

The role of loop extrusion in the dynamics of chromosome folding

INAUGURALDISSERTATION

zur Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der UNIVERSITÄT BASEL

von

PIA MACH

Basel, 2023

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Erstbetreuer: Dr. Luca Giorgetti

Zweitbetreuer: Prof. Dr. Dirk Schübeler

Externer Experte: Dr. Anton Goloborodko

Basel, den 14. November 2023

Prof. Dr. Marcel Mayor

Dekan

Acknowledgments

I would like to thank all of the people that accompanied me during my PhD journey and without whom this project would not have been possible.

First and foremost, I would like to thank Luca for giving me the opportunity to pursue my PhD in his lab. He has always been a very supportive mentor who gave me the freedom to explore, learn and discover science and this project by myself, encouraged me when I was being too critical and advised me when I could not see the forest for the trees. His expertise, enthusiasm for science and guidance have pushed me further than I could have ever anticipated. Thank you for always believing in me.

I would like to express my gratitude and appreciation to the members of my thesis committee, Dirk Schübeler and Anton Goloborodko, for your input, constructive feedback and guidance, be it during committee meetings or when asking for career advice. A special thanks also to Maxim, Pavel, Sandra, Ilya and Jana for proof-reading parts of this thesis.

Then, of course, I would like to thank all present and former members of the Giorgetti lab, that I had to pleasure to meet and work with: Julie Cramard, Pavel Kos, Jana Tünnermann, Gergely Tihanyi, Gregory Roth, Ilya Flymer, Nessim Louafi, Julie Cordier, Kristina Makasheva, Elena Testoni, Jessica Zuin, Yinxiu Zhan, Simon Gaudin, Josef Redolfi, Ewa Piskadlo, Tania Distler, Marco Michalski, Alexis Cornec and Mariya Kryzhanovska. A special thanks to Jessica and Julie who introduced me to the lab and the pleasures of genome engineering. Thank you to Julie for the occasional crafting adventure. Thank you to Jana, who I could always count on for “discussing” the ways of the TIRF and her advice on microscopy. I thank the entire 3.72 office, but especially Ilya, for trying to teach me sarcasm: Even though you did not succeed, you never gave up.

I would like to express my deepest gratitude to Pavel and Zhan, for teaching me to (sometimes) think like a physicist and for supporting me and sharing this project. I am sure I have become a better scientist because of you.

I would especially like to thank Simon and Nessim, who I had the pleasure to introduce to live-cell imaging: I am very grateful for your invaluable input, help and enthusiasm for the work we did. I wish you all the best for your future.

A big thanks to all the members of FAIM: In particular, Laurent Gelman and Laure Plantard,

who, I am sure, share my feelings towards the TIRF and who taught me more about microscopy that I could have ever imagined. A thanks to Jan Eglinger and Tim-Oliver Buchholz, who always supported me in my image analysis endeavors and taught me the importance of good data. Thank you all, for your help with my projects. Without you, it would not have been possible. A big thanks also to the TIRF for almost never letting me down and surprising me with always beautiful imaging data.

I would also like to acknowledge the support of the Functional Genomics facility, especially Sebastien Smallwood, the support from Hubertus Köhler for cell sorting, and the support from Gregory Roth for all my questions regarding mathematical modeling. I would also like to thank the members of the Chao lab for our joint lab meetings and their input in all things imaging, in particular Jeff, Franka and Tobias.

A very special thanks goes to the ChromDesign community: Thank you for all your support and feedback over the three years. Thank you to all the ChromDesigners, in particular Tina, Carla and Antonia, that made me feel part of a community with a PhD as a shared goal. A special thanks also to Giacomo Cavalli for his input as my second supervisor within the ChromDesign network. I would also like to thank Edith Heard, Geneviève Almouzni, Antonia Hauth, Tina Karagyoza, the team of the EMBL Imaging Center, Bassam Hajj and Tommaso Galgani for giving me the opportunity to perform my secondments in their labs. It was a great experience and I learned a lot from all of you.

I would like to thank my collaborators, Edoardo Marchi, Guido Tiana, Simon Grosse-Holz, Laura Caccianini, Leonid Mirny and in particular Elphège Nora, for their support and expertise throughout our many (Zoom) meetings. It was a great pleasure to share these projects with you.

I would like to express my deepest gratitude towards my friends, Jessica, Sandra and Pedro, Tobias and Claudia, Tommaso, Anjela, Paul, Lisa and Marius. Our weekend trips, gaming and movie nights, the evenings at the Rhine (or even at the FMI) were the highlights of my PhD. I am so grateful for all the fun memories we created together and your kind and helpful words, when I needed them. Whatever comes next for all of us, you made my time in Basel truly unforgettable.

Most of all, I would like to thank Maxim for being there for me, comforting and supporting me. I am so very glad that we found each other and that I could share this journey with you. Thank you for your understanding in hard times (especially in 2022), your belief in me, your encouragement, your support and your love.

Finally, I am eternally grateful to my entire family, but especially my parents, Julia and Veit, who have always supported me. Without your belief in me, I would not be where I am today. I dedicate this thesis to you for everything you have done for me.

