frozen as aliquots in liquid nitrogen as stocks. Adherent mESC E14 cells were grown in tissue culture plates coated with 0.1% Gelatin (Sigma: G1393)in PBS and cultured in Glasgow's Minimal Essential Medium (GMEM BHK-21; Gibco: 21710-082) supplemented with 10% foetal bovine serum (FBS)(Gibco: A4766801), 1 mM sodium pyruvate (Gibco: 11360-039), 1x minimum Essential Medium Non-Essential Amino Acids (MEM NEAA; Gibco: 11140-035), 2 mM L-Glutamine (Gibco: 25030024), 1000 U/mL LIF (Cambridge Biosciences: GFM200-100), 0.1 mM 2-Mercaptoethanol (Gibco: 31350-010).

### 6.1.3 Human embryonic kidney 293(T) cells

Wild type (WT) human embryonic kidney 293T (HEK293T) cells were gifted by Ryan Beveridge (WIMM, Genome Engineering Services); WT human embryonic kidney 293 (HEK293) cells were gifted from Lise Chauveau (WIMM, Rehwinkel Group), these cell lines were expanded and frozen as aliquots in liquid nitrogen as stocks. $RAD21^{TST}$ cell lines were targeted using WT cells from ATCC (ATCC: CRL-1573). HEK293 HEK293(T) cells were grown in Dulbecco's Modified Eagle Medium, high glucose (DMEM; Gibco: 41965-039) supplemented with 10% foetal bovine serum (FBS)(Gibco: A4766801), 1x minimum Essential Medium Non-Essential Amino Acids (MEM NEAA; Gibco: 11140-035) and 2 mM L-Glutamine (Gibco: 25030024).

### 6.1.4 Human HeLa cells

HeLa S3 (ATCC: CCL2.2) cells were purchased from ATCC, expanded and stored in aliquots in liquid nitrogen as stocks. HeLa cells were grown in Dulbecco's Modified Eagle Medium, high glucose (DMEM; Gibco: 41965-039) supplemented with 10% foetal bovine serum (FBS) (Gibco: A4766801).

### 6.1.5 Differentiation of mESCs into embryoid bodies (EBs) for the isolation of CD71+ erythroid cells

mESCs were differentiated into EBs, and CD71+ cells isolated on day 7 using the recently published protocol from the Higgs group (WIMM)[161] (Figure 6.1).

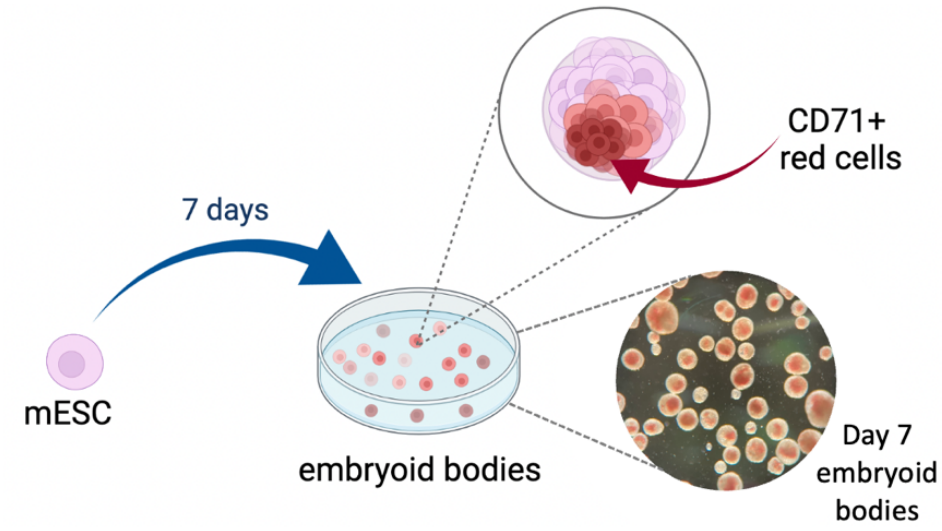



**Figure 6.1: In vitro erythroid differentiation.**
mESCs can be differentiated into embryoid bodies (EBs) from which CD71+ primitive (embryonic) erythroid cells can be isolated at day 7. The CD71+ erythroid cells isolated using this method have been extensively characterised to confirm that they are similar to the primitive erythroid cells in mouse embryonic blood [147,161], and therefore provide an excellent system in which to study the R2 insertion model that does not require the laborious generation of mouse models. Figure made using BioRender.

### 6.1.6 Expansion of single cells from FACS

Sorting media was made by combining 45% regular culture media, 45% conditioned media and 10% foetal bovine serum (FBS)(Gibco: A4766801), and then filter sterilised using a 0.22 µm filter unit (Corning: 431097). Conditioned media was obtained by culturing WT cells and collecting the culture media when the cells were in log phase of growth. Conditioned media was centrifuged to removed dead cells and debris and stored in 50 mL Falcon tubes at -80°C before use. For mES cells, sorting plates were gelatin coated prior to FACS (for details see description of mESC culture above). Using FACS, single cells were sorted into 96 round-bottom well plates, containing 100 µL sorting media, pre-warmed to 37°C. Plates were centrifuged at low speed for 1 minute prior to incubation. Wells were supplemented with 100 µL of fresh mESC culture media 5 days post-sorting.

## 6.2 Genome engineering strategies

Sequences of sgRNAs and donors used for CRISPR-cas9 editing, and primers used for screening across all three chapters are provided throughout this chapter. Additional plasmid maps are provided in Supplementary Materials (Figure S7-S9).

### 6.2.1 Artificial Locus

The clonal mESC line containing the R2 enhancer insertion at chromosome X that was used in Chapter 1 of this thesis was designed and developed by Caz Harrold[147]. Endogenous RAD21 was tagged in mESC with a twin StrepTag (TST) by Nigel Roberts (WIMM, Hughes group). Sequence for sgRNA and donor plasmids are shown in Table 6.1 and Table 6.2.

### 6.2.2 dTAG degron system

The tagging of the native RAD21 gene with FKBP12$^{F36V}$ in mESC as previously described[125] was achieved using CRISPR-Cas9 mediated HDR in WT mESCs. Philip Hublitz (WIMM, Genome Engineering Services), cloned the sgRNA into pX458[203], a plasmid encoding Cas9 from Streptococcus pyogenes and enhanced green fluorescent protein (EGFP) (Table 1). HDR templates containing the FKBP12$^{F36V}$ tag were synthesised and cloned by GeneArt (Invitrogen, ThermoFisher) into double-stranded plasmid vectors (pMX). sgRNA/Cas9 plasmid DNA and donor plasmid DNA (Table 6.2) were prepared for transfection using QIAprep Spin Midiprep Kit (Qiagen: 12943) according to manufacturer's instructions. mESCs (~1x10$^6$) were transfected using Lipofectamine LTX Reagent with PLUS Reagent (Invitrogen, ThermoFisher: 15338100) which were performed in 6-well format by plating freshly trypsinised cells in 2 mL media supplemented with the transfection mix prepared following manufacturer's instructions: 5 µL LTX reagent, a total of 2.5 µg purified plasmid DNA (1:3 ratio of sgRNA/Cas9 plasmid DNA:donor plasmid DNA, 0.625 µg and 1.875 µg respectively), 2 µL PLUS reagent, and 250 µL Opti-MEM (Gibco: 31985062). After 24 hours incubation at 37°C, transfected cells were rinsed with PBS and given fresh media. Transfected cells were selected based on the presence of EGFP by FACS after 48 hours of incubation at 37°C. EGFP+ cells were single cell sorted into gelatin coated 96 well tissue culture plates. Following 10 days of growth, surviving colonies were split across three replica-plated, once confluent, two plates were frozen in freezing media at -80°C, the third plate was used to screen for positive clones by PCR.

### 6.2.3 RAPID-release

The original plasmid (EGFP$^{TVMVx2}$-FRB-OMP25/mCherry-FKBP12-TVMV-AI) was supplied by the Bowman lab[189] (Table 6.3 and Supplementary Materials). Cloning of human RAD21 cDNA into the recipient backbone was done using InFusion technology by Philip Hublitz

(WIMM, Genome Engineering Services). To linearise the plasmid a restriction enzyme digest was performed: 50 µg of plasmid DNA was incubated with 3 µL of AseI (New England Biolabs: R0526S), 10 µL NEBuffer r3.1 (New England Biolabs: B6003S) and diluted to 100 µL with nuclease-free water (not DEPC-treated) (Ambion: AM9932). Reaction mix was incubated at 37°C for 1 hour in a thermomixer. 1 µL of digested product was run on an 1.2% agarose gel (Invitrogen: 16500-500) in TBE, to confirm complete digestion. Once complete digestion confirmed, linearised DNA was purified using Phenol:Chloroform:Isoamyl-Alcohol (IAA) (SigmaAldrich: 77617) extraction with Phase Lock Gel Light Tubes (VWR: 733-2477), purified by ethanol precipitation, and resuspended in 200 µL nuclease-free water. DNA concentration was quantified by Nanodrop and stored at -20°C.

**Table 6.1 Guide sequences for CRISPR-cas9 editing.**

| Guide | Guide Sequence (5'-3') | Target location (mm39) | Notes |
|---|---|---|---|
| Endogenous RAD21 (sg2) | CCTCAGATAATATG GAACCGTGG | chr15: 51827450-51827472 | Used for tagging endogenous RAD21 with TST and dTAG. |
| Artificial locus (AL2R2) | GGTTGGTGCTTC CGGAGACA | chrX:10728970-10728988 | Used for targeted engineering to neutral region on chrX. Used for R2 insertion and RAPID-release. |

**Table 6.2 Donor sequences for CRISPR-cas9 editing.**

| | Donor Sequence (including homology arms in bold) (5'-3') |
|---|---|
| RAD21-FKBP$^{F36V}$-3xFLAG | **GGGTTGGAAGGTTATCAGGGGCCGGTTTGATTTTGGTTTTGTTTTCACTTTAAAATCTGCTGAGTGTTTG**<br>**TTTTTGCTAACTCACATCTCTGTTGTGGCGCCTTCACTTACTTCTTGAACTCTGTATGCCTTAAGATAATA**<br>**TGCTTTAGATAATAAGGCCTTTACCTTAGTGACTAGCATGGGAACCACTTGCGTAACTACACGAATGTCC**<br>**AGTCCCAGCTCCAGCCAAACGGAGCTCACTTTGACCAGTGCCAA**AATTGCATCTTCTGGTTACTACTTTG<br>TGCGTGAGTTACTTGAAATCATCTGCTTTGTTTTGTTCTATTTCAGCGAGCTCTTGCTAAAACTGGAGCA<br>GAGTCTATCAGTTTGCTTGAGCTGTGTCGAAACACAAACCGAAAGCAGGCAGCAGCAAAGTTCTACAGCT<br>TTTTGGTTCTTAAGAAGCAGCAAGCCATCGAGCTCACACAGGAAGAGCCGTACAGTGACATCATTGCAAC<br>GCCCGGCCCACGCTTTCACATCATCACCGGTGGAGTGCAGGTGGAAACCATCTCCCCAGGAGACGGGCG<br>CACCTTCCCCAAGCGCGGCCAGACCTGCGTGGTGCACTACACCGGGATGCTTGAAGATGGAAAGAAAGT<br>TGATTCCTCCCGCGACAGAAACAAGCCCTTTAAGTTTATGCTAGGCAAGCAGGAGGTGATCCGAGGCTG<br>GGAAGAAGGGGTTGCCCAGATGAGTGTGGGTCAGAGAGCCAAACTGACTATATCTCCAGATTATGCCTA<br>TGGTGCCACTGGGCACCCAGGCATCATCCCACCACATGCCACTCTCGTCTTCGATGTGGAGCTTCTAAAA<br>CTGGAACAATTGGGCGGATCCGGGAGCGCCTGGAGCCACCCCCAGTTCGAGAAGGGCGGCGGCAGCGG<br>CGGCGGCAGCGGCGGCAGCGCCTGGAGCCACCCCCAGTTCGAGAAGGTCGACAGATCTTGATTGATTAA<br>AGATCTTCTAGA**GGAGCTAGATGTGTTCGAGCTAGTGATAACTCACTAGTACATACAAATTGCCCCCGTG**<br>**TGCAGGGCACCAAAACCCTTTAAGAAAGTTTTTAGATTTCTGTTTGTACAAAAATCTTTGCCTTTTCTTTC**<br>**TTCTTTTTCCCCCCAGTGTTTCTAATTTTGTCAACCATATTTTTAAGGGAAACTGCTTATTTGGGTTGGGT**<br>**TTGTATTCCTGGAGAAAACAGTAGCCCAAGAACCCAGAAGACTTTTAACAGTTCA** |
| RAD21$^{TST}$ | **CCCGGGTTGGAAGGTTATCAGGGGCCGGTTTGATTTTGGTTTTGTTTTCACTTTAAAATCTGCTGAGTGT**<br>**TTGTTTTTGCTAACTCACATCTCTGTTGTGGCGCCTTCACTTACTTCTTGAACTCTGTATGCCTTAAGATA**<br>**ATATGCTTTAGATAATAAGGCCTTTACCTTAGTGACTAGCATGGGAACCACTTGCGTAACTACACGAATG**<br>**TCCAGTCCCAGCTCCAGCCAAACGGAGCTCACTTTGACCAGTGCCAAAATTGCATCTTCTGGTTACTACT**<br>**TTGTGCGTGAGTTACTTGAAATCATCTGCTTTGTTTTGTTCTATTTCAGCGAGCTCTTGCTAAAACTGGA**<br>**GCAGAGTCTATCAGTTTGCTTGAGCTGTGTCGAAACACAAACCGAAAGCAGGCAGCAGCAAAGTTCTAC**<br>**AGCTTTTTGGTTCTTAAGAAGCAGCAAGCCATCGAGCTCACACAGGAAGAGCCGTACAGTGACATCATT**<br>**GCAACGCCCGGCCCACGCTTTCACATCATC**AGCGCCTGGAGCCACCCCCAGTTCGAGAAGGGCGGCGGC<br>AGCGGCGGCGGCAGCGGCGGCAGCGCCTGGAGCCACCCCCAGTTCGAGAAG**TGAGATATCGGAGCTAG**<br>**ATGTGTTCGAGCTAGTGATAACTCACTAGTACATACAAATTGCCCCCGTGTGCAGGGCACCAAAACCCTT**<br>**TAAGAAAGTTTTTAGATTTCTGTTTGTACAAAAATCTTTGCCTTTTCTTTCTTCTTTTTCCCCCCAGTGTT**<br>**TCTAATTTTGTCAACCATATTTTTAAGGGAAACTGCTTATTTGGGTTGGGTTTGTATTCCTGGAGAAAACA**<br>**GTAGCCCAAGAACCCAGAAGACTTTTAACAGTTCAGAACAGATGTGTGCAATATTGGTGCATGTAAGAAT**<br>**ATGGAGTAACAGTCAAAAGGCACCATTTTTAATGTTAGTTTTCCATTACTATGTTGAAAGGAAAACCTGC**<br>**CTAGGAAAATGCCTGACACTTTAAGAACTGTGGTTTGAGTCCCTTGACAGGAAGAGAAAAATGTCTTCCC**<br>**ATCAGTGAAACCAACGGTCTGGTTAACCACTGTAGTAGGGATAGTGTGTGAAGCATCCCGGG** |

| | |
|---|---|
| RAPID-release | **CCCACAATGGCATGGCATTGCTGAATGTATCCCAGTTTTTTGGCTTTTGATGCCTATATTGTCAGGGTCA**<br>**TATATCCCCAACCACTACTGAGAAAAACTTCCTTGGGTCAGTGTTGGATAGCAGCAGCTCCCTTATGCTT**<br>**ACTTGGAGTGGTTCTAAGATTTTGTGTTTTACATTAGGTATTTTATCTATTTGCAGTTGGTTTGGGATGTT**<br>**GTGTGAGACAAGGGTCCAGTTTCTTTCTTCTGCATGTACACATCTAGTCTTCAAACAGAGTTCAATGAAG**<br>**AAATTGTCAATTTTCTATTGTATATACTTTTGCCATATTCATGCTGGAAATGGCTTTCTTTGCTTTTCCCA**<br>**AGGTAATCTGCATTTAATGTGGATGAGAGGATGCTGGCCATAACCCCCCCCCCCTTCCATCCCTGCACTG**<br>**GAAAATATTCTGGGCATCTGACAGTCTATAAGACCATGCTCTTCCCAGGTAGTTATCTTCTTATGACTGA**<br>**GCACAGTTGGTTG**ATAACTTCGTATAGCATACATTATACGAAGTTATATTAATGGGCAGAGCGCACATCG<br>CCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGG<br>GGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT<br>AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGCTAGCGCTAC<br>CGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG<br>GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAA<br>GCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCT<br>GACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC<br>CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC<br>CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA<br>CGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA<br>GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC<br>CGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAG<br>CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC<br>CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATACTCAGATCTAGTCGA<br>CATGTTCTACGCACATTTTGTTCTCAGTAAAAGAGGGCCTCTGGCCAAAATTTGGCTAGCGGCCCATTGG<br>GATAAGAAGCTAACCAAAGCCCATGTGTTCGAGTGTAATTTAGAGAGCAGCGTGGAGAGTATCATCTCAC<br>CAAAGGTGAAAATGGCATTACGGACATCAGGACATCTCTTACTGGGAGTAGTTCGAATCTATCACAGGAA<br>AGCCAAATACCTTCTTGCAGACTGTAATGAAGCATTCATTAAGATAAAGATGGCTTTTCGGCCAGGTGTG<br>GTTGACCTGCCTGAGGAAAATCGGGAAGCAGCTTATAATGCCATTACTTTACCTGAAGAATTTCATGACT<br>TTGATCAGCCACTGCCTGACTTAGATGACATCGATGTGGCCCAGCAGTTCAGCTTGAATCAGAGTAGAGT<br>GGAAGAGATAACCATGAGAGAAGAAGTTGGGAACATCAGTATTTTACAAGAAAATGATTTTGGTGATTTT<br>GGAATGGATGATCGTGAGATAATGAGAGAAGGCAGTGCTTTTGAGGATGACGACATGTTAGTAAGCACT<br>ACTACTTCTAACCTCCTATTAGAGTCTGAACAGAGCACCAGCAATCTGAATGAGAAAATTAACCATTTAG<br>AATATGAAGATCAATATAAGGATGATAATTTTGGAGAAGGAAATGATGGTGGAATATTAGATGACAAACT<br>TATTAGTAATAATGATGGCGGTATCTTTGATGATCCCCCTGCCCTCTCTGAGGCAGGGGTGATGTTGCCA<br>GAGCAGCCTGCACATGACGATATGGATGAGGATGATAATGTATCAATGGGTGGGCCTGATAGTCCTGAT<br>TCAGTGGATCCCGTTGAACCAATGCCAACCATGACTGATCAAACAACACTTGTTCCAAATGAGGAAGAAG<br>CATTTGCATTGGAGCCTATTGATATAACTGTTAAAGAAACAAAAGCCAAGAGGAAGAGGAAGCTAATTGT<br>TGACAGTGTCAAAGAGTTGGATAGCAAGACAATTAGAGCCCAACTTAGTGATTATTCAGATATTGTTACT<br>ACTTTGGATCTGGCACCGCCCACCAAGAAATTGATGATGTGGAAAGAGACAGGAGGAGTAGAAAAACTG<br>TTTTCTTTACCTGCTCAGCCTTTGTGGAATAACAGACTACTGAAGCTCTTTACACGCTGTCTTACACCGC<br>TTGTACCAGAAGACCTTAGAAAAAGGAGGAAAGGAGGAGAGGCAGATAATTTGGATGAATTCCTCAAAG<br>AATTTGAAAATCCAGAGGTTCCTAGAGAGGACCAGCAACAGCAGCATCAGCAGCGTGATGTTATCGATG<br>AGCCCATTATTGAAGAGCCAAGCCGCCTCCAGGAGTCAGTGATGGAGGCCAGCAGAACAAACATAGATG<br>AGTCAGCTATGCCTCCACCACCACCTCAGGGAGTTAAGCGAAAAGCTGGACAAATTGACCCAGAGCCTG |