Table of Contents

| | |
|---|------------|
| List of Abbreviations | VI |
| List of Figures | X |
| List of Tables | XII |
| Abstract | 1 |
| Outline | 2 |
| 1 Introduction | 5 |
| 1.1 Transcriptional regulation in the mammalian genome | 5 |
| 1.1.1 Enhancers and their role in transcriptional regulation | 6 |
| 1.1.2 TF binding in a nucleosomal context | 7 |
| 1.1.3 Interactions of TFs with cofactors | 8 |
| 1.1.4 Modulation of transcriptional output by enhancer sequence | 10 |
| 1.1.5 Models of E-P communication | 11 |
| 1.2 Genome architecture and chromosome folding | 15 |
| 1.2.1 Methods to study 3D genome architecture | 15 |
| 1.2.2 “Contacts” in DNA-FISH and 3C methods | 18 |
| 1.2.3 Hierarchical organization of the 3D genome | 19 |
| 1.2.4 Topologically associating domains | 22 |
| 1.3 Cohesin and the loop extrusion model | 25 |
| 1.3.1 The cohesin complex | 26 |
| 1.3.2 SMC complexes during cell cycle progression | 28 |
| 1.3.3 Loop extrusion | 29 |
| 1.3.4 Barrier elements to extrusion: CTCF | 33 |
| 1.3.5 Regulators of cohesin | 36 |
| 1.4 Dynamics of chromosome folding processes | 39 |
| 1.4.1 Methods to study dynamics of chromosomes | 39 |
| 1.4.2 Time scales of chromosomal interactions | 45 |
| 1.5 Aim of this thesis | 45 |

| | | |
|----------|---|------------|
| 2 | Cohesin and CTCF control the dynamics of chromosome folding | 47 |
| 2.1 | Abstract | 48 |
| 2.2 | Introduction | 48 |
| 2.3 | Results | 49 |
| 2.3.1 | Cohesin decreases chromosome mobility independently of CTCF | 49 |
| 2.3.2 | Loop extrusion can explain reduced chromosome dynamics | 52 |
| 2.3.3 | Cohesin and CTCF constrain the dynamics of sequences in <i>cis</i> | 55 |
| 2.3.4 | Chromosomal contacts are transient | 58 |
| 2.4 | Discussion | 63 |
| 2.5 | Methods | 64 |
| 2.6 | Extended Data | 72 |
| 3 | Cohesin dynamics on DNA: Cohesion or Extrusion? | 85 |
| 3.1 | Abstract | 86 |
| 3.2 | Introduction | 86 |
| 3.3 | Results | 87 |
| 3.3.1 | Cohesin and DNA dynamics in living cells | 87 |
| 3.3.2 | Cohesive and extrusive cohesin impose constraints on DNA motion | 90 |
| 3.3.3 | Comparison of cohesin motion and DNA motion | 92 |
| 3.4 | Discussion and outlook | 94 |
| 3.5 | Methods | 95 |
| 4 | Discussion | 103 |
| 4.1 | Cohesin globally slows down chromosome dynamics | 103 |
| 4.2 | Global chromosome motion is not influenced by CTCF | 105 |
| 4.3 | CTCF and cohesin control looping dynamics | 105 |
| 4.4 | What is a contact? | 107 |
| 4.5 | Inference of the cohesin-mediated looped state | 108 |
| 4.6 | Looping dynamics and regulation of transcription | 110 |
| 4.7 | Long-term single-particle tracking of cohesin on DNA | 111 |
| 4.8 | Conclusion | 113 |
| | References | 115 |
| | Supplementary Information | 145 |
| | Chapter 2 – Cohesin and CTCF control the dynamics of chromosome folding | 145 |
| | Methods | 145 |
| | Polymer modeling | 154 |
| | Supplementary Figures | 160 |
| | Supplementary Videos | 161 |
| | Supplementary Tables | 161 |
| | Flow cytometry gating strategy | 163 |

| | |
|---|------------|
| Chapter 3 – Cohesin dynamics on DNA: Cohesion or Extrusion? | 164 |
| Genotypes of degron cell lines | 164 |
| Overview on image analysis pipeline | 165 |
| Appendix | 166 |
| Integrative approaches to study enhancer-promoter communication | 168 |
| Cohesin and CTCF control the dynamics of chromosome folding | 178 |
| Nonlinear control of transcription through enhancer-promoter interactions | 205 |
| Curriculum Vitae | 226 |

List of Abbreviations

- 2D** two-dimensional. 16, 67–69, 98, 100, 153
- 3C** chromosome conformation capture. 7, 11, 15–20, 39, 48, 63, 103, 105, 107, 149, 150, VI, VII, X
- 3D** three-dimensional. 5, 7, 10, 15, 17, 18, 29, 36, 43, 51, 58, 67–69, 79, 86, 89, 100, 103, 105, 107, 109, 153
- 4C** chromosome conformation capture-on-chip (also circular 3C). 15, 47, 51, 64, 71, 73, 150, 162
- 5C** chromosome conformation capture carbon copy. 15
- ADP** adenosine diphosphate. 35
- AID** auxin-inducible degron. 23, 38, 50–52, 64, 72, 73, 75, 87, 88, 91, 92, 94–96, 148, 161, 164
- APC** anaphase promoting complex. 29
- ATAC-seq** assay for transposase-accessible chromatin using sequencing. 7
- ATP** adenosine triphosphate. 11, 21, 27, 29, 30, 32, 33, 36
- ATPase** adenosine 5'-triphosphatase. 26, 27, 32, 33, 36
- BE** barrier element. 29, 31, 105, 106, 108
- BILD** Bayesian inference of looping dynamics. 109
- bp** base pair. 7, 8, 20, 73, 148, 150, 152
- CDF** cumulative distribution function. 18
- CdLS** Cornelia de Lange syndrome. 25
- ChIP-seq** chromatin immunoprecipitation followed by sequencing. 6, 31, 34, 49
- CID** chromosomal interaction domain. 22
- CLEM** correlative light and electron microscopy. 14, 17
- CNN** convolutional neural net. 98
- CpG** 5'-C-phosphate-G-3'. 7, 8, 35
- CRISPR** clustered regularly interspaced short palindromic repeats. 40, 41
- cryo-EM** cryo-electron microscopy. 13, 32
- CTCF** CCCTC-binding factor. 1, 2, 9, 15, 20, 22–25, 27–31, 33–38, 42, 45, 46, 48–66, 70, 72–75, 79–83, 86, 87, 94, 96, 103–110, 113, 145–148, 151, 153, 154, 161, X, XII
- CuO** cumate operator. 40
- DNA** deoxyribonucleic acid. 1, 3, 5–9, 11–13, 15–19, 21–23, 25, 27–46, 48, 52, 55, 56, 59, 65, 86–95, 100, 103–108, 111–113, 145, 146, 148–151, 165, X, XI

DNase I deoxyribonuclease I. 6

E-P enhancer-promoter. 9–11, 13–15, 20, 22, 23, 39, 40, 44, 45, 106–108, 110, 111, 113

EF3 elongation factor 3. VII

eGFP enhanced green fluorescent protein. 40, 50, 51, 56–59, 64, 66, 67, 69, 72, 73, 79, 87, 88, 91, 92, 95, 96, 145–148, 161, 164

EM electron microscopy. 17, 18, 31, VI

EPHA4 ephrin type-A receptor 4. 23

eRNA enhancer RNA. 6

ESCO establishment of cohesion homolog. 25, 27, 28, 36–38

eve even-skipped. 105, 110

FACS fluorescence activated cell sorting. 64, 65, 96

FCS fluorescence correlation spectroscopy. 39, 42

FISH fluorescence *in situ* hybridization. 15, 17–20, 22, 25, 39, 44, 45, 103, 106, 107, X

FKBP FK506-binding protein. 56, 66, 78, 94–96, 147, 148, 164

FOV field of view. 89, 90, 97–99

FRAP fluorescence recovery after photobleaching. 38, 39, 42, 45, 86, 108

FROS fluorescein repressor operator system. 40, 41, 103, 104

FUCCI fluorescence ubiquitin cell cycle indicator. 111

GAM genome architecture mapping. 16

gRNA guide RNA. 65, 78, 146–149, 162

H3K27ac histone H3 lysine 27 acetylation. 8

H3K27me3 histone H3 lysine 27 trimethylation. 8, 13, 22

H3K4me3 histone H3 lysine 4 trimethylation. 8, 22

H3K9me3 histone H3 lysine 9 trimethylation. 9, 13

HAWK HEAT repeated proteins associated with kleisins. 26–28, 32, 36

HDAC8 Hos1 in budding yeast 8. 37

HEAT acronym from four proteins: Huntingtin, EF3, PP2A and yeast kinase TOR1. VII

Hi-C high-throughput chromosome conformation capture. 15–20, 22, 23, 25, 29–31, 34–36, 39, 40, 43, 44, 110, IX

HILO highly inclined and laminated optical sheet. 89, 97

HMM hidden Markov model. 59–62, 70, 81–83, 109, 160–162

HoxB Homeobox B protein. 13

HP1 heterochromatin protein 1. 13, 22

IDR intrinsically disordered domain. 22

IGF2 insulin-like growth factor 2. 35

IGH immunoglobulin heavy chain. 25

ITR inverted terminal repeats. 50, 65, 79, 145, 146, 148, 164

kb kilobase. 11, 20, 22, 23, 41, 45, 46, 49, 51, 52, 54–63, 76–80, 82, 83, 86, 94, 103, 105, 106, 108, 145, 149, 152, 153, 160, 161

LacI LacI repressor. 40, 56–58, 66, 145–147, 151

LacO Lac operator. 40, 56–60, 65, 66, 68, 78–81, 146–149, 152, 153, 161

LAD lamina-associated domain. 106

Lbd1 LIM domain-binding family protein. 13

LCR locus control region. 6, 11

LEF loop-extruding factor. 29–31

LIM acronym from three proteins: LIN-11, Isl-1 and MEC-3. VIII

LoG Laplacian of Gaussian. 99, 100

Mb megabase. 6, 10, 11, 17, 20, 22, 43–45, 52, 53, 55, 57, 61, 76, 78, 80, 154, 161

MCM minichromosome maintenance. 26, 28, 31

mESC mouse embryonic stem cell. 1, 24, 35, 45, 46, 48–51, 55–58, 63–65, 67, 68, 72, 79, 80, 87, 95, 104, 106, 107, 109, XII

MNase micrococcal nuclease. 16

MPRA massive parallel reporter assay. 10, 11

MSD mean squared displacement. 51, 52, 54, 55, 68, 69, 89–93, 101, 106, 109, 153, 157, 165