TGATGCCTCCTCAGCAGGTAGAGCAGATGGAAATACCACCTGTAGAGCTTCCCCCAGAAGAACCTCCAA
ATATCTGTCAGCTAATACCAGAGTTAGAACTTCTGCCAGAAAAAGAGAAGGAGAAAGAGAAGGAAAAAGA
AGATGATGAAGAGGAAGAGGATGAAGATGCATCAGGGGGCGATCAAGATCAGGAAGAAAGAAGATGGAA
CAAAAGGACTCAGCAGATGCTTCATGGTCTTCAGCGTGCTCTTGCTAAAACTGGAGCTGAATCTATCAGT
TTGCTTGAGTTATGTCGAAATACGAACAGAAAACAAGCTGCCGCAAAGTTCTACAGCTTCTTGGTTCTTA
AAAAGCAGCAAGCTATTGAGCTGACACAGGAAGAACCGTACAGTGACATCATCGCAACACCTGGACCAA
GGTTCCATATTATAGTCGACAGCGCCTGGAGCCACCCCCAGTTCGAGAAGGGCGGCGGCAGCGGCGGCG
GCAGCGGCGGCAGCGCCTGGAGCCACCCCCAGTTCGAGAAGAAAGATCTAGAAGGAGGTTCCGGCGGG
AGCGGAGGAGAAACTGTAAGATTCCAAAGCGGAGGGTCAGAAACCGTGCGCTTTCAGAGCGGTGGCGGG
GGAAGCGGCGCGCCAGGTGGCGCGGGAGGTGGCTCTGGAGGTAGCGTGGCGATTCTGTGGCATGAAAT
GTGGCATGAAGGCCTGGAAGAAGCGAGCCGCCTGTATTTTGGCGAACGCAACGTGAAAGGCATGTTTGA
AGTGCTGGAACCGCTGCATGCGATGATGGAACGCGGCCCGCAGACCCTGAAAGAAACCAGCTTTAATCA
GGCGTATGGCCGCGATCTGATGGAAGCGCAGGAATGGTGCCGCAAATATATGAAAAGCGGCAACGTGAA
GGACCTGACCCAGGCGTGGGATCTGTATTATCATGTGTTTCGCCGCATTGGCGGGATCCCCGTCCCACT
TCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGCAATGATCAAGGATCTAAGCTTCGCGGCGATGGC
GAACCGAGCGGCGTGCCGGTGGCGGTGGTGCTGCTGCCGGTGTTTGCGCTGACCCTGGTGGCGGTGTG
GGCGTTTGTGCGCTATCGCAAACAGCTGTAGACTAGTTCTGCAGGCTACGGTACAGCGTGCCCGGGAGT
ACTGCAGTTGAAGGTACTGCTGGCCCGGGGCCCCTCTCCCTCCCCCGGGATCGATCCGCCCCTCTCCCT
CCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATT
TTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATT
CCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTC
TGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCG
ACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCC
ACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAA
GGATGCCCAGAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTT
TAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGAT
GATAATATGGCCACAACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATG
CGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGG
CCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTG
GGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGA
CTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGT
GGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCAC
CAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGAT
GTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACT
ACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCA
ACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGG
GCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCCCAATTGCCGTCCCACT
TCAGTTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGCGAATTCAACGAGGAGTTCGTGCAGGTGGAAAC
CATTAGCCCGGGCGATGGCCGCACCTTTCCGAAACGCGGCCAGACCTGCGTGGTGCATTATACCGGCAT
GCTGGAAGATGGCAAAAAATTTGATAGCAGCCGCGATCGCAACAAACCGTTTAAATTTATGCTGGGCAAA
CAGGAAGTGATTCGCGGCTGGGAAGAAGGCGTGGCGCAGATGAGCGTGGGCCAACGCACGAAACTGAC
CATTAGCCCGGATTATGCGTATGGCGCGACCGGCCATCCGGGCATTATTCCGCCGCATGCGACCCTGGT
GTTTGATGTGGAACTGCTGAAACTGGAAGGCGGCTCCGGCGGCAGCGGTCCTCGAGGCGGGTCTGGAG

| | |
|---|---|
| | GTAGCGGCGGTTCAAAGGCGCTATTGAAGGGAGTGCGAGATTTTAATCCAATCTCTGCTTGCGTATGCCT<br>CCTTGAGAACTCCTCGGATGGGCATAGTGAGAGACTGTTTGGCATTGGTTTTGGCCCATATATCATTGCC<br>AACCAACATCTTTTTAGAAGGAACAATGGAGAGTTGACTATCAAAACCATGCATGGTGAGTTCAAAGTCA<br>AGAACTCAACACAGCTGCAGATGAAACCAGTTGAGGGCAGAGACATAATAGTTATCAAAATGGCTAAGG<br>ACTTCCCACCATTCCCTCAAAAACTAAAATTCAGACAGCCTACCATCAAAGATAGAGTGTGCATGGTATC<br>CACAAATTTTCAGCAGAAAAGTGTCTCTAGTCTAGTGTCTGAGTCATCACACATTGTGCATAAAGAGGAC<br>ACTTCATTCTGGCAACACTGGATAACAACAAAGGATGGACAATGTGGAAGTCCGCTGGTTTCAATCATTG<br>ATGGAAATATTTTGGGAATACACAGCCTGACGCATACGACCAATGGTAGCAATTACTTCGTGGAATTTCC<br>TGAGAAGTTCGTAGCTACATATCTTGATGCCGCTGATGGTTGGTGCAAGAATTGGAAGTTCAATGCTGAT<br>AAGATCAGTTGGGGTTCCTTTATTCTGTGGGAAGATGCGCCGGAAGATTTTATGAGCGGCCTGGTGCCA<br>CGCGGTGTGGGTCGCGAATATGTGCGCTTTGCGCCGGGCAGCTAGGATATCAACTTGTTTATTGCAGCT<br>TATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAG<br>TTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTTAGCTGTGCCTTCTAGTTGCCAGCCATCTGT<br>TGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT<br>GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC<br>AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGATAACTTCGTA<br>TAGTATACATTATACGAAGTTA**ACTTATGACTTTTCTGATGGGCAGTTTTTCCCCTGGTCTCCACTGGCC<br>TGAAGAGACGTTTTCAAAGTGATGCTTTTCAGCTGAGGTGTATTCCATCCTTCCCTCTGCTTCCACAGGT<br>CTTGGATGGACAACCTCTGTTTGTTCCATTTATCTTTTTGCAGGTTTCTAGCCCCATCTAGGTGTCTGCT<br>TCCAGGAGGACAAGAACTGGCATACATAACCATAAAGTCTCAACAACGCAGTAGGTGTTTTCCAGATATC<br>CTCCCCTGAGTTCTTTCCAGCTGAATGAGGGAGCCACTGAGGTCTGTGCCTTTGCTTACCTCCCCAACTA<br>CCATGCAGCAGCCTTGACAACAGCACTTCTGTTCTGCTATGCAAATTTGTTTTCTCTCTCTACTGAATTC<br>CCTGAAGGTAAGAACAGGAATTATCTTATCTCTAAACCAGGTATTGGGCAATTCAGGTGGATCATGCATG<br>ATACTCTGGGTCACACATCCCCC** |

**Table 6.3 Plasmid used for RAPID-release**

| Plasmid | Components | Selection marker | Promoter |
|---|---|---|---|
| H3.1 dual vector | H3.1-EGFP$^{TVMVx2}$-FRB-NES-OMP25-IRES-mCherry-NES-FKBP-TVMV-thr-AI | NeoR/KanR | CMV |
| H3.1 single vector | H3.1-EGFP$^{TVMVx2}$-FRB-NES-OMP25 | NeoR/KanR | CMV |
| EV dual vector | EGFP$^{TVMVx2}$-FRB-NES-OMP25-IRES-mCherry-NES-FKBP-TVMV-thr-AI | NeoR/KanR | CMV |
| RAD21 dual vector | EGFP-RAD21$^{TST-TVMVx2}$-FRB-NES-OMP25-IRES-mCherry-NES-FKBP-TVMV-thr-AI | NeoR/KanR | CMV |

### 6.2.4 Transient transfection of HeLa and HEK293 cell lines with RAPID-release

HeLa/HEK293(T) cells were seeded at a density of $0.2x10^6$ in 35 mm treated µ -dishes (Ibidi: 81156), containing 2 mL HEK293 culture medium 24 hours prior to transfection. Transient transfections were performed using jetPRIME (Polyplus: 101000027) transfection reagent and jetPRIME buffer (Polyplus: 201000003). For a 35mm dish, 1 µg DNA and 2 µL of jetPRIME reagent was diluted in 200 µL of jetPRIME buffer following the manufacturers protocol and added to the cell media dropwise. Cells were incubated at 37°C for 6 hours, cells were rinsed once with PBS before changing to fresh culture media.

### 6.2.5 Generation of RAPID-release stable lines in HEK293(T) cells

A kill curve analysis was performed to find the optimal concentration of G418 (Merck: G8168) to select for positive clones. Cells were seeded in a 24 well plated at a density of $0.5x10^5$ cells/well. After 24 hours, G418 was added to wells in duplicate at 12 incremental concentrations between 0-1200 µg/mL. Media was changed in the wells daily, replacing with fresh pre-warmed media containing matched concentrations of G418. After 10 days, cell confluency was observed by eye and 600 µg/mL was selected as the minimum concentration that resulted in > 90 % cell death.

HEK293 transfections were performed following the PEI-pro (Polyplus: 115-0015) manufacturer supplied protocol. Cells were transfected whilst in suspension and plated into standard untreated 10 cm cell-culture dishes. Transfection mixture for $1x10^6$ cells contained 8 µL PEI-pro transfection reagent 500 µL Opti-MEM reduced serum media (Gibco: 31985062) and 8 µg linearised plasmid DNA. Cells were rinsed once with PBS and media was replaced with fresh pre-warmed media 6 hours post transfection. Cells were sorted into single cells into 96 well plates by fluorescent-activated cells sorting (FACS) after 10 days growing under G418 selection at 600 µg/mL, based on co-expression of EGFP

and mCherry. Individual clones were cultured in conditioned media for 10 days, expanded and frozen at -80°C for temporary storage prior to screening and in liquid nitrogen for long-term storage.

### 6.2.6 Generation of RAPID-release stable lines in mESC E14 cells

mESC transfections were performed following the LTX Lipofectamine LTX (ThermoFisher: A12621) manufacturer supplied protocol. Cells were passaged and plated at $1x10^6$ cells per 10 cm dish, transfections were carried out 24 hours after passaging. Transfection mixture for $1x10^6$ cells contained 6 µL Lipofectamine, 300 µL Opti-MEM reduced serum media (Gibco: 31985062), 2.5 µL PLUS reagent (ThermoFisher: A12621) and 2.5 µg plasmid DNA. Individual clones were isolated by fluorescent-activated cells sorting (FACS) after 10 days growing under G418 selection at 200 µg/mL, based on co-expression of EGFP and mCherry. Individual clones were cultured in conditioned media for 10 days, expanded and frozen at -80°C for temporary storage prior to screening and in liquid nitrogen for long-term storage.

## 6.3 Identification of positive clones

### 6.3.1 Agarose gel electrophoresis

Agarose (Invitrogen: 16500-500) was dissolved in 150 mL TBE to a 1.2% gel, 2 µL EtBr (Bio-Rad: 161-0433) was added. Gel percentage was modified based on expected fragment sizes. Loading dye was added to all samples (New England Biolabs: B7025S), and 8 µL of sample was loaded per well. 100 bp and 1 kb ladders were loaded for reference (New England Biolabs: N0551S/N0552S). Gels were run in TBE buffer at 100 mA for 1 hour and imaged on the Bio-Rad ChemiDoc system using UV transillumination.

### 6.3.2 PCR screening

PCR screening was used to screen all stably transfected cell lines described in this thesis in 96 well-plate format. Cells were cultured until approximately 50% confluent, media was removed and wells rinsed twice with 100 µL PBS. 50 µL of lysis buffer was added per well (buffer components listed in Table 6.4). Using a multichannel pipette, cells were resuspended and lysate transferred to a 96 well PCR micro plate (Axygen: PCR-96-FLT-C). Plates were incubated at 65°C for 3 hours, followed by 95°C for 10 minutes for proteinase K deactivation. For colony screening, genomic DNA from the crude lysate was used as a template for PCR. For the PCR reaction, per well: 1 µL crude lysate was combined with 0.25 µL of combined 10 µM forward and reverse primer, and 11.25 µL Platinum PCR supermix (Invitrogen: 12532-016). PCR was run on the BIORAD T100 thermocyler: 94°C 2 minutes, 32 cycles (94°C 30 seconds, 56°C 30 seconds, 68°C 60 seconds), 10 minutes at 72°C. PCR products were resolved on a 1.2% agarose gel, as described above. Screening primers listed in Tables 6.5 (RAPID-release) and Table 6.6 (dTAG). Diagrams of screening strategy shown in relevant chapters (dTAG: Figure 3.4, RAPID-release: Figure 4.19).