N2V2 Noise2Void2. 89, 90, 98, 99, 112, 165

nCATS Nanopore Cas9-targeted sequencing. 56, 149

NGS next-generation sequencing. 6, 15

NIPBL nipped-B-like protein (SCC2 in mice). 1, 23–25, 27, 28, 32, 36–38, 46, 86, 87, 90–96, 103, 112, 164

nm nanometer. 13, 14, 16, 17, 21, 26, 28, 42, 43, 56, 58, 59, 61, 63, 64, 66–70, 79, 89, 96–98, 106, 107, 109

PALM photo-activated localization microscopy. 17, 42

PAPA proximity-assisted photoactivation. 14

PCA principle component analysis. 20

pHi-C promoter capture Hi-C. 16

PCR polymerase chain reaction. 15, 96

PDS5 precocious dissociation of sisters 5 protein. 24, 27, 36–38, 86, 94–96, 98, 113, 164, XI

Pgk 1 Phosphoglycerate kinase 1. 58

PIC pre-initiation complex. 8, 13, 26

PolII RNA Polymerase II. 5, 7, 11, 13, 14, 26, 28, 31, 37, 42, 111

PP2A protein phosphatase 2A. 29, VII

PRC1 Polycomb repressive complex 1. 22, 42

PSF point-spread-function. 42

PTM post-translational modification. 27, 35

px pixel. 98–100

radial MSD MSD of radial distances. 50, 51, 55, 56, 58, 69, 78, 80, 91, 106
RAG recombination activating gene. 25
RBS Roberts/SC phocomelia syndrome. 25
RNA ribonucleic acid. 11, 13, 26, 40, 41, 48, 111, 149, VII, VIII

SA stromalin antigen. 26–29, 34–36, 38
SAW self-avoiding walk. 44
SBR signal-to-background ratio. 100
scHi-C single-cell Hi-C. 16, 25, 39, 45
scsHiC sister-chromatid-sensitive Hi-C. 24, 38
scSPRITE single-cell SPRITE. 16
SGO1 shugoshin 1. 29
SIM structured illumination microscopy. 17
SMC structural maintenance of chromosome. 25–33, 37, 38, 86, 87
SMLM single-molecule localization microscopy. 17, 39, 41–43, 45, 86
SNR signal-to-noise ratio. 40–43, 65, 66, 96, 104, 148
SPRITE split-pool recognition of interactions by tag extension. 16, IX
SPT single-particle tracking. 1, 3, 40, 46, 87–89, 94, 95, 103, 104, 112, XI
STED stimulated emission depletion. 17
STORM stochastic optical reconstruction microscopy. 17, 22, 42
SUMO small ubiquitin-like modifier. 9
SV40 simian virus 40. 6, 13
SWI/SNF switch/sucrose non-fermentable. 9

TAD topologically associating domain. 1, 20, 22–25, 29–31, 34, 35, 37, 45, 46, 48, 49, 55–57, 59, 60, 63, 64, 80, 81, 86, 106, 108–110, 113, 151, X
TALE transcription activator-like effectors. 40, 41
TetO Tet operator. 40, 46, 49–52, 54–60, 64–69, 73–75, 78–81, 87, 88, 90, 93, 96–98, 100, 101, 104, 145–150, 152, 153, 161, 164
TetR Tet repressor. 40, 49, 50, 56–59, 64–66, 69, 79, 87, 96, 98, 104, 145, 146, 148, 151, 161, 164
TF transcription factor. 5–13, 19, 41, 107, 108, 110, 113, IX
TFBS TF binding site. 6, 8–10, 110, 111
TSS transcriptional start site. 5, 6, 11

V(D)J variable, diversity and joining. 25

WAPL wings apart-like protein. 23, 24, 27–29, 34, 36–38, 46, 104
WT wild-type. 25, 36, 78, 89–91, 93, 94, 152, 153

Xist X-inactive-specific transcript. 13

ZF zinc finger. 33

List of Figures

| | | |
|----------|---|-----------|
| 1 | Introduction | 5 |
| 1.1 | Variables of enhancer-promoter communication | 7 |
| 1.2 | Models of enhancer-promoter communication | 12 |
| 1.3 | Models of information transfer between enhancer and promoter | 14 |
| 1.4 | Relationship between 3C contacts and spatial distances in DNA-FISH experiments | 19 |
| 1.5 | Levels of higher order chromosome organization | 21 |
| 1.6 | TADs and CTCF loops upon depletion of CTCF, cohesin or its regulators | 24 |
| 1.7 | Structure of the cohesin complex in comparison to RNA Pol II and Mediator | 26 |
| 1.8 | The mechanism of DNA loop extrusion | 30 |
| 1.9 | A partial structure of the loop-extruding cohesin complex | 32 |
| 1.10 | The molecular basis of the “CTCF convergence rule” | 34 |
| | | |
| 2 | Cohesin and CTCF control the dynamics of chromosome folding | 47 |
| 2.1 | Cohesin slows down chromosome dynamics in living cells | 50 |
| 2.2 | Loop extrusion generally slows down polymer motion | 53 |
| 2.3 | Convergent CTCF sites further constrain polymer dynamics | 55 |
| 2.4 | Cohesin and CTCF reduce variability in chromosome folding dynamics | 57 |
| 2.5 | Cohesin and CTCF control contact dynamics inside a TAD | 60 |
| 2.6 | Estimation of frequency and duration of cohesin-mediated CTCF loops | 62 |
| | Extended Data | 72 |
| E1 | Chromosome structure is altered upon degradation of factors involved in loop extrusion | 72 |
| E2 | Chromosome dynamics is modulated by degradation of factors involved in loop extrusion | 74 |
| E3 | Simulations of chromosome dynamics and effects of loop extrusion | 76 |
| E4 | MSDs of systems for two extruder speeds | 77 |
| E5 | Characterization of TetO and LacO array integrations | 78 |
| E6 | Correction of chromatic aberrations and characterization of mESC lines with promoters flanking TetO and LacO arrays | 79 |
| E7 | Polymer simulations of two genomic locations within the same TAD | 80 |
| E8 | Live-cell imaging of two genomic locations within the same TAD | 81 |

| | | |
|----------|--|------------|
| E9 | HMM analysis of simulations compared to experimental data | 82 |
| E10 | Polymer simulations of landscapes with two barriers at different distances | 83 |
| 3 | Cohesin dynamics on DNA: Cohesion or Extrusion? | 85 |
| 3.1 | SPT of cohesin and DNA in the same cell to observe signatures of loop extrusion | 88 |
| 3.2 | Denoising allows for extended imaging of cohesin molecules | 90 |
| 3.3 | Set-up allows for accurate observation of dynamics of DNA loci | 91 |
| 3.4 | Cohesion and loop extrusion contribute to constraining chromosome motion | 92 |
| 3.5 | Cohesin moves slowly on DNA | 93 |
| 3.6 | Increasing the velocity of cohesin on DNA: PDS5 and partial RAD21 depletion | 95 |
| 4 | Discussion | 103 |
| 4.1 | Temporal dynamics of E-P communication | 111 |
| | Supplementary Information | 145 |
| S1 | rMSD of two beads separated by 100 other beads in Rouse chain | 158 |
| S2 | rMSD of two beads separated by 100 other beads in Rouse chain with excluded volume | 159 |
| S3 | Features of HMM states called on simulations across the parameter space | 160 |
| S4 | Genotypes of degron cell lines used for SPT | 164 |
| S5 | Overview on the image analysis pipeline | 165 |

List of Tables

| | | |
|----------|--|------------|
| 1 | Introduction | 5 |
| 1.1 | Cohesin complex composition in loop extrusion or sister-chromatid cohesion | 27 |
| 1.2 | Techniques for visualizing DNA in living cells | 41 |
| 3 | Cohesin dynamics on DNA: Cohesion or Extrusion? | 85 |
| 3.1 | Parameters for spot detection | 99 |
| 3.2 | Parameters used for tracking with LAPtrack | 100 |
| 4 | Discussion | 103 |
| 4.1 | Comparison of studies on CTCF/cohesin-mediated looping dynamics in mESCs | 109 |
| | Supplementary Information | 145 |
| S1 | Statistics of live-cell imaging data | 161 |
| S2 | Statistical tests for duration and frequency of the HMM-called states | 162 |
| S3 | Oligonucleotides | 162 |
| S4 | Spot detection and tracking parameters | 162 |

Abstract

In mammalian genomes, physical interactions between chromosomal loci and notably between enhancers and promoters, enable control of gene expression across large genomic distances. However, it remains unknown how frequent and stable contacts between chromosomal sequences are over the course of the cell cycle and how they depend on the underlying dynamics of chromosome folding and on the loop extrusion activity of cohesin.

To address these questions, we engineered mouse embryonic stem cells (mESCs) to carry tens of random integrations of bacterial operator arrays that can be visualized using live-cell fluorescence microscopy. By imaging these in the context of auxin-inducible degron systems for factors such as CCCTC-binding factor (CTCF) and RAD21, a subunit of the cohesin complex, we observed that on a global scale cohesin serves to constrain chromosome motion, while the loss of CTCF does not impact global chromosome dynamics.

However, by imaging two distinguishable chromosomal locations separated by 150 kb in *cis* within a topologically associating domain (TAD), we discovered that their interactions are transient events occurring frequently during the course of a cell cycle and that these interactions become substantially more frequent and longer (~ 16 minutes) in the presence of convergent CTCF sites. By contrast, cohesin depletion results in shorter and less frequent *cis*-contacts. These observations thereby indicate a role for cohesin and CTCF in the suppression of variability in chromosome folding across time. Comparison of the experimental results to physical models of chromosome dynamics additionally suggests that individual CTCF-anchored, cohesin-mediated loops last around 10 minutes. These measurements therefore indicate that long-range transcriptional regulation might rely on transient physical proximity, and show that cohesin and CTCF stabilize otherwise highly dynamic chromosome structures to facilitate selected subsets of chromosomal interactions.

In a complimentary approach, we further investigated what function of cohesin leads to this effect: its role in sister-chromatid cohesion or in loop extrusion. We therefore employed a single-particle tracking (SPT) approach following individual RAD21 molecules bound to DNA at the same time as the motion of DNA loci. We found that cohesin motion on DNA is characterized by a more sub-diffusive behavior than that of DNA itself. Furthermore, by simultaneously depleting regulators of cohesin in loop extrusion (NIPBL) or sister-chromatid cohesion (Sororin), we showed that both functions contribute to imposing constraints on the dynamics of chromosomes.