**Table 6.4 DNA extraction lysis buffer for colony screening**

| Stock solution | Catalogue no. | Concentration required | Volumes for 5ml (µl) |
|---|---|---|---|
| 1M Tris-HCl (pH8) | Invitrogen: 15568-025 | 50mM | 250 |
| Triton-X-100 | Sigma: T8787-50ML | 0.5% | 25 |
| Proteinase K (20mg/ml) | ThermoFisher: EO0491 | 1mg/ml | 250 |
| Sigma water | Sigma: W4502-1L | NA | 4475 |

**Table 6.5 Screening primers for RAPID-release EF1a HDR targeting to chrX**

| Primer | Sequence | Target (mm39) | GC (%) | MW | Tm (°C) | Species |
|---|---|---|---|---|---|---|
| chrX 5′ EXT FW | ACTTCTTGAACAGTGCTGCA | chrx:11239182-11239201 | 45 | 6092 | 61.3 | mouse |
| JH_023 EF1a RV | GGGACTGTGGGCGATGTG | EF1a core promoter | 66.6 | 5652 | 68.1 | synthetic |
| JH_023 3′INT FW1 | TGGGAAGACAATAGCAGGCA | 3′ internal (bGH poly(A) signal) | 50 | 6224 | 65.6 | synthetic |
| JH_023 3′INT FW2 | CGGTGGGCTCTATGGATAACT | 3′ internal (bGH poly(A) signal/loxP) | 52.3 | 6477 | 63.9 | synthetic |
| JH_023 3′INT FW3 | GTAGGTGTCATTCTATTCTGGGG | 3′ internal (bGH poly(A) signal) | 47.8 | 7117 | 62.4 | synthetic |
| JH_023 3′INT FW4 | GGGGAGGATTGGGAAGACA | 3′ internal (bGH poly(A) signal) | 57.8 | 6007 | 66.2 | synthetic |
| chrX 3′ EXT RV1 | GGGAAGGGAGAGAGACATGG | chrx:11240323-11240342 | 60 | 6345 | 64.9 | mouse |
| chrX 3′ EXT RV2 | TATGTGCAGAGATGGAGGGG | chrx:11240341-11240360 | 55 | 6302 | 65 | mouse |
| chrX 3′ EXT RV3 | AGAAACAGAGACATAGGAGGGG | chrx:11240277-11240298 | 50 | 6916 | 62.3 | mouse |
| chrX 3′ EXT RV4 | AGGGAGAGAGACATGGAGAGA | chrx:11240318-11240338 | 52.3 | 6642 | 62 | mouse |
| 5HOM 1 FW | GACCATGCTCTTCCCAGGTA | chrX:11239715-11239734 | 55 | 6053 | 64 | mouse |
| 5HOM 2 FW | TGTGGATGAGAGGATGCTGG | chrx:11239631-11239650 | 55 | 6293 | 66.4 | mouse |

**Table 6.6 Screening primers for dTAG HDR targeting to chrX.**

| Primer | Sequence | Target (mm39) | GC (%) | MW | Tm (°C) | Species |
|---|---|---|---|---|---|---|
| RAD21 ExR | GATACAGGTGAGGCTGGAGC | chr15: 51828029-51828048 | 60 | 6247 | 63.8 | mouse |
| RAD21 ExF | CAAATTCTTGCTCAGTGTGG | chr15: 51826912-51826931 | 45 | 6123 | 60.9 | mouse |
| JH_025 HOM F | TGTCGAAACACAAACCGAAA | chr15: 51827576-51827595 | 40 | 6112 | 64.1 | mouse |
| JH_025 FKBP F | ATGCCACTCTCGTCTTCGAT | Synthetic FKBP region | 50 | 6019 | 63.8 | synthetic |
| JH_025 HOM R | GGCAGGTTTTCCTTTCAACA | chr15: 51827091-51827110 | 45 | 6083 | 63.9 | mouse |
| JH_025 FLAG R | TGTAGTCACGCGTTCCTTTG | Binds FLAG-tag region | 50 | 6090 | 63.8 | synthetic |

### 6.3.3 Sanger sequencing

Clones that passed the PCR screening were further validated using Sanger sequencing. ExoSAP-IT (ThermoFisher: 78205) was used to enzymatically purify the PCR reaction before proceeding to the Sanger sequencing PCR reaction. To do so, 2 µL PCR product was combined with 2 µL ExoSAP-IT and 3 µL water, and incubated at 37°C for 1 hour, followed by 85°C for 15 minutes. For the Sanger sequencing PCR: 7 µL of purified PCR product was mixed with 0.5 µL 10 nM primer, 0.4 µL BigDye (LifeTechnologies: 4337455) and 2 µL 5x sequencing buffer (LifeTechnologies: 4337455) and incubated at 95°C 5 minutes for 1 cycle, followed by 25 cycles of 95°C, 15secs > 66°C, 10 secs > 60°C, 4 minutes. DNA was precipitated with 80% ethanol for a minimum of 15 minutes at room temperature, before centrifuging at 3000rpm, 4°C for 30 minutes to pellet the DNA. DNA pellets were washed in 70% ethanol and centrifuged for a further 10 minutes (3000 rpm, 4°C). Ethanol was then tipped off and allowed to air-dry at room temperature for 10 minutes to remove residual ethanol before adding 15 µL of HiDi Formamide (Life Technologies: 4311320), transferring to a sequencing plate (ThermoFisher: AB1400) and submitting to our internal sequencing facility (John Frankland/Tim Rostron, WIMM, Sequencing Facility).

## 6.4 Induction methods

To perform the induction of both RAPID-release and dTAG cell lines, culture medium was replaced with media containing: 200 nM rapamycin (Cambridge Bioscience: 301.3016) for RAPID-release, or 500 nM dTAG-13 (Bio-Techne: 6605/5) for the dTAG degron. Both regents were diluted from the stock concentration to their working concentrations in sterile water. Time of induction was recorded from the time of the media change. Cells were harvested for downstream experiments following the standard protocol after the relevant duration of treatment, duration of induction included up until the media was rinsed from the cells prior to trypsinisation.

## 6.5 Live cell imaging

### 6.5.1 Microscopy

All images were captured using a Zeiss spinning disk confocal microscope using 63x oil immersion objective with 488 nm and 561 nm lasers for detection of EGFP and mCherry respectively. For live cell imaging transfected cells were cultured in 35mm treated µ-dishes (Ibidi: 81156). For HEK293(T) and HeLa cell imaging, 1 hour prior to imaging culture media was replaced with pre-warmed DMEM without phenol red (Gibco: 31053028) in order to reduce background fluorescence. For RAPID-release experiments, 500 µL culture medium containing 1 µM Rapamycin (Cambridge Bioscience: 301.3016) was added directly to the well resulting in a final concentration of 200nM Rapamycin. An initial 5 images were acquired prior to rapamycin additions, subsequent images were acquired every 5 seconds for up to 30 minutes. All imaging data was stored on the OMERO imaging data management platform.

### 6.5.2 Image analysis

Fluorescence analysis was performed using Fiji (v2.3.0) and Python (v3.8.9). Using Fiji: (1) CZI movie files were split by channel, (2) a region of interest (ROI) was drawn within the cell nucleus, and the cytoplasm containing the mitochondria, (3) using the ROI manager `'multi-measure'` function, fluorescence measurements were taken at the ROIs in each frame of the movie, (4) then repeated for the red (mCherry) and green (EGFP) channels. Results saved as CSV files were then used to plot changes in fluorescence using Python.

## 6.6 Protein methods

Aliquots of 5-10 x$10^6$ cells were used for protein extraction. Protein concentrations of the lysates were quantified using the Qubit protein quantification kit (Invitrogen: Q33212).

### 6.6.1 Whole-cell protein extraction

For whole cell lysates, cultured cells were trypsinised, washed and pelleted, rinsed in cold PBS containing protease inhibitor complex (PIC)(Sigma: P8340), resuspended in 200µL RIPA buffer (Life Tech: 89900) containing PIC and incubated end-over-end for 30 minutes at 4°C. Cell lysate was centrifuged for 30 minutes at max speed at 4°C to clear the supernatant. The cleared supernatant was transferred to a clean tube, protein concentration was immediately quantified and then stored at -80°C unless proceeding directly with western blot.

### 6.6.2 Extraction of proteins from subcellular compartments

For subcellular fractionation a subcellular protein fractionation kit for cultured cells (ThermoFisher: 78840) was used according to manufacturer's instructions, assuming a confluent T75 of adherent cells equated to a packed cell volume of 50 µL. Slight

modifications were made to the protocol as follows: centrifugation speed for the cytoplasmic fraction was changed to 700g and 37°C incubation steps was carried out using a thermomixer with a 10 second vortex halfway through the incubation period.

### 6.6.3 Fluorescent western blot

9-12 µg of protein lysate per sample was mixed with loading dye (125mM Tris-HCl, 4% SDS, 50% glycerol, 0.2 g OrangeG (Sigma: O7252-25G)) and 1M DTT (ThermoFisher: R0861) in the ratio 45:45:10 respectively. Samples were boiled for 10mins before loading. 3 µL Chameleon Duo-pre stained protein ladder (Li-Cor: 928-60000) was loaded per ladder. NuPAGE Bis-Tris 4-12% 1.5 mm protein gels (ThermoFisher: NP0335BOX) were used with NuPAGE MOPS Running Buffer (ThermoFisher: NP0001) and ran at 170 V for 3 minutes to stack the samples and then a further 90 minutes at 120V for SDS-PAGE. Gel was then transferred onto a 0.45 µm nitrocellulose membrane (Amersham: 10600008) for 90 minutes at 30 V in transfer buffer (NuPAGE Transfer buffer (ThermoFisher: NP00061), MetOH and distilled water using the XCell II™ Blot Module (ThermoFisher: EI9051). Once transferred the nitrocellulose membrane was blocked with 5% milk in PBS for 1 hour under agitation and incubated overnight in primary antibody in 5% milk with 0.2% Tween-20 at 4°C. After incubation period, primary antibody was removed and membrane rinsed once with PBS, followed by 3 washes in PBS and 0.1% tween-20 and a final PBS wash. Rinsed membrane was then incubated in species matched IRDye® conjugated secondary antibody in 5% milk and 0.2% tween-20 for 1 hour, room temp in the dark. A full list of primary and secondary antibodies and dilutions used for western blot is provided in Table 6.7. Membrane was then rinsed as before and imaged using the iBright™ FL1000 Imaging System for fluorescent blots (ThermoFisher: A44241). Raw, unprocessed images of western blot membranes for all western blot experiments presented in this thesis are provided in Supplementary Materials.

**Table 6.7** **Antibodies used for ChIP and western blot experiments.**
C = ChIP, WB = Western blot.

| Target | C | WB | Species | Clonality | Supplier | Cat. number | Dilution for WB |
|---|---|---|---|---|---|---|---|
| **Alpha-tubulin** | | X | rabbit | mono | abcam | ab52866 | 1:10,000 |
| **Actin** | | X | rabbit | mono | abcam | ab179467 | 1:10,000 |
| **EGFP** | | X | rabbit | poly | abcam | ab290 | 1:2500 |
| **FLAG-M2** | X | X | mouse | mono | Merck | F1804 | 1:1000 |
| **FLAG** | X | | rabbit | poly | Bethyl | A190-102A | NA |
| **Golgin-97** | | X | mouse | mono | ThermoFisher | A21270 | 1:500 |
| **Histone 2B** | | X | rabbit | poly | Merck | 07-371 | 1:5000 |
| **IRDye® 680RD** | | X | mouse | poly | abcam | ab216776 | 1:5000 |
| **IRDye® 800CW** | | X | rabbit | poly | abcam | ab216773 | 1:10,000 |
| **mCherry** | | X | mouse | mono | abcam | ab125096 | 1:1000 |
| **RAD21** | X | X | rabbit | poly | abcam | ab992 | 1:1000 |
| **StrepTag II** | X | X | mouse | mono | Qiagen | 34850 | 1:1000 |
| **Tomm20 *** | | X | rabbit | mono | Abcam | ab186734 | 1:5000 |
| **TopoII-beta** | | X | mouse | mono | BD Bioscience | 611492 | 1:1000 |
| **VDAC *** | | X | Rabbit | mono | Cell signalling Technologies | 4661 | 1:1000 |
| **VDAC *** | | X | Rabbit | poly | Cell signalling Technologies | 4866 | 1:1000 |

* No signal detected using these antibodies. Further optimisation required.

### 6.6.4 Western blot analysis

Western blots were analysed using iBright™ Analysis Software (v5.0).

Equal sized frames were manually added over each band. Using the inbuilt analysis tool, the local background corrected ratio of each band compared to reference band was calculated and where applicable normalised to the loading control. Data was exported and plotted using Microsoft Excel.

# 6.7 RNA isolation and gene expression analysis

## 6.7.1 RNA isolation and cDNA synthesis

Cultured cells were frozen at -80°C in aliquots of $1x10^6$ per 100 µL TRI-reagent (Zymo: R2050-1-50). RNA was extracted using the Direct-zol RNA Miniprep Kit (Zymo: R2051) following the manufacturers protocol, excluding the DNase I treatment step. Nanodrop was used to quantify RNA concentration. For DNase I treatment: to 1-1.5 µg RNA, 0.1 µL DNase I, 2 µL DNase I buffer (Roche: 04716728001), 0.5 µL RNase inhibitor (Roche: 3335399001) and RNase-free water was added to make a total volume of 20 µL and incubated for 30 minutes at 37°C in a thermomixer. 1 µL 50 mM EDTA, pH 8 (ThermoFisher:AM9260G) was added to each sample and incubate at 75°C for 10 minutes to inactivate the DNase I. From 10 µL of treated RNA, cDNA was synthesised using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems: 4368814) following the manufacturers protocol.

## 6.7.2 RT-qPCR expression analysis

As a proxy for assessing copy number, gene expression by qPCR was performed. cDNA was diluted 1:16 in water, KAPA SYBR® Fast Universal with low ROX (Kapa Biosystems: KK4602) and the primers listed in Table 6.8 at 10 µM were used to perform qPCR following the manufacturers protocol. All sample were run in triplicate. For relative quantification of RNA abundance, the ΔΔCt method was used.

**Tables 6.8** **RT-qPCR primers for copy number analysis of RAPID-release clones.**
All primers were ordered from SigmaAldridge.

| Primer | Sequence | Target | GC (%) | MW | Tm (°C) |
|---|---|---|---|---|---|
| FKPB F | ATTTATGCTGGGCAAACAGG | Synthetic FKBP | 45 | 6181 | 63.6 |
| FKBP R | TAATCCGGGCTAATGGTCAG | Synthetic FKBP | 50 | 6157 | 63.6 |
| FRB F | CGATTCTGTGGCATGAAATG | Synthetic FRB | 45 | 6172 | 64.1 |
| FRB R | CAAACATGCCTTTCACGTTG | Synthetic FRB | 45 | 6052 | 64.1 |

# 6.8 Fixation condition optimisation

## 6.8.1 Testing different combinations of fixatives

1% and 2% formaldehyde (FA) was prepared from 37% FA stock solution (Sigma: 252549) in PBS and disuccinimidyl glutarate (DSG) (Sigma: 80424-50MG-F) was diluted to 2 mM in PBS. Aliquots of 5 x $10^6$ mESC $RAD21^{TST}$ were used per replicate, with three replicates per condition. For each fixation, 1 mL of fixative was added to cell pellets which were resuspended by vortexing. Cells were incubated end-over-end at room temperature for the duration of the fixation time (see Figure 4.9), after which cells were centrifuged at 1500 rpm for 3 minutes, fixative removed, pellets rinsed with PBS and centrifuged again at 1500 rpm for 3 minutes. For consecutive fixation (i.e., DSG followed by FA) the rinse and centrifugation steps were applied between fixes. Fixed cell pellets were frozen at -80°C prior to proceeding from the lysis step of the ChIP protocol. StrepTag antibody (see Table 6.7) was used for chromatin-immunoprecipitation.

## 6.8.2 Testing quenching reagents

Sterile stock solutions of 2.5 M glycine (Invitrogen: 15527013) and 2M Tris, pH 8 (Invitrogen: 15504020) were prepared. A 2% solution of formaldehyde (FA) was prepared from 37% FA stock solution (Sigma: 252549) in PBS. Aliquots of FA, Tris and glycine solutions were equilibrated to 37°C in a water bath prior to beginning the experiment, and

aliquots of PBS chilled on ice for post-quenching washes. Aliquots of 5 x $10^6$ mESC $RAD21^{TST}$ were used per replicate, with three replicates per condition and cells were pelleted in 15 mL falcon tubes. After rinsing with PBS each cell pellet was resuspended in 5 mL of 2 % FA and incubated at room temperature on a roller for 1 minute. After 1 minute, quencher was added, either: 7.5ml 2M Tris (for 1.5 M) or 255 µL 2.5M glycine (for 125mM), and incubated at room temperature on a roller (see Figure 4.10). Once the quencher was added, a 10-minute timer was started: after 1 minute, 5 minutes and 10 minutes the cells were centrifuged at 4°C, 1500 rpm for 3 minutes. Cell pellets were rinsed with ice-cold PBS and centrifuged again at 4°C, 1500 rpm for 3 minutes to remove fixative. Fixed cell pellets were frozen at -80°C prior to proceeding from the lysis step of the ChIP protocol. StrepTag antibody (see Table 6.7) was used for chromatin-immunoprecipitation.

## 6.9 Chromatin Immuno-precipitation (ChIP)

Chromatin immunoprecipitation was performed using 1x$10^7$ cells per replicate (unless specified otherwise) following a protocol supplied by the Milne Lab with either a 1% or 4% spike-in of WT HEK293 cells for mES E14 ChIP, or mES E14 cells for HEK293 ChIP for calibration. All buffer recipes are listed in Tables 6.9-6.14. For double cross-linking: cells were resuspended in 1 mL PBS containing 2 mM disuccinimidyl glutarate (DSG)(Sigma: 80424-50MG-F) for 30 minutes at room temperature (RT), pellets rinsed with PBS, before the second fixation with 1% formaldehyde (Sigma: 252549) for an additional 30 minutes at RT. Single cross-linking: cells were fixed with 2% formaldehyde (Sigma: 252549) in PBS for 10 minutes. Once fixed, the cells were washed in cold PBS and frozen at -80°C. Cells were thawed on ice and lysed in 120 µL SDS lysis buffer* with cOmplete Proteinase Inhibitor Cocktail (PIC)(Sigma: P8340) and syringe passaged using a 27G needle into Covaris micro-tubes (Covaris: 520045). Fixed chromatin samples were fragmented using the Covaris

sonicator (ME220) for 10 minutes total sonication time (175 power, 1000 cycles per burst, 25% duty factor) at 4°C, yielding an average fragment size between 200 and 400 bp. Sonicated chromatin samples were diluted to 1.1mL total volume using ChIP dilution buffer* containing PIC. Before the addition of antibody, 50 µL was taken from each sample as an input control. Protein A/G Dynabeads (InVitrogen: 100-01D/100-03D) were mixed in 1:1 ratio, rinsed and resuspended in ChIP dilution buffer* and 10 µL was used to pre-clear per 1 mL samples. Antibody was added at a concentration of 1 µg/µL per sample and immunoprecipitation was performed at 4°C overnight, see Table 6.7 for a full list of antibodies used. 20 µL Protein A/G Dynabeads were added per 1 mL sample and incubated for 5 hours at 4°C. The beads were then washed 4x in RIPA buffer* on a magnetic stand, followed by one wash with 1mL TE (Sigma Aldridge: 93283) with 50mM NaCl (ThermoFisher: 207790010). Bound chromatin fragments were eluted from the beads by the addition of 105µL of elution buffer and then incubated for 30 minutes on a thermomixer set at 65°C, 1000 rpm. Input samples were diluted 1:1 with elution buffer*, each 100 µL sample/input was then incubated at 65°C overnight to reverse the crosslinking. Samples were RNase (Roche: 12849100) treated at 37°C for 60 minutes, then proteinase K (BioLabs: P81075) treated at 45°C for 60 minutes. DNA fragments were purified using Zymo ChIP DNA Clean & Concentrator kit (Zymo Research: D5205) and eluted in water. Samples were stored at -20°C.

### 6.9.1 Modified chromatin immunoprecipitation protocol for FLAG antibodies

It is widely reported that SDS denatures the FLAG-M2 antibody, and Sodium-deoxycholate inhibits the FLAG-M2 antibody from binding to FLAG fusion proteins. I therefore modified the buffers (*) to remove these components so that the protocol could be used for ChIP using FLAG antibodies. The overall protocol is the same, buffers are swapped as shown in Table 6.9.