Outline

The thesis describes both published and un-published work with collaborations and co-authorships indicated at the beginning of each chapter. It consists of four chapters:

Chapter 1 is an introduction to the topic of transcriptional regulation through chromosome architecture with a focus on enhancer-promoter communication in mammals. Sections and the figures have been reproduced or adapted from the following publication:

Integrative approaches to study enhancer–promoter communication

Mach, P., & Giorgetti, L.

Current Opinion in Genetics & Development, 80, 102052. (2023) <https://doi.org/10.1016/j.gde.2023.102052>

The second part of this chapter then introduces different features of chromosome architecture, the dynamics of their folding processes and the methods to study them. It further gives an introduction on the cohesin complex and its roles in regulating chromosome folding, namely by DNA loop extrusion. It then describes the questions and the scope addressed in the experimental chapters of this dissertation.

Chapter 2 and **Chapter 3** summarize the results of this dissertation. **Chapter 2** describes the dynamics of chromosomes inside living cells using live-cell imaging and how these are controlled by the cohesin complex and CTCF. The contents of this chapter are reproduced or adapted from the following publication:

Cohesin and CTCF control the dynamics of chromosome folding

Mach, P.,* Kos, P. I.*, Zhan, Y.*, Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., & Giorgetti, L.

Nature Genetics, 54(12), 1907–1918. (2022) <https://doi.org/10.1038/s41588-022-01232-7>

*contributed equally

Chapter 3 focuses on a characterization of the motion of DNA and of the cohesin complex itself when bound to DNA and upon perturbation of its functions (DNA loop extrusion or sister-chromatid cohesion) employing a SPT approach. The results of this chapter have not been published.

Chapter 4 is a discussion of the results described in **Chapters 2** and **3** placing the results in the context of the current understanding in the field. Some parts of this chapter have been adapted from the following publication:

Integrative approaches to study enhancer–promoter communication

Mach, P., & Giorgetti, L.

Current Opinion in Genetics & Development, 80, 102052. (2023) <https://doi.org/10.1016/j.gde.2023.102052>

The references for all chapters are summarized in a final **References** chapter. The **Supplementary Information** contain all further data related to **Chapter 2** and **3**. The above-mentioned manuscripts in its published form as well as other manuscript published during my PhD studies can be found in the **Appendix**.

Chapter 1

Introduction

1.1 Transcriptional regulation in the mammalian genome

In multicellular organisms, one deoxyribonucleic acid (DNA) sequence present in all cells serves as the building plan for the entire organism, which means that humans approximately consist of 10^{13} cells with over 200 distinct cell types (Bianconi *et al.*, 2013). During development, distinct gene expression patterns therefore ensure correct differentiation into cell types. However, this complex regulatory system is not encoded by the number of genes present in the DNA, as the number of genes does not vary considerably with approximately 20,000 genes across species, from *C. elegans* (C. elegans Sequencing Consortium, 1998) over mice to humans (International Human Genome Sequencing Consortium, 2004; Breschi *et al.*, 2017). What varies is the genome size and thereby the complexity of regulatory mechanisms that modulate transcriptional changes of developmental genes in a temporal and tissue-specific manner.

In mammalian cells, multiple processes that regulate transcription have been identified: Transcription relies not only on DNA sequences proximal to the transcriptional start site (TSS), called promoters, but also on distal *cis*-regulatory elements known as enhancers. Promoters as well as enhancers are bound by transcription factors (TFs), which ultimately lead to recruitment of RNA Polymerase II (PolII) and initiation of transcription. However, this process can be modulated by other factors such as epigenetic marks, chromatin re-modelers and cofactors, other *cis*-regulatory elements, such as insulators, and the three-dimensional (3D) organization of the genome. Misregulation in any of these processes is directly related to changes in transcriptional activation at the promoter which in turn can lead to disease-related phenotypes, such as craniofacial malformations during development (Minoux *et al.*, 2017; Long *et al.*, 2020) or tumorigenesis (Flavahan *et al.*, 2016; Cho *et al.*, 2018).

It is therefore instrumental for understanding mammalian gene expression patterns to identify and characterize such *cis*-regulatory elements and the mechanisms that give rise to quantitative changes during transcriptional regulation.

1. INTRODUCTION

1.1.1 Enhancers and their role in transcriptional regulation

Whilst promoters had already been identified, the idea of distal *cis*-regulatory elements that regulate transcription was first explored in Banerji *et al.*, 1981. They showed that DNA sequences, in this case a non-coding region of the simian virus 40 (SV40) genome, distal to the TSS or even 3' to the coding sequence within an episomal reporter vector could enhance transcription (Banerji *et al.*, 1981). Importantly, these so-called “enhancers” were able to modulate transcription independent of their orientation. Soon after, enhancers were also found to act within the mammalian genome (Hanahan, 1985) and in a cell-type specific manner (Banerji *et al.*, 1983; Gillies *et al.*, 1983; Mercola *et al.*, 1983). To date, one of the most studied examples in the mammalian genome is the β -globin locus control region (LCR) which controls the expression of a set of globin genes during development (Van der Ploeg *et al.*, 1980; Kioussis *et al.*, 1983; Driscoll *et al.*, 1989; Levings & Bungert, 2002). Studies at the LCR identified important enhancer features, such as their sensitivity to deoxyribonuclease I (DNase I) (Tuan *et al.*, 1985), indicating a depletion of nucleosomes, and their prevalence for TF binding sites (TFBSs) (Philipsen *et al.*, 1990; Talbot *et al.*, 1990). They also showed that enhancers are highly conserved across evolution (Margot *et al.*, 1989; Moon & Ley, 1990; Reitman & Felsenfeld, 1990; Q. Li *et al.*, 1991).

However, it remained technically challenging to identify enhancers in the genome (Loots *et al.*, 2000) until the advent of next-generation sequencing (NGS) techniques enabled the identification of putative enhancers based on the following characteristics: Whole-genome sequencing was used to search for conserved *cis*-regulatory elements across species and known TFBSs (Hardison *et al.*, 1997; Dermitzakis *et al.*, 2002; Pennacchio *et al.*, 2006); chromatin accessibility was assessed genome-wide by DNase I hypersensitivity (Thurman *et al.*, 2012); chromatin immunoprecipitation followed by sequencing (ChIP-seq) enabled mapping of sites bound by TFs (Visel *et al.*, 2009) and certain histone marks associated with active transcription (Heintzman *et al.*, 2007); bisulfite sequencing allowed to investigate DNA methylation genome-wide (Lister *et al.*, 2009) and finally whole transcriptome sequencing showed that enhancers themselves are transcribed into enhancer RNAs (eRNAs) (Andersson *et al.*, 2014).

To date, these descriptors are used to identify and annotate putative enhancers sequences (ENCODE Consortium, 2012). So far, approximately 1 million sequences were identified indicating that up to 16 % of the human genome potentially functions as enhancers (Gasperini *et al.*, 2020). The regulatory complexity conferred by *cis*-regulatory elements can in turn explain the discrepancy between organismal complexity and the number of genes encoded: Greater complexity in mammalian development would hence be the consequence of greater *cis*-regulatory complexity encoded in great part in enhancers (Dermitzakis *et al.*, 2002). However, for most of these putative enhancer sequences, it is not known what target genes they act upon, since they can be located more than 1 megabase (Mb) from their respective target gene and separated from it by other intervening regulatory elements and other potential target genes. If target genes are identified, then it is often not known in what specific context (what developmental time, what cell type) they modulate the transcription at the promoter.

1.1. TRANSCRIPTIONAL REGULATION IN THE MAMMALIAN GENOME

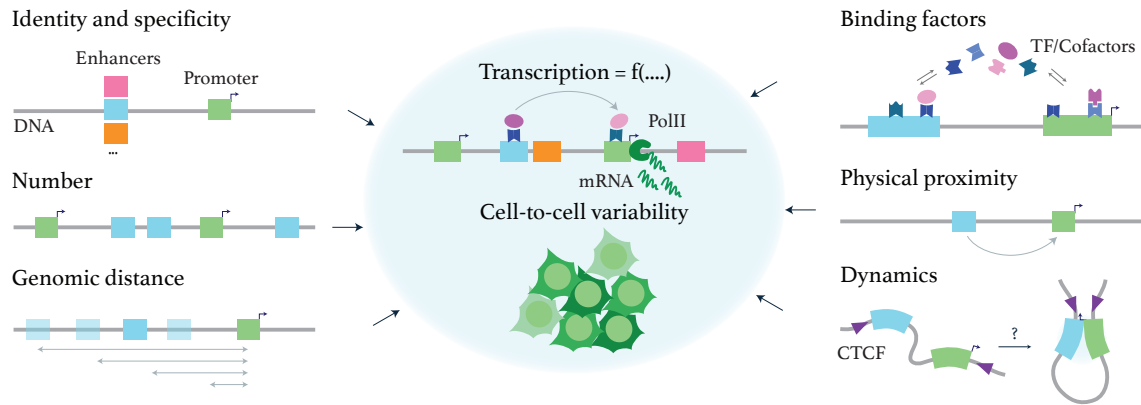


Figure 1.1: Enhancer-promoter communication and transcriptional outputs depend on a number of variables including enhancer and promoter sequence, number, mutual genomic distance, in a way that depends on binding of transcription factors and interactions with cofactors, the physical proximity between regulatory sequences and its dynamics (TF=transcription factor, PolII=RNA Polymerase II). *Fig. adapted from Mach and Giorgetti, 2023.*

New approaches that test endogenous enhancer function are therefore key to dissecting the mechanism by which enhancers function *in vivo* and to revealing further key features and their effects on transcription. For example, knockout studies in mice showed that enhancer redundancy provides a means to overcome loss-of-function mutations in single enhancers (Osterwalder *et al.*, 2018). The key challenge of identifying the target gene of putative enhancers can now be addressed with chromosome conformation capture (3C) techniques (Dixon *et al.*, 2015), specifically capturing promoter sequences and mapping interacting distant sequences (Mifsud *et al.*, 2015). Fig. 1.1 summarizes characteristics that are known to modulate transcriptional levels (Mach & Giorgetti, 2023). However, there is little insight into the interplay between these factors and underlying molecular mechanisms that could solve the *cis*-regulatory “code”. In the following, I will give a brief overview of possible mechanisms on how some of these features might function and the techniques used to study them, with a focus on mechanisms that involve the 3D conformation of the genome and the dynamics of it.