**Table 6.9 Buffer substitution for 'gentle' ChIP protocol**

| Step | Standard ChIP | Gentle ChIP |
|---|---|---|
| Lysis | SDS lysis buffer | Gentle lysis buffer |
| Washes | RIPA wash buffer | Gentle wash buffer |
| Dilution | ChIP dilution buffer | Gentle lysis buffer |
| Elution | SDS lysis buffer | SDS lysis buffer |

**Table 6.10 RIPA wash buffer**

Store at 4°C.

| Reagent | Supplier | Cat. number | Stock conc. (M) | Required conc. (mM) | Volume (mL) |
|---|---|---|---|---|---|
| Hepes-KOH (pH 7.6) | Sigma | H0527-100g | 1 | 50 | 25 |
| LiCl | Sigma | 310468-100g | 5 | 500 | 50 |
| EDTA | ThermoFisher | AM9260G | 0.5 | 1 | 1 |
| NP-40 (IGEPAL) | Sigma | I8896-50ML | 10% | 1% | 50 |
| Na-Deoxycholate | Sigma | D6750-25g | 10% | 0.70% | 35 |
| Sigma water | Sigma | W4502-1L | - | - | 339 |
| | | | TOTAL VOLUME | | 500 |

**Table 6.11 ChIP dilution Buffer.**

Once made, filter sterilise into 2x 50ml falcons, store at 4°C.

| Reagent | Supplier | Cat. number | Stock conc. (M) | Required conc. (mM) | volume (mL) |
|---|---|---|---|---|---|
| SDS | Severn Biotech | 20-4002-02 | 20% | 0.01% | 0.05 |
| Triton X-100 | Sigma | T8787-50ml | | 1.10% | 1.1 |
| EDTA | ThermoFisher | AM9260G | 0.5 | 1.2 | 0.24 |
| Tris-HCl (pH 8) | Life Tech. | 15568025 | 1 | 16.7 | 1.67 |
| NaCl | ThermoFisher | 207790010 | 5 | 167 | 3.34 |
| Sigma water | Sigma | W4502-1L | - | - | 93.6 |
| | | | TOTAL VOLUME | | 100 |

**Table 6.12** **SDS Lysis/Elution Buffer.**

Make fresh, can be stored for up to 2 days at 4°C.

| Reagent | Supplier | Cat. number | Stock conc. (M) | Required conc. (mM) | volume (mL) |
|---|---|---|---|---|---|
| SDS | Severn Biotech | 20-4002-02 | 20% | 1% | 0.5 |
| EDTA | ThermoFisher | AM9260G | 0.5 | 10 | 0.2 |
| Tris-HCl (pH 8) | Life Tech. | 15568025 | 1 | 50 | 0.5 |
| Sigma water | Sigma | W4502-1L | - | - | 8.8 |
| | | | TOTAL VOLUME | | 10 |

**Table 6.13** **Gentle lysis buffer.**

Store at 4°C.

| Reagent | Supplier | Cat. number | Stock conc. (M) | Required conc. (mM) | volume (mL) |
|---|---|---|---|---|---|
| Tris-HCl (pH 8) | Life Technologies | 15568025 | 1 | 50 | 2.5 |
| NaCl | ThermoFisher | 207790010 | 5 | 150 | 1.5 |
| NP-40 | Sigma | I8896-50ML | 20% | 0.5% | 1.25 |
| Glycerol | ThermoFisher | 17904 | 100% | 5% | 2.5 |
| Sigma water | Sigma | W4502-1L | - | - | 42.5 |
| | | | TOTAL VOLUME | | 50 |

**Table 6.14** **Gentle wash buffer.**

Once made, filter sterilise into 2x 50ml falcons, store at 4°C.

| Reagent | Supplier | Cat. number | Stock conc. (M) | Required conc. (mM) | volume (mL) |
|---|---|---|---|---|---|
| Tris-HCl (pH 8) | Life Technologies | 15568025 | 1 | 50 | 5 |
| NaCl | ThermoFisher | 207790010 | 5 | 150 | 3 |
| EDTA | ThermoFisher | AM9260G | 0.5 | 1 | 0.2 |
| Glycerol | ThermoFisher | 17904 | 100% | 0.5% | 0.5 |
| Sigma water | Sigma | W4502-1L | - | - | 91.3 |
| | | | TOTAL VOLUME | | 100 |

### 6.9.2 Chromatin immunoprecipitation using pre-conjugated magnetic FLAG beads

For the use of magnetic FLAG bead in ChIP, the 'gentle' ChIP protocol was used. Instead of adding antibody after the pre-clearing steps, the desired volume of Anti-FLAG M2 magnetic beads (Millipore: M8823) was added to each 1mL aliquot and incubated for 5 hours at 4°C. The usual wash steps were then performed, and the protocol proceeded as described.

### 6.9.2 Quality control steps

DNA concentration was quantified using the Qubit dsDNA HS assay (Invitrogen: Q32851) as per manufacturers protocol using 1 µL of sample. To assess sonication efficiency the D1000 Tapestation (Agilent: 50675583 and 50675582) assay was performed.

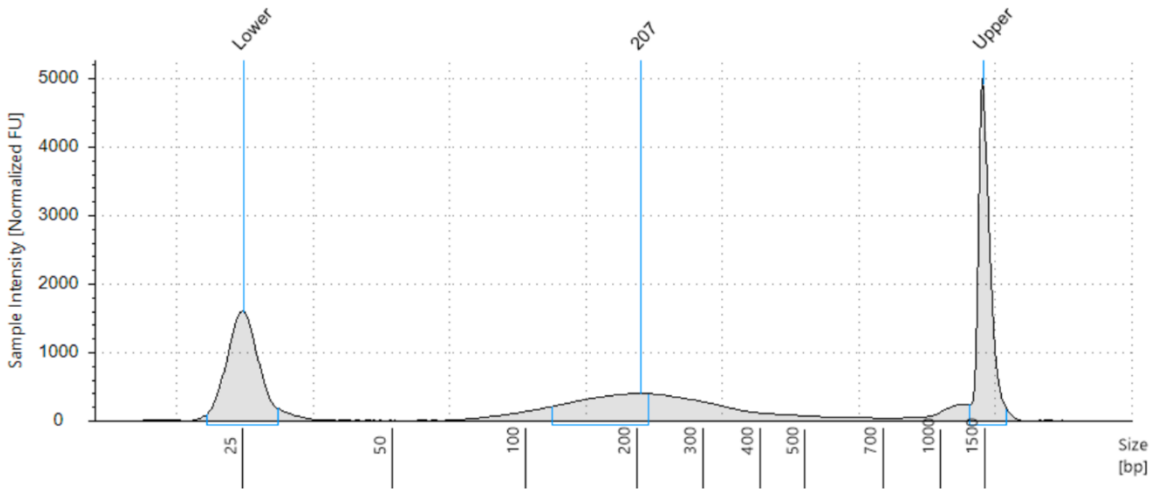



**Figure 6.2: Quality control of ChIP-seq sonication efficiency.**
Assessment of processed ChIP-seq libraries after sonication using the Tapestation (D1000). Optimal fragment size ~200bp.

### 6.9.3 qPCR

Primers were designed for two region types: (1) intergenic background control, (2) RAD21 binding sites. For each region type, three sets of primers were designed and tested. The best two sets of primers were then used for all further experiments (listed in Table 6.15). To perform the ChIP-qPCR KAPA SYBR FAST qPCR Kit (Universal) (Roche: 07959389001) with low ROX was used according to the manufacturer supplied protocol. Each sample was run in triplicate. Ct values were normalised by input values and fold enrichment calculated as ratio of sample versus intergenic control.

**Table 6.15 Primer sequences for RAD21 ChIP-qPCR**

| Primer | Direction | Sequence | Target (mm39) |
|---|---|---|---|
| Intergenic (Background Control) | 5′F | GTCTTATAGCCCAAGCCAACA | chr6:122652386-122652406 |
| | 3′R | CATGCTCAAGATGCACAGAAATAC | chr6: 122652457-122652480 |
| Nanog (Rad21 binding) | 5′F | TGATGTGCAGTTCCTTGTCTC | chr6:122646563-122646583 |
| | 3′R | ACTTCAGCTTCTTCCGCTTC | chr6:122646648-122646667 |

### 6.9.4 Next Generation Sequencing

An approximately equal mass of input and IP DNA was used for indexing (0.5 ng – 1 µg). NEBNext Ultra II DNA Library Prep Kit (New England Biolabs: NEB: 7645S/L) was used to prepare indexed sequencing libraries following the manufacturer supplied protocol. PCR amplification was performed for 7-11 cycles (depending on input DNA concentration) using NEBNext Mulitplex Oligos (New England Biolabs: E7335S/L and E7500S/L). Indexed sample concentration was quantified using the KAPA Library Quantification Complete Kit (Universal)(Roche: KK4824). Samples were pooled as a 2 nM or 4 nM library and sequenced with a high-Output v2 75 cycle kits on the Illumina NextSeq platform.

## 6.10 Tiled-C

Tiled-C data presented in this thesis was generated in collaboration with Marieke Oudelaar's laboratory (Max Planck Institute). 3C material and probe design was performed in the Hughes lab. Tiled-C experiments were performed by Marieke Oudelaar, as previously described[56].

## 6.11 ATAC-seq

ATAC-seq data presented in this thesis was generated by Caz Harrold[147] in mESC and CD71+ EB cells. 75,000 cells were used per replicate and processed as previously described[235]. ATAC-seq libraries were sequenced with a high-Output v2 75 cycle kits on the Illumina NextSeq platform.

## 6.12 Bioinformatic analysis

All genomic data presented in this thesis was aligned to the most recent genome builds (mm39 and hg38) unless otherwise stated. A brief description of each analysis pipeline is given, specific details and code are available on GitHub. Package versions used are listed in Table 6.16. Data generated in-house and publicly available has been analysed using these pipelines. Publicly available datasets were downloaded from the NCBI Gene Expression Omnibus (GEO) and converted into FASTQ file format using SRAtoolkit. For a full list of publicly available datasets used in this thesis, see Table 6.17.

**Table 6.16 Packages and versions used for bioinformatic analysis**

| Package | Version |
|---|---|
| Bedtools [236] | 2.29.2 |
| Bowtie2[237] | 2.4.2 |
| deepTools [181] | 3.5.1 |
| Hi-C Pro[238] | 3.1.0 |
| HiCPlotter [239] | 0.6.6 |
| HTSlib[240] | 1.10 |
| Matplotlib[241] | 3.4.3 |
| Python | 3.10.7 |
| Sambamba[242] | 0.8.0 |
| Samtools[243] | 1.10 |
| SRAToolkit[244] | 3.0.0 |

**Table 6.17 Publicly available genomic data used in analyses**

| ChIP target | Cell type /isogenic background | GEO accession | NCBI SRA | Antibody | Publication source/DOI |
|---|---|---|---|---|---|
| DNase I | mESC E14 (129/Ola) | GSE37074 | SRR578997.2 SRR578998.2 | - | ENCODE |
| CTCF | mESC (C57BL/6-129) | GSE30203 | SRR299029 SRR299030 SRR299031 | SantaCruz: sc15914 | 10.1038/ng.946[245] |
| | HEK293 | GSM749687 | SRR299276 SRR299295 | Millipore: unknown | ENCODE |
| H3K27me3 | mESC E14 | GSE30203 | SRR299034 SRR299035 SRR299036 | Upstate: 07-449 | 10.1038/ng.946[245] |
| H3K4me1 | mESC E14 | GSE120376 | SRR7895915 SRR7895916 | Diagenode: C15410194 | 10.1186/s13059-019-1860-7[246] |
| H3K4me3 | mESC E14 | GSE120376 | SRR7895917 SRR7895918 | Diagenode: C15410003 | 10.1186/s13059-019-1860-7[246] |
| H3K27Ac | mESC E14 | GSE120376 | SRR7895919 SRR7895920 | Diagenode: C15410196 | 10.1186/s13059-019-1860-7[246] |
| NIPBL | mESC (C57BL/6-129) | GSE22557 | SRR058989.3 SRR058990.3 | Bethyl: A301-779A | 10.1038/nature09380[75] |
| MED1 | mESC (C57BL/6-129) | GSE22557 | SRR058987.3 SRR058988.3 | Bethyl: A300-793A | 10.1038/nature09380[75] |
| RAD21 | HEK293 | DRA001232 | DRS012990 | Abcam: ab992 | 10.1093/nar/gku600[247] |
| SMC3 | mESC (C57BL/6-129) | GSE22557 | SRR058983 SRR058984 | Abcam: ab9263 | 10.1038/nature09380[75] |
| SMC1 | mESC (C57BL/6-129) | GSE22557 | SRR058981 SRR058982 | Bethyl: A300-055A | 10.1038/nature09380[75] |
| WAPL | mESC (129/Ola) | GSE135180 | SRR9870458.1 | Proteintech: 16370-1-AP | 10.1038/s41588-020-00744-4[248] |

### 6.12.1 ChIP-seq and ATAC-seq

Single- or paired-end reads from ChIP-seq were aligned to the target genome using Bowtie2. Sambamba was used to remove duplicates and index bamfiles. A coverage track of the aligned reads was generated using deepTools '`bamCoverage`', coverage was calculated for a single base pair window and replicates were normalised to Reads Per Kilobase per Million mapped reads (RPKM) using flags: `-bs 1 --normalizeUsing RPKM`. The output coverage files in bigWig format were visualised using the UCSC Genome Browser[249].

### 6.12.2 Calibrated ChIP-seq

The pipeline I wrote for analysis of calibrated ChIP-seq datasets was adapted from the scripts written by Nadya Fursova (University of Oxford, Klose lab)[250]. In brief: (1) Whole genome fasta files for the target and spike-in species were concatenated and used to build Bowtie 2 index files. (2) Paired-end reads were then aligned to the concatenated genome using Bowtie 2. (3) Reads were separated and counted based on alignment to target, or spike-in genome and a down-sampling factor was calculated based on the total number of spike-in reads per sample as follows:

$$\textit{Down-sampling factor} = a \times \frac{1}{total\ reads\ spike\text{-}in\ ChIP} \times \frac{total\ reads\ spike\text{-}in\ Input}{total\ reads\ target\ Input}$$

Where $a$ = coefficient for bulk normalisation of samples within a single experiment to enable the value of the largest down-sampling factor to equal 1[250]. (4) Using the calculated down-sampling factor, reads for each ChIP sample were randomly subsampled using Sambamba. (5) Input samples were used to correct the ratio of spike-in to target read counts across the biological replicates to account for minor variations in the mixing of spike-in cells. (6) Sub-sampled bamfiles were then processed according to the standard ChIP-seq analysis pipeline described in 6.12.1.

### 6.12.2 Custom genome builder

To correctly align reads to genetically engineered cell lines a custom genome is required. For this work I developed an updated version of the original script developed in the Hughes lab (unpublished) that is compatible with the latest genome builds. This script was designed to take an input file containing the genomic sequence in fasta format of the sequence being inserted into the genome, and a BED file containing the coordinates of the edited region and used `'SED'` to create a custom genome from this information using a cut-and-paste type approach. Samtools was then used to index the custom genome fasta file, and Bowtie2 used to create new indices.

### 6.12.3 Tiled-C

Tiled-C data was analysed using Hi-C Pro[238], with DpnII digestion of the genome using the built-in `'digest-genome'` utility. Capture target was specified using an input BED file containing region coordinates, min/max fragment sizes were specified in the config file as 20 and 100,000 respectively, and ICE normalisation selected. All other parameters applied were the default. HiCPlotter[239] was used to visualised the interaction matrices.

### 6.12.4 Peak calling using Lanceotron

Publicly accessible BigWig tracks were uploaded to the online version of Lanceotron[180] (https://lanceotron.molbiol.ox.ac.uk). Using the `'find and score peaks'` function, peaks in each file were called using default parameters using the built-in machine learning approach.

### 6.12.5 CTCF MOTIF analysis

The fasta sequence for the CTCF peaks called by Lanceotron was extracted using the Bedtools `'getfasta'` utility. The MEME frequency matrix for CTCF

(CTCF_MA0139.1.meme) was downloaded from JASPAR[251,252]. Using the FIMO (Find Individual Motif Occurrences)[253] tool from The MEME suite[254], occurrences of MA0139.1 were determined within the CTCF peak calls using default parameters. Peaks with identified CTCF motifs were annotated with the orientation.

### 6.12.6 RAD21 degradation analysis

Plots (Figure 3.17 and Figure 3.18B) were created using the Matplotlib library in Python. For the classification of *cis*-regulatory elements shown in Figure 3.18A, coverage scores were calculated for H3K27Ac, H3K4me1 and H3K4me3 publicly available datasets (see Table 6.16) and CTCF ChIP-seq generated in the lab. For histone ChIP-seq, an arbitrary cut-off of 10 was applied to the coverage scores; for CTCF ChIP-seq, an arbitrary cut-off of 20 was applied. Below threshold is indicated by cross, and above threshold indicated by a tick (Figure 3.18A).

### 6.12.7 Hi-C visualisation

Using the Juicebox Web App[182] (https://aidenlab.org/juicebox/), the Bonev *et al.* mESC Hi-C dataset[54] was loaded from the Juicebox archive and visualised. Publicly accessible BED files of persistent-RAD21, total RAD21 and CTCF peaks were uploaded in URL format.

### 6.12.8 Code availability

Quick-ChIP: https://github.com/EGeorgia/Quick-ChIP

Calibrated ChIP: https://github.com/EGeorgia/CalibratedChIP.

Custom genome: https://github.com/EGeorgia/custom-genome-builder

ATAC-seq: https://github.com/EGeorgia/ATAC-seq-analysis

# Supplementary Materials

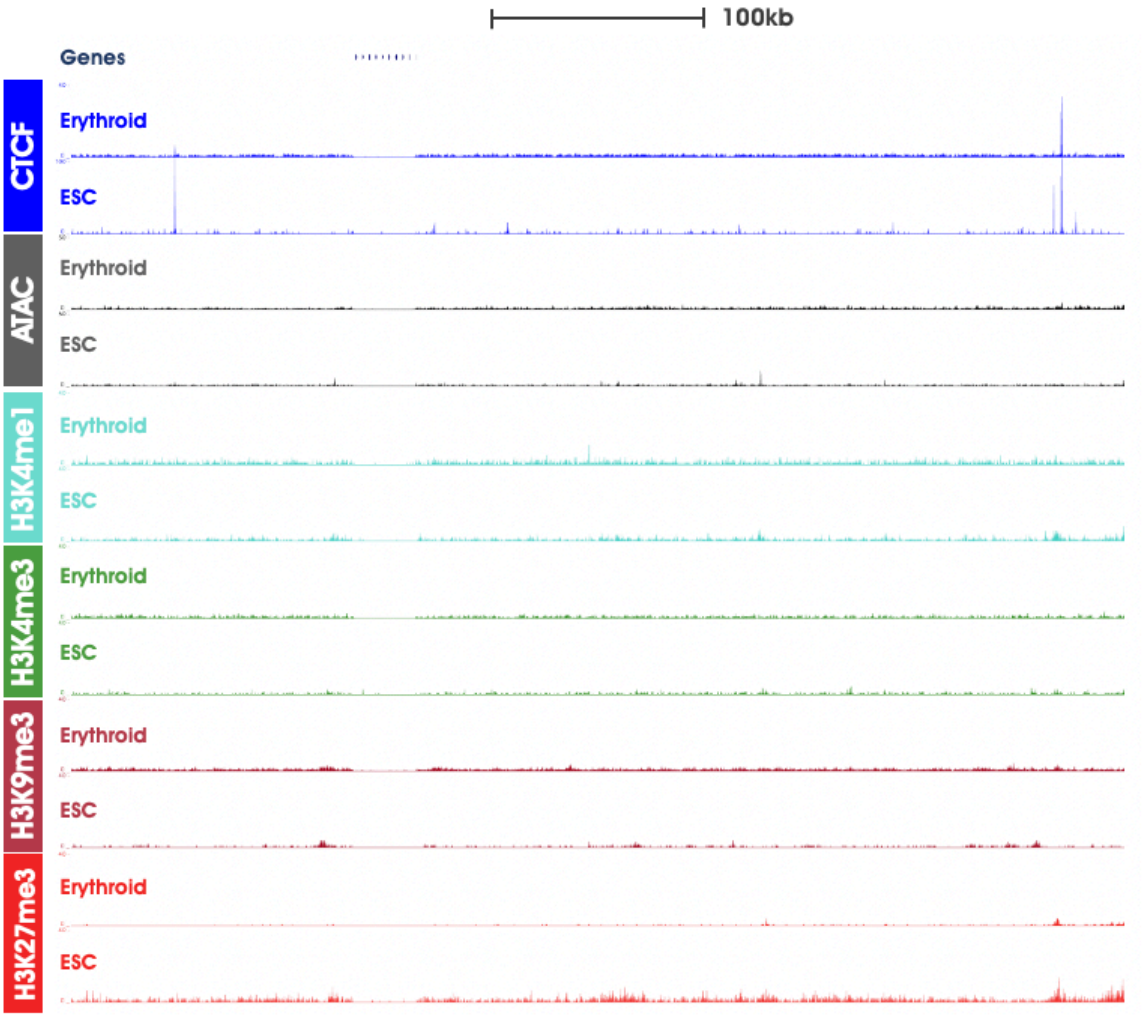



**Figure S1: Defining a neutral region of the genome.**

A neutral region is defined as a region devoid of histone marks (H3K4me1, H3K4me3, H3K9me3 and H3K27me3), open chromatin sites, annotated genes and without an underlying TAD structure. Shown is the neutral region selected for use in this chapter, chrX:10,935,857-11,046,248 (mm9). Figure adapted from Caz Harrold's thesis[147].

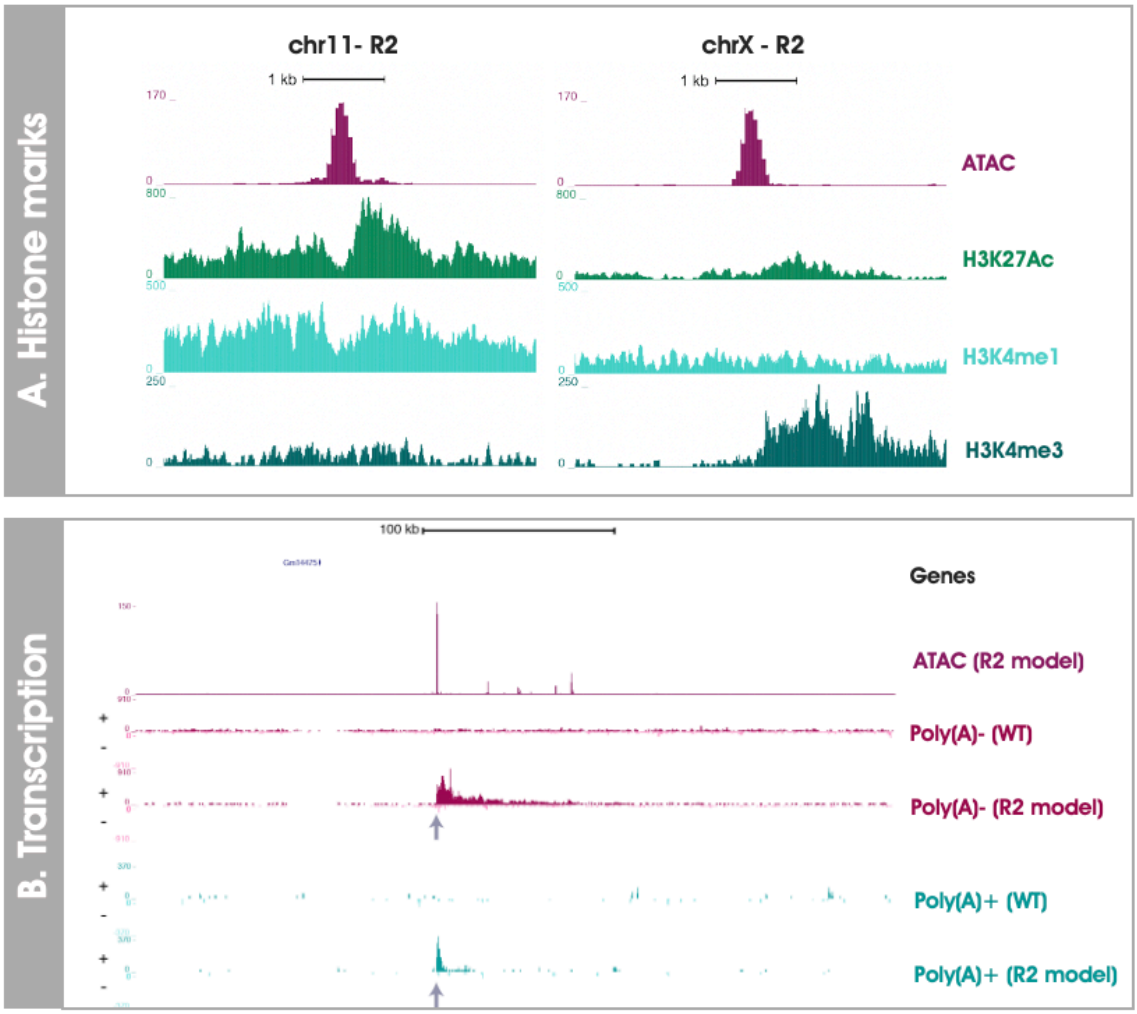



**Figure S2: Comparison of the wild-type (WT/chr11-R2) and R2 enhancer model (AL2R2).**
**a** | The open chromatin profiles of the R2 enhancer in the native (chr11-R2) versus artificial insertion in neutral chrX region (AL2-R2) are almost identical. The histone marks at the native R2 are indicative of an enhancer, however when inserted in neutral chrX locus the R2 associated chromatin displays promoter-like histone marks. **b** | Poly(A)- RNA-seq data shows that insertion of R2 results in transcription from the R2 insertion site which is not seen in the WT. Figures adapted from Caz Harrold's thesis [147].

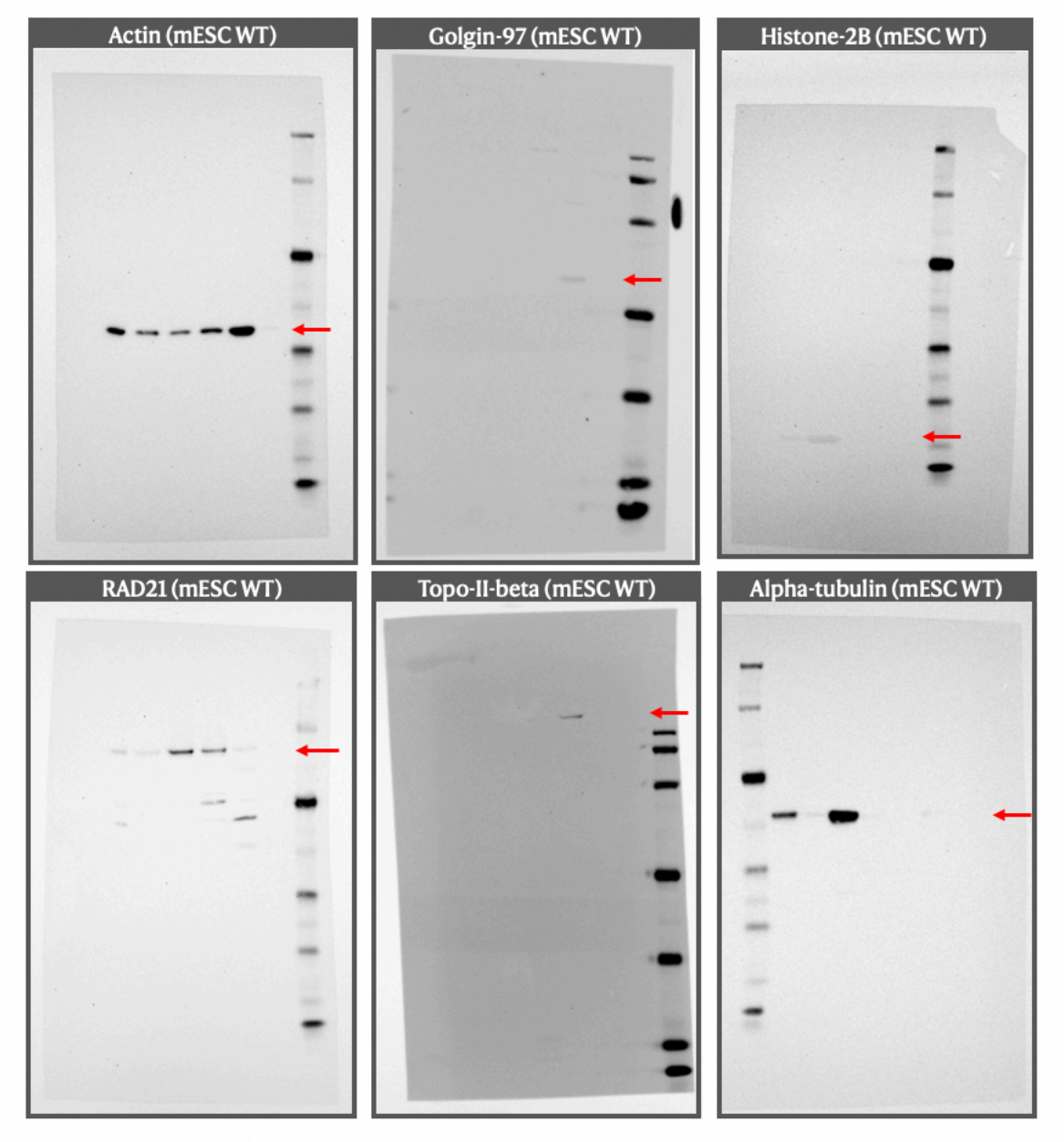



**Figure S3: Raw greyscale western blot membrane images for Figure 3.6.**
Subcellular fractionation of WT mESC cells.

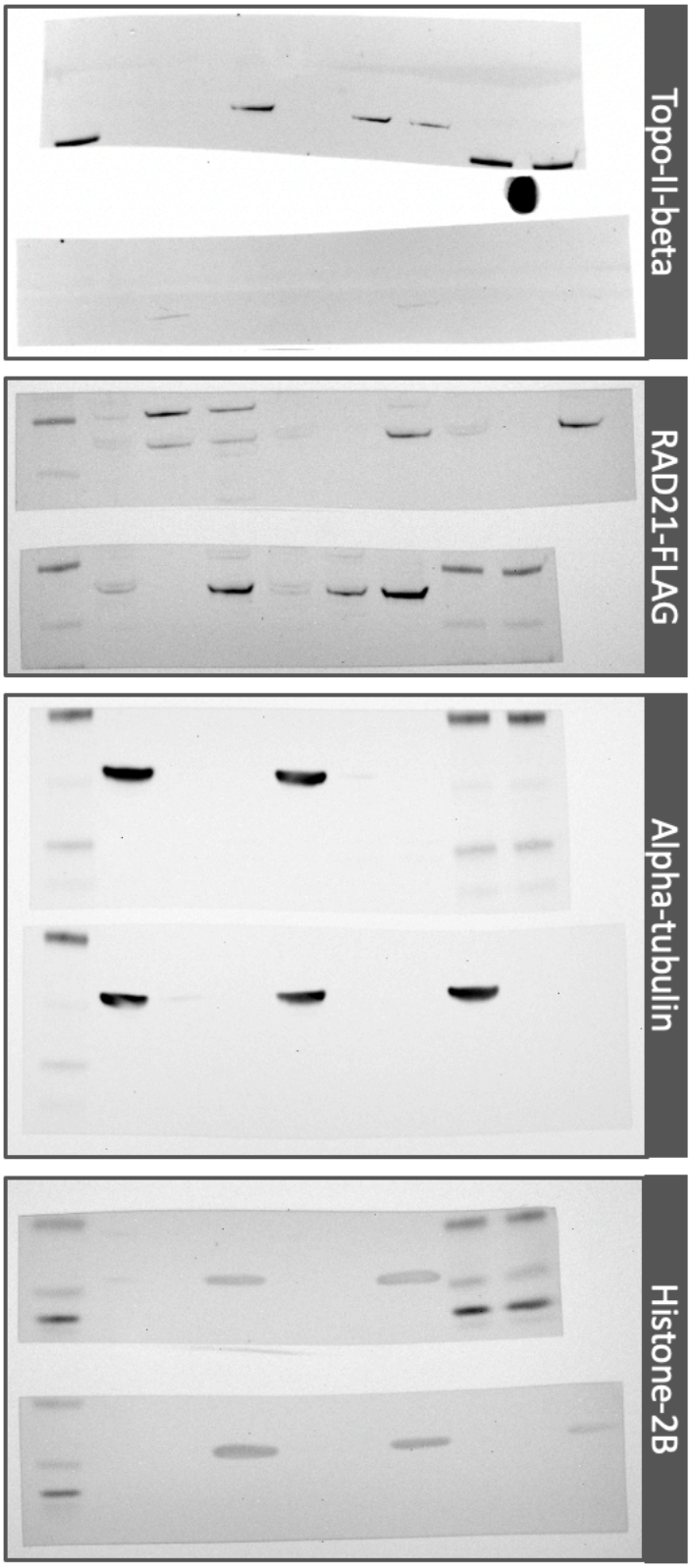



**Figure S4: Raw greyscale western blot membrane images for Figure 3.7.**
Fraction controls and RAD21-FLAG during 24 degradation time-course.

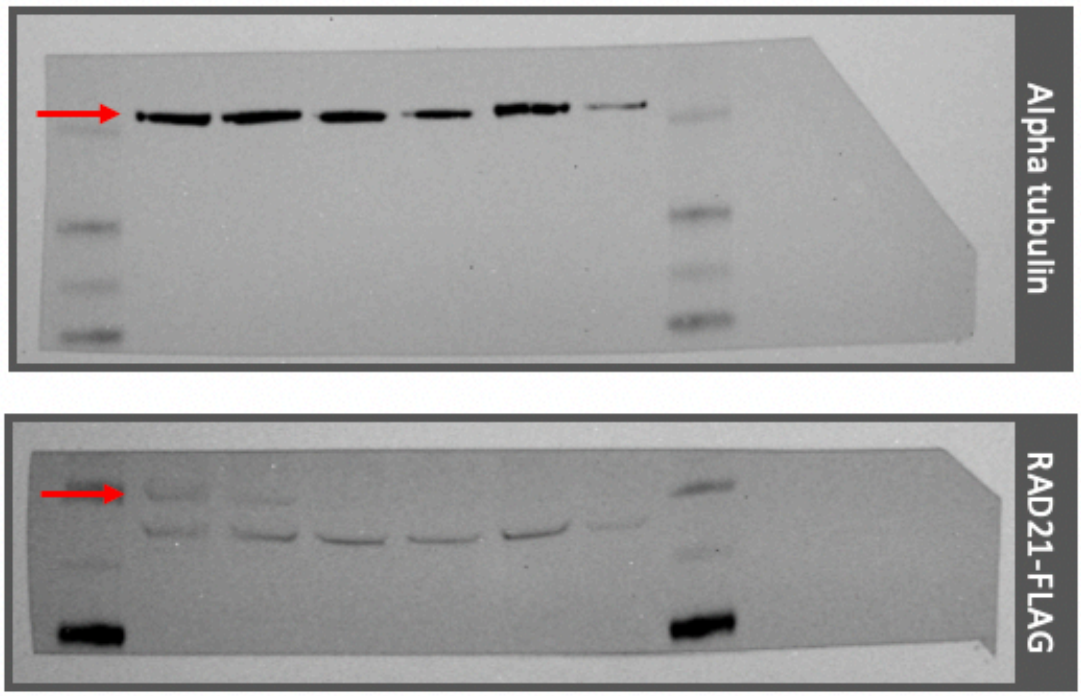



**Figure S5: Raw greyscale western blot membrane images for Figure 3.8.**
Titration of dTAG-13 concentration for a 4-hour degradation.

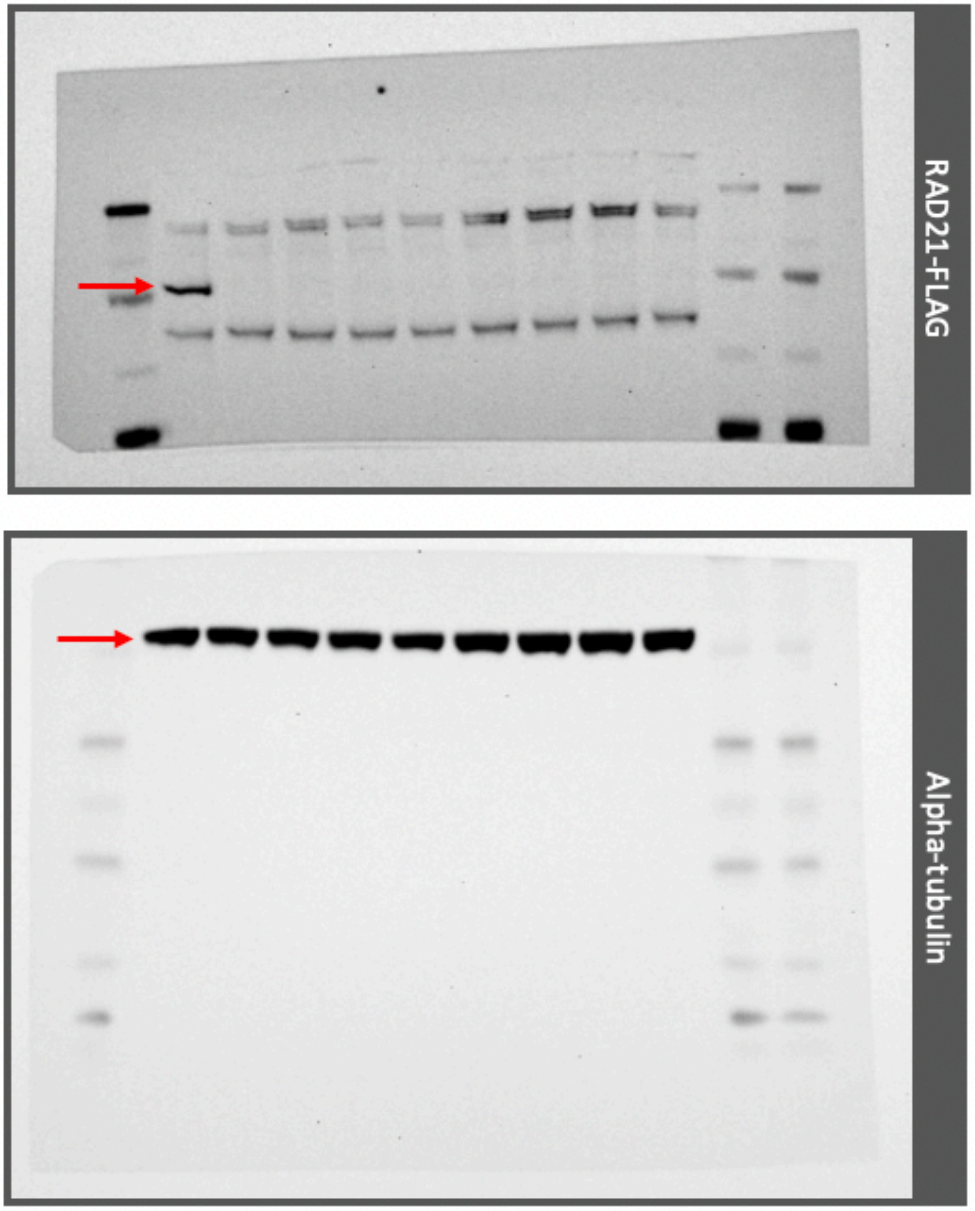



**Figure S6: Raw western blot membrane images for Figure 3.9.**

RAD21 does not rapidly reaccumulate following dTAG-13 treatment.

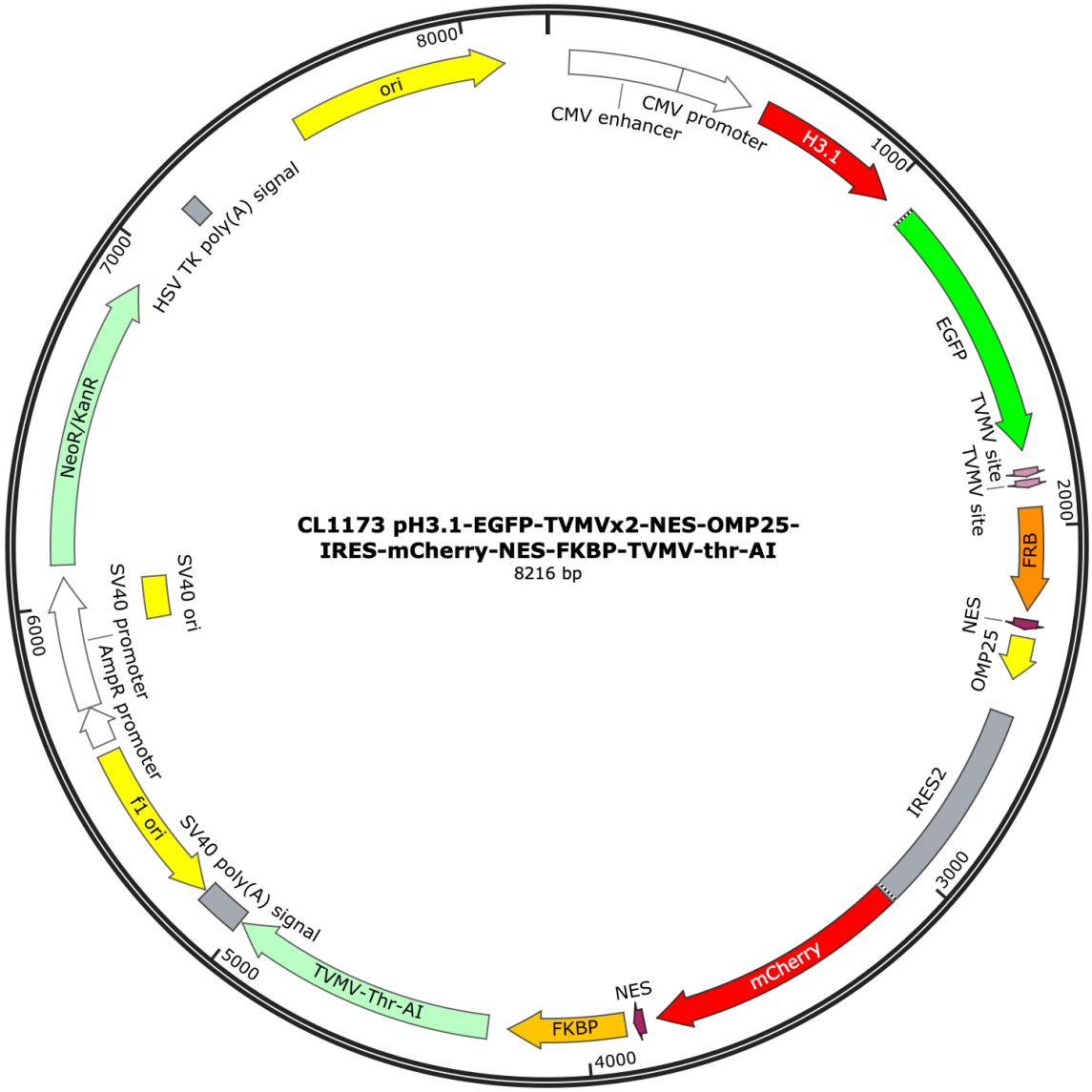



**Figure S7: Plasmid map of the original dual vector H3.1 RAPID-release vector from Andrew Bowman's lab (University of Warwick)[189]**

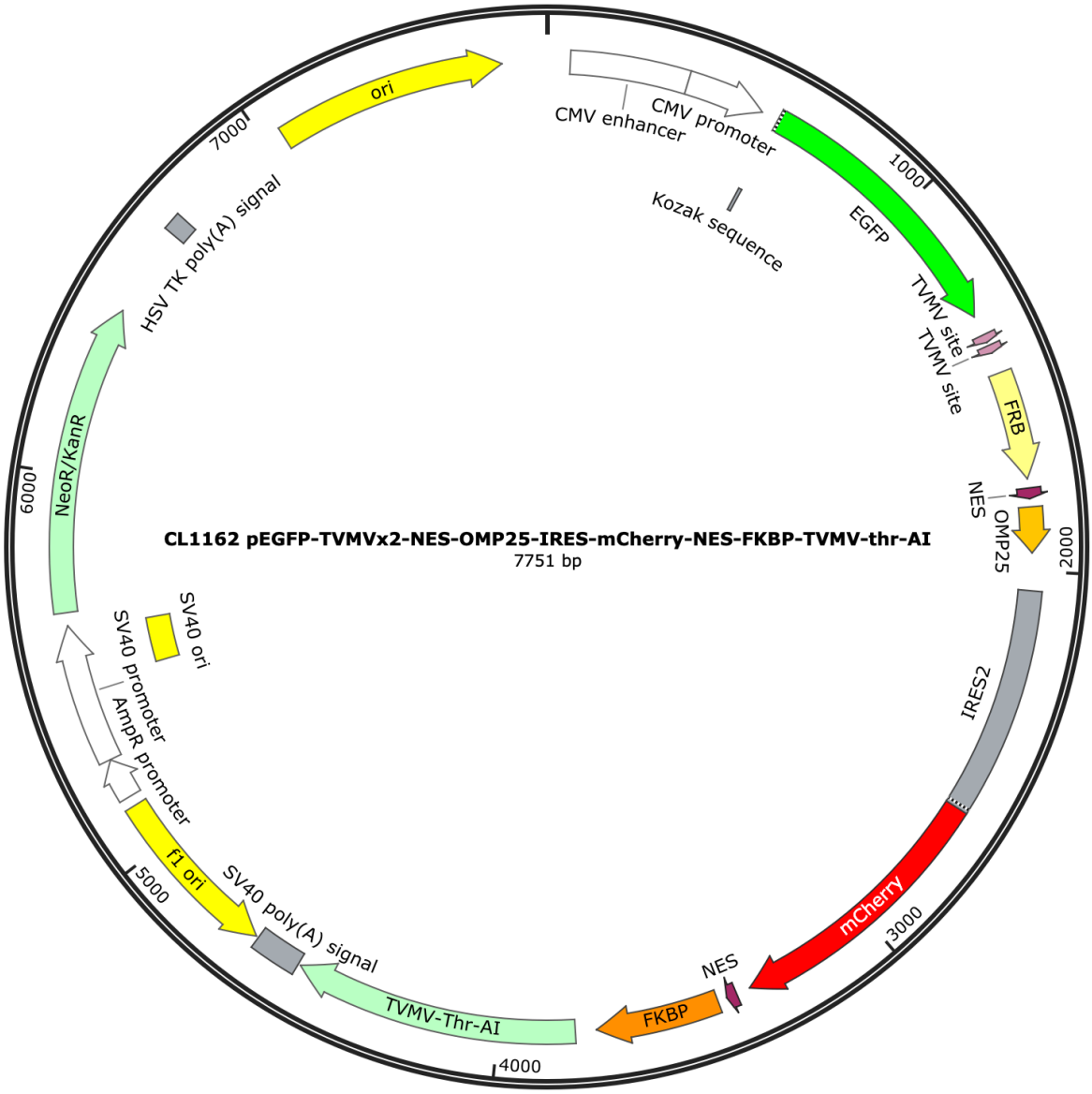



**Figure S8: Plasmid map of the single plasmid (dual vector) empty RAPID-release vector from Andrew Bowman's lab (University of Warwick)[189].**

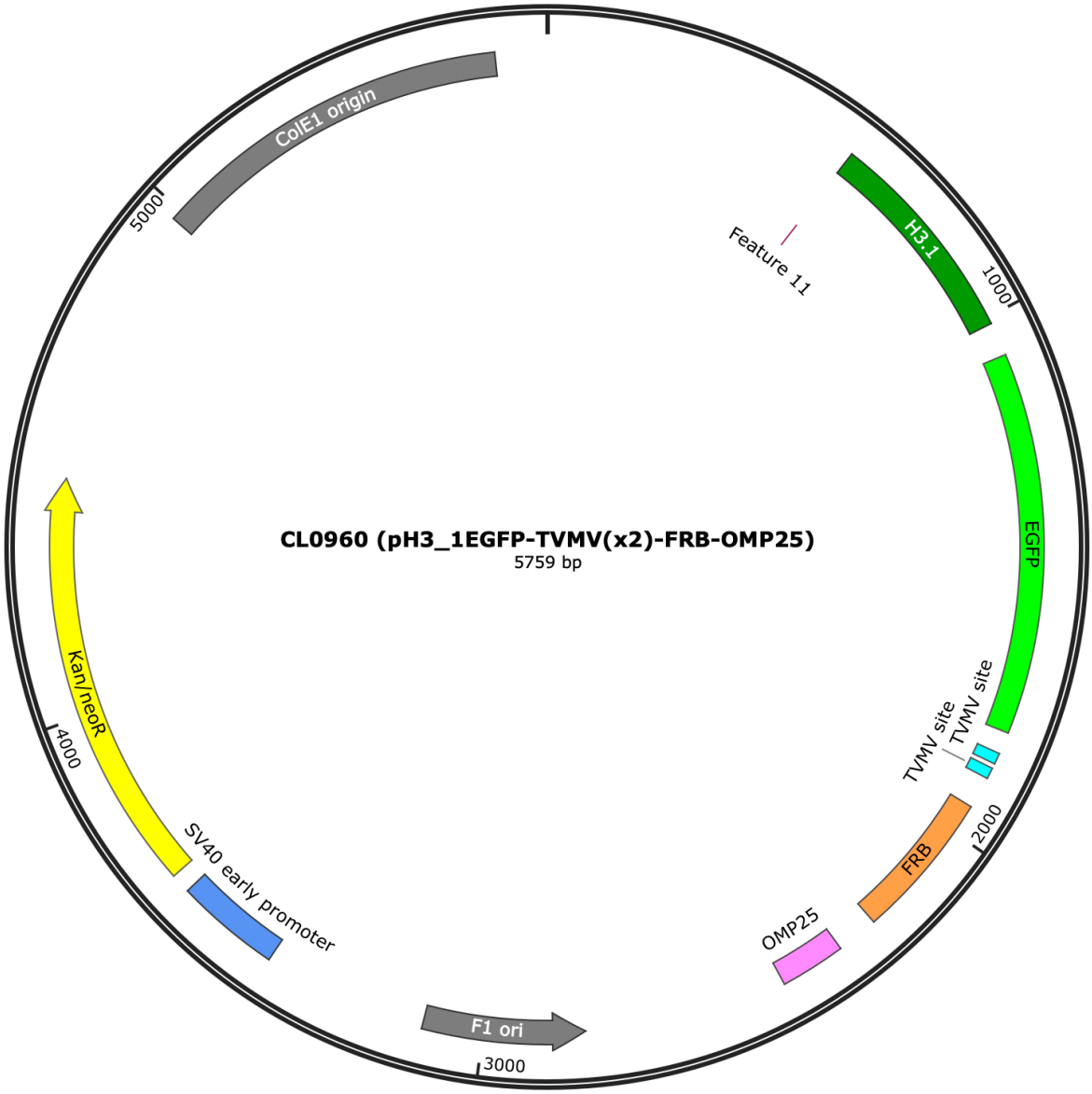



**Figure S9: Plasmid map of the single vector empty RAPID-release vector from Andrew Bowman's lab (University of Warwick)[189].**

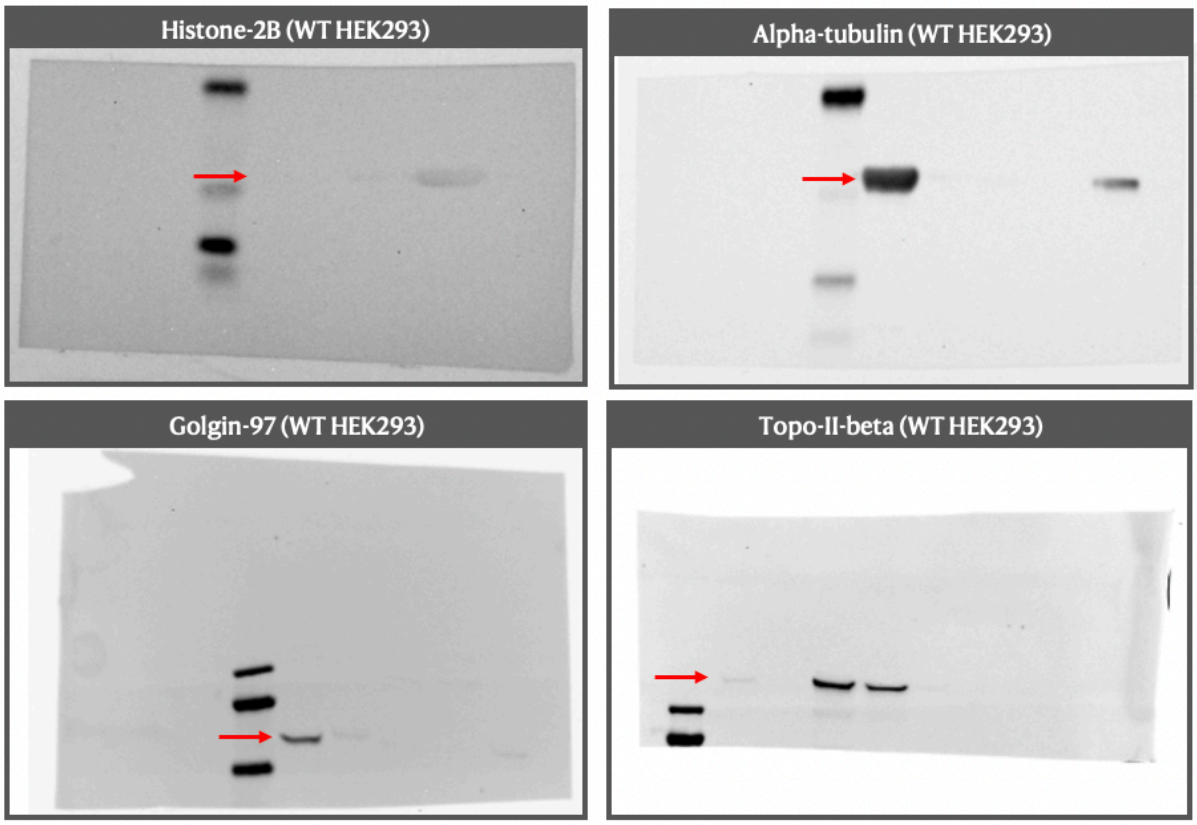



**Figure S10: Raw western blot membrane images for Figure 4.14A.**
Subcellular fractionation of WT HEK293 cells.

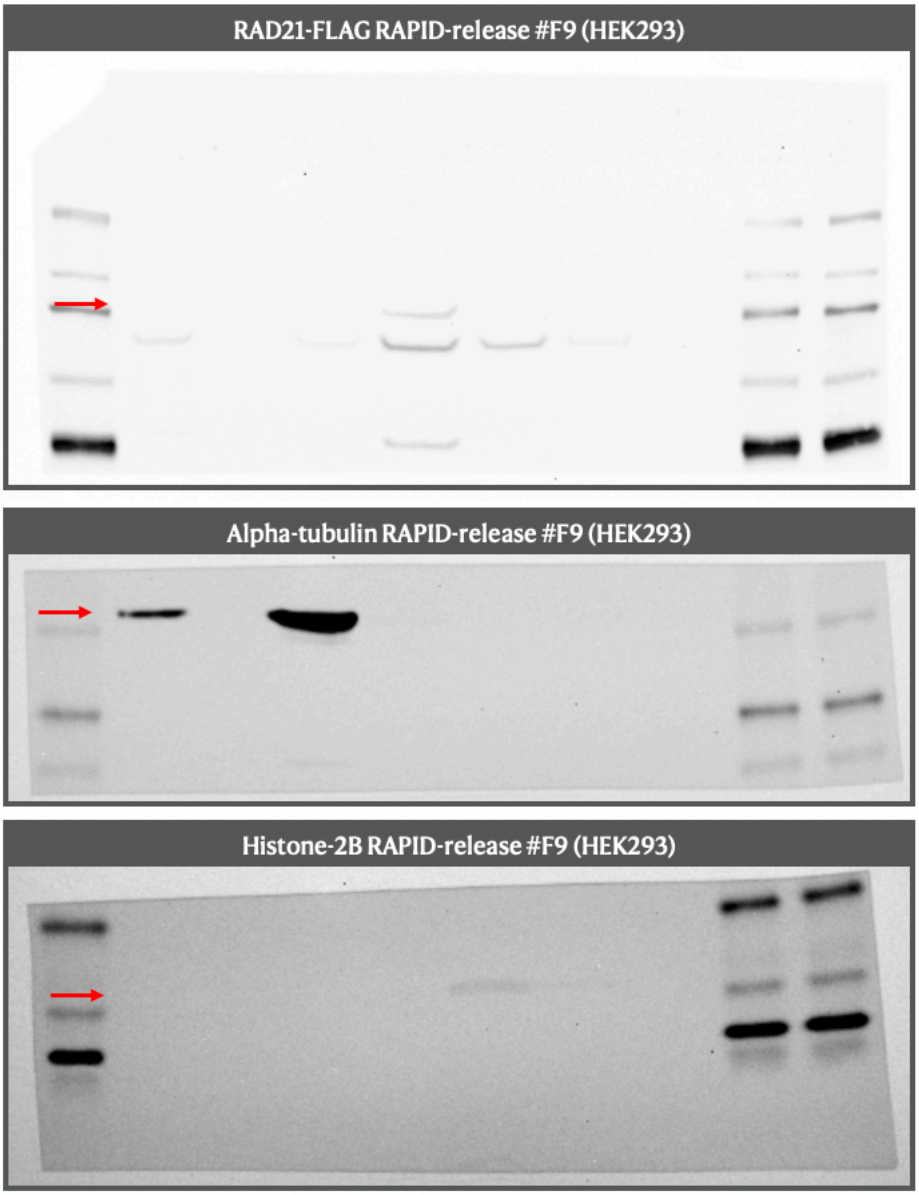



**Figure S11: Raw western blot membrane images for Figure 4.14B.**
Subcellular fractionation of HEK293 RAPID-release clone #F9 cells.

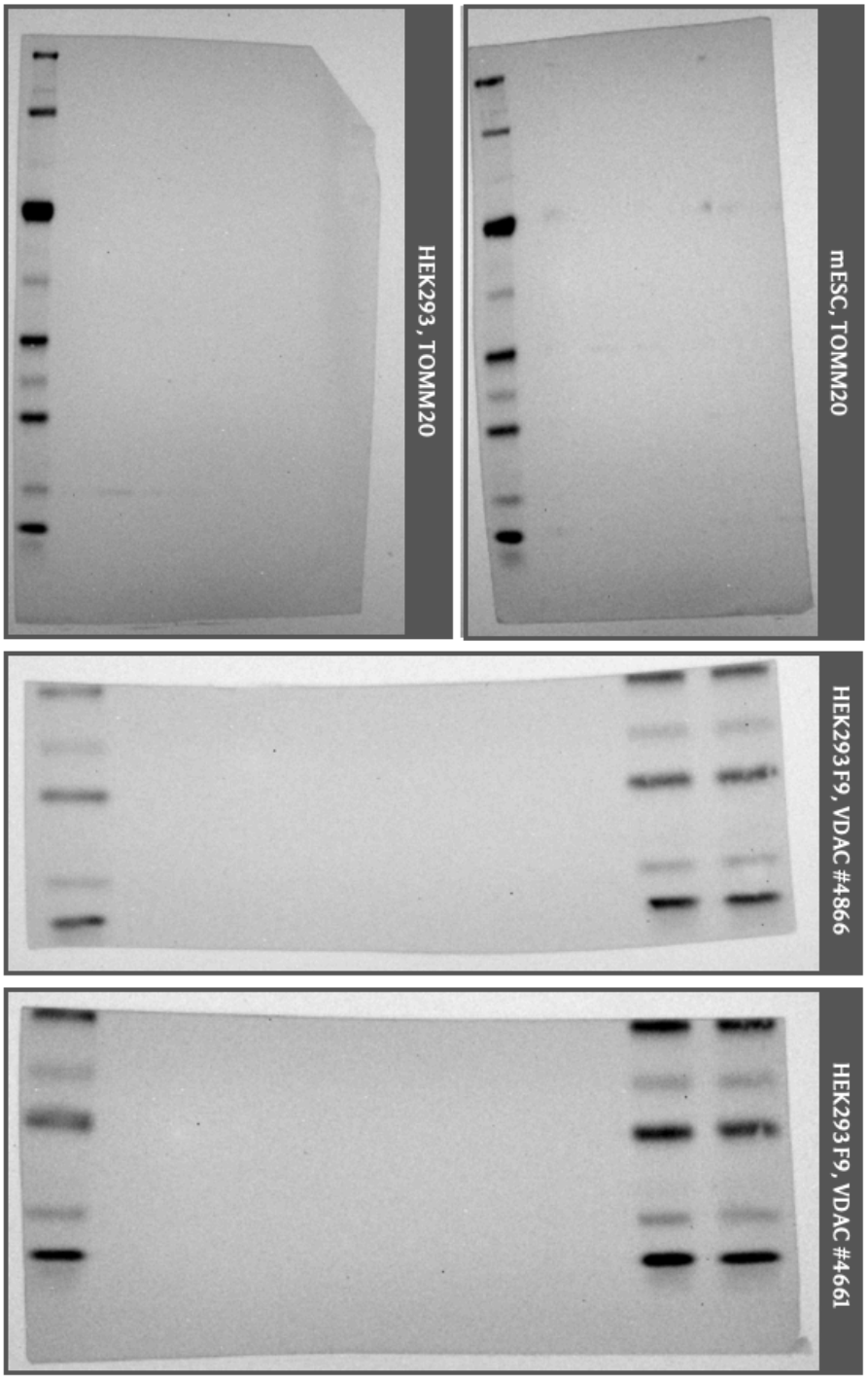



**Figure S12: Raw greyscale membrane images of TOMM20 and VDAC western blots.**
Two antibodies were tested as potential membrane-protein fraction controls, specifically mitochondrial markers. TOMM20 antibody (ab186734) should detect TOMM20 protein at ~15kDa. This antibody was tested in both HEK293 and mESC cells in whole-cell lysate and subcellular fractions. Two different VDAC antibodies were tested (Cell Signalling Technologies: 4661 and 4866), should detect VDAC protein at ~ 30kDa.

# References


1. Hatano, A. *et al.* CELLPEDIA: a repository for human cell information for cell studies and differentiation analyses. *Database (Oxford)* **2011**, bar046 (2011).

2. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).

3. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).

4. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).

5. Bernstein, B.E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences* **99**, 8695-8700 (2002).

6. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D. & Lieb, J.D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics* **36**, 900-905 (2004).

7. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).

8. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931-21936 (2010).

9. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).

10. Blackwood, E.M. & Kadonaga, J.T. Going the distance: a current view of enhancer action. *Science* **281**, 60-3 (1998).

11. Hatzis, P. & Talianidis, I. Dynamics of Enhancer-Promoter Communication during Differentiation-Induced Gene Activation. *Molecular Cell* **10**, 1467-1477 (2002).

12. Wang, Q., Carroll, J.S. & Brown, M. Spatial and Temporal Recruitment of Androgen Receptor and Its Coactivators Involves Chromosomal Looping and Polymerase Tracking. *Molecular Cell* **19**, 631-642 (2005).

13. Kong, S., Bohl, D., Li, C. & Tuan, D. Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Molecular and Cellular Biology* **17**, 3955-3965 (1997).

14. Tchurikov, N.A., Kretova, O.V., Moiseeva, E.D. & Sosin, D.V. Evidence for RNA synthesis in the intergenic region between enhancer and promoter and its inhibition by insulators in Drosophila melanogaster. *Nucleic Acids Res* **37**, 111-22 (2009).

15. Zhu, X. *et al.* A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res* **35**, 5532-44 (2007).

16. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.-F. & Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nature Genetics* **32**, 623-626 (2002).

17. Cook, P.R. & Marenduzzo, D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res* **46**, 9895-9906 (2018).

18. De Laat, W. & Grosveld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome Research* **11**, 447-459 (2003).

19. Deng, W. *et al.* Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell* **149**, 1233-1244 (2012).

20. Deng, W. *et al.* Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell* **158**, 849-860 (2014).

21. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell* **70**, 287-296.e6 (2018).

22. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358-368 (2016).

23. Crump, N.T. *et al.* BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nature Communications* **12**(2021).

24. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038-2049 (2016).

25. Dowen, M., Jill *et al.* Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell* **159**, 374-387 (2014).

26. Ing-Simmons, E. *et al.* Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Research* **25**, 504-513 (2015).

27. Stadhouders, R., Filion, G.J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345-354 (2019).

28. Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125-129 (2021).

29. Benko, S. *et al.* Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics* **41**, 359-364 (2009).

30. Bhatia, S. *et al.* Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved PAX6 Enhancer Causes Aniridia. *The American Journal of Human Genetics* **93**, 1126-1134 (2013).

31. Uslu, V.V. *et al.* Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics* **46**, 753-758 (2014).

32. Lupiáñez, G., Darío *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell***161**, 1012-1025 (2015).

33. Gorkin, U., David, Leung, D. & Ren, B. The 3D Genome in Transcriptional Regulation and Pluripotency. *Cell Stem Cell***14**, 762-775 (2014).

34. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nature Reviews Genetics***17**, 661-678 (2016).

35. *Johnson, G.* All Wound Up: DNA sequencing reveals the 3D structure of our DNA (2018).

36. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harbor Perspectives in Biology***2**, a003889-a003889 (2010).

37. Rabl, C. *Über Zelltheilung*, (Rabl C. 1885. U ¨ ber Zelltheilung. Morph Jb 10: 214–, Morph Jb., 1885).

38. Misteli, T. Chromosome territories: The arrangement of chromosomes in the nucleus. *Nature Education***1**(2008).

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

40. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *Journal of Cell Biology***217**, 4025-4048 (2018).

41. Chen, W. *et al.* RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome. *iScience***4**, 204-215 (2018).

42. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science***326**, 289-293 (2009).

43. Oudelaar, A.M. & Higgs, D.R. The relationship between genome structure and function. *Nature Reviews Genetics***22**, 154-168 (2021).

44. Rowley, M.J. & Corces, V.G. Organizational principles of 3D genome architecture. *Nature Reviews Genetics***19**, 789-800 (2018).

45. Van Steensel, B. & Furlong, E.E.M. The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology***20**, 327-337 (2019).

46. Rada-Iglesias, A., Grosveld, F.G. & Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Molecular Systems Biology***14**, e8214 (2018).

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

48. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature***485**, 381-385 (2012).

49. Hou, C., Li, L., Qin, S., Zhaohui & Corces, G., Victor. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell* **48**, 471-484 (2012).

50. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458-472 (2012).

51. De Wit, E. & Nora, E.P. New insights into genome folding by loop extrusion from inducible degron technologies. *Nature Reviews Genetics* (2022).

52. Zhan, Y. *et al.* Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Research* **27**, 479-490 (2017).

53. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Research* **24**, 390-400 (2014).

54. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

55. Phillips-Cremins, J. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**, 1281-1295 (2013).

56. Oudelaar, A.M. *et al.* Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nature Communications* **11**, 2722 (2020).

57. Goel, V.Y., Huseyin, M.K. & Hansen, A.S. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. (Cold Spring Harbor Laboratory, 2022).

58. Rao, S. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665-1680 (2014).

59. Monfils, K. & Barakat, T.S. Models behind the mystery of establishing enhancer-promoter interactions. *Eur J Cell Biol* **100**, 151170 (2021).

60. Holzmann, J., Fuchs, J., Pichler, P., Peters, J.-M. & Mechtler, K. Lesson from the Stoichiometry Determination of the Cohesin Complex: A Short Protease Mediated Elution Increases the Recovery from Cross-Linked Antibody-Conjugated Beads. *Journal of Proteome Research* **10**, 780-789 (2011).

61. Shi, Z., Gao, H., Bai, X.C. & Yu, H. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science* **368**, 1454-1459 (2020).

62. Gligoris, T.G. *et al.* Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science* **346**, 963-7 (2014).

63. Huis in 't Veld, P.J. *et al.* Characterization of a DNA exit gate in the human cohesin ring. *Science* **346**, 968-72 (2014).

64. Haering, C.H. *et al.* Structure and Stability of Cohesin's Smc1-Kleisin Interaction. *Molecular Cell* **15**, 951-964 (2004).

65. Li, Y. *et al.* The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 472-476 (2020).

66. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910 (2015).

67. Pugacheva, E.M. *et al.* CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc Natl Acad Sci U S A* **117**, 2020-2031 (2020).

68. Alipour, E. & Marko, J.F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res* **40**, 11202-12 (2012).

69. Wendt, K.S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801 (2008).

70. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell* **60**, 676-684 (2015).

71. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627 (2015).

72. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10**, 1297-309 (2015).

73. Chao, C.H., William *et al.* Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. *Cell Reports* **12**, 719-725 (2015).

74. Hinshaw, S.M., Makrantoni, V., Kerr, A., Marston, A.L. & Harrison, S.C. Structural evidence for Scc4-dependent localization of cohesin loading. *eLife* **4**(2015).

75. Kagey, M.H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430 (2010).

76. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **175**, 292-294 (2018).

77. Zuin, J. *et al.* A Cohesin-Independent Role for NIPBL at Promoters Provides Insights in CdLS. *PLoS Genetics* **10**, e1004153 (2014).

78. Zhu, Y., Denholtz, M., Lu, H. & Murre, C. Calcium signaling instructs NIPBL recruitment at active enhancers and promoters via distinct mechanisms to reconstruct genome compartmentalization. *Genes Dev* **35**, 65-81 (2021).

79. Boudaoud, I. *et al.* Connected Gene Communities Underlie Transcriptional Changes in Cornelia de Lange Syndrome. *Genetics* **207**, 139-151 (2017).

80. Rinzema, N.J. *et al.* Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat Struct Mol Biol* **29**, 563-574 (2022).

81. Sasca, D. *et al.* Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood* **134**, 2195-2208 (2019).

82. Busslinger, G.A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503-507 (2017).

83. Banigan, E.J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. (Cold Spring Harbor Laboratory, 2022).

84. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349 (2019).

85. Petela, N.J. *et al.* Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Molecular Cell* **70**, 1134-1148.e7 (2018).

86. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367-371 (2014).

87. Kanke, M., Tahara, E., Huis In'T Veld, P.J. & Nishiyama, T. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO Journal* **35**, 2686-2698 (2016).

88. Collier, J.E. *et al.* Transport of DNA within cohesin involves clamping on top of engaged heads by Scc2 and entrapment within the ring by Scc3. *eLife* **9**(2020).

89. Kanke, M., Tahara, E., Huis In'T Veld, P.J. & Nishiyama, T. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO Journal* **35**, 2686-2698 (2016).

90. Rhodes, J., Mazza, D., Nasmyth, K. & Uphoff, S. Scc2/Nipbl hops between chromosomal cohesin rings after loading. *eLife* **6**(2017).

91. Luppino, J.M. *et al.* NIPBL and WAPL balance cohesin activity to regulate chromatin folding and gene expression. (Cold Spring Harbor Laboratory, 2022).

92. Kanke, M., Tahara, E., Huis In'T Veld, P.J. & Nishiyama, T. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO Journal* **35**, 2686-2698 (2016).

93. Davidson, I.F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nature Reviews Molecular Cell Biology* **22**, 445-464 (2021).

94. Mayerova, N., Cipak, L. & Gregan, J. Cohesin Biology: From Passive Rings to Molecular Motors. *Trends Genet* **36**, 387-389 (2020).

95. Bauer, B.W. *et al.* Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**, 5448-5464.e22 (2021).

96. Kueng, S. *et al.* Wapl Controls the Dynamic Association of Cohesin with Chromatin. *Cell* **127**, 955-967 (2006).

97. Tedeschi, A. *et al.* Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564-568 (2013).

98. Gandhi, R., Gillespie, P.J. & Hirano, T. Human Wapl Is a Cohesin-Binding Protein that Promotes Sister-Chromatid Resolution in Mitotic Prophase. *Current Biology* **16**, 2406-2417 (2006).

99. Beckouët, F. *et al.* Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Molecular Cell* **61**, 563-574 (2016).

100. Ouyang, Z. *et al.* Structure of the human cohesin inhibitor Wapl. *Proceedings of the National Academy of Sciences* **110**, 11355-11360 (2013).

101. Morales, C. *et al.* PDS5 proteins are required for proper cohesin dynamics and participate in replication fork protection. *Journal of Biological Chemistry* **295**, 146-157 (2020).

102. Nishiyama, T. *et al.* Sororin Mediates Sister Chromatid Cohesion by Antagonizing Wapl. *Cell* **143**, 737-749 (2010).

103. Kikuchi, S., Borek, D.M., Otwinowski, Z., Tomchick, D.R. & Yu, H. Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proceedings of the National Academy of Sciences* **113**, 12444-12449 (2016).

104. Van Ruiten, M.S. *et al.* The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. *Nature Structural & Molecular Biology* **29**, 586-591 (2022).

105. Gerlich, D., Koch, B., Dupeux, F., Peters, J. & Ellenberg, J. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication. *Current Biology* **16**, 1571-1578 (2006).

106. Gause, M., Misulovin, Z., Bilyeu, A. & Dorsett, D. Dosage-Sensitive Regulation of Cohesin Chromosome Binding and Dynamics by Nipped-B, Pds5, and Wapl. *Molecular and Cellular Biology* **30**, 4940-4951 (2010).

107. Ladurner, R. *et al.* Cohesin's ATPase Activity Couples Cohesin Loading onto DNA with Smc3 Acetylation. *Current Biology* **24**, 2228-2237 (2014).

108. Wutz, G. *et al.* ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *eLife* **9**(2020).

109. Casa, V. *et al.* Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. *Genome Research* **30**, 515-527 (2020).

110. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**(2017).

111. Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. Vol. 42 23 (Journal of Physics A: Mathematical and Theoretical, 2009).

112. Van Der Lelij, P. *et al.* Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. *eLife* **6**(2017).

113. Cuadrado, A. & Losada, A. Specialized functions of cohesins STAG1 and STAG2 in 3D genome architecture. *Curr Opin Genet Dev* **61**, 9-16 (2020).

114. Kojic, A. *et al.* Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nature Structural & Molecular Biology* **25**, 496-504 (2018).

115. Countryman, P. *et al.* Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates. *Journal of Biological Chemistry* **293**, 1054-1069 (2018).

116. Sanborn, A.L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences* **112**, E6456-E6465 (2015).

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

118. Davidson, I.F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).

119. Xi, W. & Beer, M.A. Loop competition and extrusion model predicts CTCF interaction specificity. *Nature Communications* **12**(2021).

120. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498 (2001).

121. Bosch, P.S., Pepperl, J. & Basler, K. Anchor Away - A Fast, Reliable and Reversible Technique To Inhibit Proteins in. *G3 (Bethesda)* **10**, 1745-1752 (2020).

122. Nora, E.P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).

123. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods* **6**, 917-922 (2009).

124. Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nature Communications* **11**, 5701 (2020).

125. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nature Chemical Biology* **14**, 431-441 (2018).

126. Békés, M., Langley, D.R. & Crews, C.M. PROTAC targeted protein degraders: the past is prologue. *Nature Reviews Drug Discovery* **21**, 181-200 (2022).

127. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO Journal* **36**, 3573-3599 (2017).

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

129. Thiecke, M.J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Reports* **32**, 107929 (2020).

130. Rhodes, J.D.P. *et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Reports* **30**, 820-835.e10 (2020).

131. Kriz, A.J., Colognori, D., Sunwoo, H., Nabet, B. & Lee, J.T. Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation. *Mol Cell* **81**, 1970-1987.e9 (2021).

132. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal* **32**, 3119-3129 (2013).

133. Luppino, J.M. *et al.* Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nature Genetics* **52**, 840-848 (2020).

134. Hafner, A., Park, M., Berger, S.E., Nora, E.P. & Boettiger, A.N. Loop stacking organizes genome folding from TADs to chromosomes. (Cold Spring Harbor Laboratory, 2022).

135. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496-501 (2022).

136. Mach, P. *et al.* Live-cell imaging and physical modeling reveal control of chromosome folding dynamics by cohesin and CTCF. (Cold Spring Harbor Laboratory, 2022).

137. Liu, N.Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics* **53**, 100-109 (2021).

138. Lupiáñez, D.G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237 (2016).

139. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017-1021 (2015).

140. Flavahan, W.A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229-233 (2019).

141. Despang, A. *et al.* Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nature Genetics* **51**, 1263-1271 (2019).

142. Xu, B. *et al.* Acute depletion of CTCF rewires genome-wide chromatin accessibility. *Genome Biology* **22**(2021).

143. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res* **47**, 6699-6713 (2019).

144. Haarhuis, J.H.I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e14 (2017).

145. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).

146. Oudelaar, A.M., Beagrie, R.A., Kassouf, M.T. & Higgs, D.R. The mouse alpha-globin cluster: a paradigm for studying genome regulation and organization. *Curr Opin Genet Dev* **67**, 18-24 (2021).

147. Harrold, C. University of Oxford (2019).

148. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).

149. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics* **38**, 1341-1347 (2006).

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

151. Hughes, J.R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics* **46**, 205-212 (2014).

152. Davies, J.O.J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nature Methods* **13**, 74-80 (2016).

153. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research* **16**, 1299-1309 (2006).

154. McCord, R.P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell* **77**, 688-708 (2020).

155. Davies, J.O.J., Oudelaar, A.M., Higgs, D.R. & Hughes, J.R. How best to identify chromosomal interactions: a comparison of approaches. *Nature Methods* **14**, 125-134 (2017).

156. Downes, D.J. *et al.* Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. *Nature Protocols* **17**, 445-475 (2022).

157. Hanssen, L.L.P. *et al.* Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology* **19**, 952-961 (2017).

158. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120 (2012).

159. Phillips, J.E. & Corces, V.G. CTCF: Master Weaver of the Genome. *Cell* **137**, 1194-1211 (2009).

160. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L.A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harbor Symposia on Quantitative Biology* **82**, 45-55 (2017).

161. Francis, H.S. *et al.* Scalable in vitro production of defined mouse erythroblasts. *PLOS ONE* **17**, e0261950 (2022).

162. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biology* **7**, S4 (2006).

163. Harrow, J. *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research* **22**, 1760-1774 (2012).

164. Arruda, N.L. *et al.* Distinct and overlapping roles of STAG1 and STAG2 in cohesin localization and gene expression in embryonic stem cells. *Epigenetics & Chromatin* **13**(2020).

165. Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences* **107**, 3651-3656 (2010).

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

167. Bailey, T.L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202-W208 (2009).

168. Anguita, E. *et al.* Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J* **23**, 2841-52 (2004).

169. Dubois-Chevalier, J. *et al.* A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res* **42**, 10943-59 (2014).

170. Brown, J.M. *et al.* A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nature Communications* **9**(2018).

171. Nguyen, T.A. *et al.* High-throughput functional comparison of promoter and enhancer activities. *Genome Research* **26**, 1023-1033 (2016).

172. Van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nature Biotechnology* **35**, 145-153 (2017).

173. Dao, L.T.M. *et al.* Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics* **49**, 1073-1081 (2017).

174. Arnold, C.D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-7 (2013).

175. Bediaga, N.G. *et al.* Multi-level remodelling of chromatin underlying activation of human T cells. *Scientific Reports* **11**(2021).

176. Zhang, C. *et al.* tagHi-C Reveals 3D Chromatin Architecture Dynamics during Mouse Hematopoiesis. *Cell Reports* **32**, 108206 (2020).

177. Gil, D.A., Cuadrado, A., Giménez-Llorente, D., Rodríguez-Corsino, M. & Losada, A. Different NIPBL requirements of cohesin-STAG1 and cohesin-STAG2. (Cold Spring Harbor Laboratory, 2022).

178. Oh, K. Technical Considerations for Contemporary Western Blot Techniques. in *Methods in Molecular Biology* 457-479 (Springer US, 2021).

179. Holzmann, J. *et al.* Absolute quantification of cohesin, CTCF and their regulators in human cells. *eLife* **8**, e46269 (2019).

180. Hentges, L.D. *et al.* LanceOtron: a deep learning peak caller for genome sequencing experiments. *Bioinformatics* (2022).

181. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).

182. Robinson, J.T. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Systems* **6**, 256-258.e1 (2018).

183. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

184. Liu, N.Q. *et al.* Rapid depletion of CTCF and cohesin proteins reveals dynamic features of chromosome architecture. (Cold Spring Harbor Laboratory, 2021).

185. Ocampo-Hafalla, M.T. & Uhlmann, F. Cohesin loading and sliding. *Journal of Cell Science* **124**, 685-691 (2011).

186. Higashi, T.L., Pobegalov, G., Tang, M., Molodtsov, M.I. & Uhlmann, F. A Brownian ratchet model for DNA loop extrusion by the cohesin complex. *eLife* **10**(2021).

187. Hansen, A.S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132-148 (2020).

188. Balleza, E., Kim, J.M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods* **15**, 47-51 (2018).

189. Apta-Smith, M.J., Hernandez-Fernaud, J.R. & Bowman, A.J. Evidence for the nuclear import of histones H3.1 and H4 as monomers. *The EMBO Journal* **37**, e98714 (2018).

190. Singh, V.P. & Gerton, J.L. Cohesin and human disease: lessons from mouse models. *Curr Opin Cell Biol* **37**, 9-17 (2015).

191. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).

192. Perez-Pinera, P., Ousterout, D.G., Brown, M.T. & Gersbach, C.A. Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res* **40**, 3741-52 (2012).

193. McLenachan, S., Sarsero, J.P. & Ioannou, P.A. Flow-cytometric analysis of mouse embryonic stem cell lipofection using small and large DNA constructs. *Genomics* **89**, 708-20 (2007).

194. von Groll, A., Levin, Y., Barbosa, M.C. & Ravazzolo, A.P. Linear DNA low efficiency transfection by liposome can be improved by the use of cationic lipid as charge neutralizer. *Biotechnol Prog* **22**, 1220-4 (2006).

195. Hardee, C., Arévalo-Soliz, L., Hornstein, B. & Zechiedrich, L. Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes* **8**, 65 (2017).

196. Tian, B., Yang, J. & Brasier, A.R. Two-Step Cross-linking for Analysis of Protein–Chromatin Interactions. in *Methods in Molecular Biology* 105-120 (Springer New York, 2012).

197. de Jonge, W.J., Brok, M., Kemmeren, P. & Holstege, F.C.P. An extensively optimized chromatin immunoprecipitation protocol for quantitatively comparable and robust results. *bioRxiv*, 835926 (2019).

198. Zaidi, H., Hoffman, E.A., Shetty, S.J., Bekiranov, S. & Auble, D.T. Second-generation method for analysis of chromatin binding with formaldehyde–cross-linking kinetics. *Journal of Biological Chemistry* **292**, 19338-19355 (2017).

199. Thomas, P. & Smart, T.G. HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods* **51**, 187-200 (2005).

200. Qin, J.Y. *et al.* Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PloS one* **5**, e10611 (2010).

201. Meilinger, D. *et al.* Np95 interacts withde novoDNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO reports* **10**, 1259-1264 (2009).

202. Mehta, A.K., Majumdar, S.S., Alam, P., Gulati, N. & Brahmachari, V. Epigenetic regulation of cytomegalovirus major immediate-early promoter activity in transgenic mice. *Gene* **428**, 20-4 (2009).

203. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281-2308 (2013).

204. Posakony, J., England, J. & Attardi, G. Mitochondrial growth and division during the cell cycle in HeLa cells. *Journal of Cell Biology* **74**, 468-491 (1977).

205. Henderson, B.R. & Eleftheriou, A. A comparison of the activity, sequence specificity, and CRM1-dependence of different nuclear export signals. *Exp Cell Res* **256**, 213-24 (2000).

206. Xu, D., Farmer, A., Collett, G., Grishin, N.V. & Chook, Y.M. Sequence and structural analyses of nuclear export signals in the NESdb database. *Molecular Biology of the Cell* **23**, 3677-3693 (2012).

207. Lattore, A. & Silva, F.J. The size of the genome and the complexity of living beings. (Mètode Science Studies Journal, 4., 2014).

208. Salzberg, S.L. Open questions: How many genes do we have? *BMC Biology* **16**(2018).

209. Spielmann, M., Lupiáñez, D.G. & Mundlos, S. Structural variation in the 3D genome. *Nature Reviews Genetics* **19**, 453-467 (2018).

210. Ouyang, W., Xiong, D., Li, G. & Li, X. Unraveling the 3D Genome Architecture in Plants: Present and Future. *Molecular Plant* **13**, 1676-1693 (2020).

211. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244 (2015).

212. Hsieh, S., Tsung-Han *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108-119 (2015).

213. Lyu, X., Rowley, M.J. & Corces, V.G. Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to Temperature Stress. *Molecular Cell* **71**, 940-955.e7 (2018).

214. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics* **21**, 71-87 (2020).

215. Blayney, J. *et al.* Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression. (Cold Spring Harbor Laboratory, 2022).

216. Dequeker, B.J.H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197-203 (2022).

217. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).

218. Uhlmann, F. SMC complexes: from DNA to chromosomes. *Nature Reviews Molecular Cell Biology* **17**, 399-412 (2016).

219. Marsden, M.P. & Laemmli, U.K. Metaphase chromosome structure: evidence for a radial loop model. *Cell* **17**, 849-58 (1979).

220. Golfier, S., Quail, T., Kimura, H. & Brugués, J. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *eLife* **9**(2020).

221. Liu, L., Michowski, W., Kolodziejczyk, A. & Sicinski, P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nature Cell Biology* **21**, 1060-1067 (2019).

222. Liu, Y. & Dekker, J. CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nature Cell Biology* **24**, 1516-1527 (2022).

223. Miura, H. & Hiratani, I. Cell cycle dynamics and developmental dynamics of the 3D genome: toward linking the two timescales. *Curr Opin Genet Dev* **73**, 101898 (2022).

224. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics* **20**, 437-455 (2019).

225. Hsieh, T.-H.S. *et al.* Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nature Genetics* **54**, 1919-1932 (2022).

226. Szutorisz, H., Dillon, N. & Tora, L. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci* **30**, 593-9 (2005).

227. Johnson, K.D. *et al.* Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. *Proceedings of the National Academy of Sciences* **99**, 11760-11765 (2002).

228. Louie, M.C. *et al.* Androgen-induced recruitment of RNA polymerase II to a nuclear receptor–p160 coactivator complex. *Proceedings of the National Academy of Sciences* **100**, 2226-2230 (2003).

229. Shang, Y., Myers, M. & Brown, M. Formation of the Androgen Receptor Transcription Complex. *Molecular Cell* **9**, 601-610 (2002).

230. Wurmser, A. & Basu, S. Enhancer-Promoter Communication: It's Not Just About Contact. *Front Mol Biosci* **9**, 867303 (2022).

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

232. Cardozo Gizzi, A.M., Cattoni, D.I. & Nollmann, M. TADs or no TADS: Lessons From Single-cell Imaging of Chromosome Architecture. *J Mol Biol* **432**, 682-693 (2020).

233. Nichols, J., Evans, E.P. & Smith, A.G. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341-1348 (1990).

234. Smith, A.G. Culture and differentiation of embryonic stem cells. *Journal of Tissue Culture Methods* **13**, 89-94 (1991).

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

236. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

237. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).

238. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16**(2015).

239. Akdemir, K.C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biology* **16**(2015).

240. Bonfield, J.K. *et al.* HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **10**(2021).

241. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95 (2007).

242. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-4 (2015).

243. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2021).

244. Team, S.T.D. SRA Toolkit

245. Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics* **43**, 1091-1097 (2011).

246. Ramisch, A. *et al.* CRUP: a comprehensive framework to predict condition-specific regulatory units. *Genome Biology* **20**(2019).

247. Matsumoto, K., Suzuki, A., Wakaguri, H., Sugano, S. & Suzuki, Y. Construction of mate pair full-length cDNAs libraries and characterization of transcriptional start sites and termination sites. *Nucleic Acids Res* **42**, e125 (2014).

248. Liu, N.Q. *et al.* WAPL maintains dynamic cohesin to preserve lineage specific distal gene regulation. *bioRxiv*, 731141 (2019).

249. Lee, B.T. *et al.* The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* **50**, D1115-D1122 (2022).

250. Fursova, N.A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Molecular Cell* **74**, 1020-1036.e8 (2019).

251. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).

252. Castro-Mondragon, J.A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165-D173 (2022).

253. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-8 (2011).

254. Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. The MEME Suite. *Nucleic Acids Res* **43**, W39-49 (2015).