1.1.2 TF binding in a nucleosomal context

TFs are DNA-binding proteins that bind evolutionarily conserved sequence motifs of 6-12 base pair (bp). In humans roughly 1,600 TFs have been identified so far (Lambert *et al.*, 2018) with different tendencies to bind their target sequence dependent on the genomic context (J. Wang *et al.*, 2012): Most TFs will only bind a subset of their binding motifs present in the genome depending on their binding affinity which in vertebrates is repressed by DNA methylation at 5'-C-phosphate-G-3' (CpG) dinucleotides (Bird *et al.*, 1985; Isbel *et al.*, 2022). TF binding is further modulated by chromatin accessibility (as measured by DNase-seq or an assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Boyle *et al.*, 2008; Thurman *et al.*, 2012)). Only very few TFs have the capability to bind DNA independent of the underlying chromatin state, i. e. independent of the first regulatory layer of DNA folding: the nucleosomes binding to DNA which wrap approx. 147 bp of DNA around a core histone octamer complex to package it (Luger *et al.*, 1997) (see 1.2.3). TFs with

1. INTRODUCTION

such a capability are called pioneer TFs and can evict nucleosomes from their binding sites and thereby enable binding of other TFs (Zaret, 2020). The resulting change in transcriptional regulation can potentially explain changes in cell fate decisions. However, an alternative, yet non-exclusive model, for the establishment of cell fate decisions by TFs is that TFs cooperate to out-compete nucleosomal binding and thereby open up chromatin (Mirny, 2010; Jolma *et al.*, 2015; Sönmezer *et al.*, 2021). It is currently not well understood how this cooperativity is mediated, be it via protein-protein interactions, DNA-binding of the TF or via nucleosome eviction.

The use of neural nets that learn rules of TF cooperativity from training data sets such as TF binding data (Avsec *et al.*, 2021) or chromatin accessibility (Minnoye *et al.*, 2020; D. S. Kim *et al.*, 2021) has identified a wide range of modes of cooperativity, which is often restricted to the immediate vicinity of the TF along the DNA sequence spanning approximately 150 bp, the size of the nucleosomal DNA. These models are however agnostic to the underlying mechanism that mediates the cooperativity, e.g. a cooperativity between TF-binding that relies on a periodicity of the DNA-helix of approximately 10.5 bp maybe due to the DNA-binding nature of the TF or due to protein-protein interactions.

A more mechanistic understanding of TF cooperativity has recently emerged from studies measuring TF and nucleosome occupancy at single-molecule resolution using DNA methyltransferase footprinting followed by high-throughput sequencing: These revealed that nucleosome eviction following TF binding to enhancer DNA often results in indirect binding cooperativity (e. g. not mediated by direct protein-protein interactions) (S. Rao *et al.*, 2021; Sönmezer *et al.*, 2021). The single-molecule resolution of such assays was also key to demonstrating that CpG methylation within enhancer regions directly prevents binding of at least some TFs (Kreibich *et al.*, 2023). Coupled to long-read DNA sequencing, single-molecule footprinting methods hold great promise for revealing correlations between TFBS occupancy at distal promoter and enhancer sites (Stergachis *et al.*, 2020) and downstream transcription events (Krebs *et al.*, 2017), and even inferring their temporal order (Battaglia *et al.*, 2022).

1.1.3 Interactions of TFs with cofactors

Besides TF binding patterns, the activity of cofactors, such as the Mediator complex, nucleosome remodellers, histone modifiers and DNA methyltransferases, affect the binding activity of TFs to DNA, which ultimately changes transcriptional output at promoters. While great strides have been made in the identification and characterization of TFs, less is known about how transcriptional co-activators or -repressors modify transcriptional output by modulating TF binding and mediating information exchange between enhancers and promoters. Nevertheless, some are known to be essential for transcription at most promoters, e. g. the Mediator complex serves to phosphorylate and activate the pre-initiation complex (PIC) (Richter *et al.*, 2022). Other activating cofactors can have varying contributions, such as nucleosome remodellers that maintain DNA accessibility (Iurlaro *et al.*, 2021; Hendy *et al.*, 2022) and histone modifiers depositing active marks at enhancers and promoters (i. e. histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 acetylation (H3K27ac)). Many co-repressors have also been described, such as histone modifiers placing marks like histone H3

lysine 27 trimethylation (H3K27me3) and histone H3 lysine 9 trimethylation (H3K9me3) or removing histone marks associated with an active state (B. Li *et al.*, 2007).

Since the aforementioned inputs work in concert and might interact in a yet unknown manner to modulate the quantitative levels of transcription initiation at the promoter, defining the transcriptional levels of an enhancer-promoter (E-P) pair remains challenging. Certain specificities could also be achieved through a non-enzymatic function of histone modifiers or remodellers (Dorigi *et al.*, 2017; Rickels *et al.*, 2017; Hunt *et al.*, 2022) and a given TF might interact with multiple cofactors, which in turn could recruit other cofactors (DelRosso *et al.*, 2023). Considering that cofactors have been shown to interact with a multitude of TFs, the question further arises of how specificity for a certain E-P pair is achieved and what the molecular mechanisms that ultimately transfer the information from the enhancer to the promoter are.

Identifying cofactors that mediate the communication between a specific E-P pair in the context of chromatin has proven difficult, because contrary to TFs, co-activators and co-repressors do not bind DNA in a sequence-specific manner and no specific 'interaction code' with TFs has been identified so far (S. Kim & Wysocka, 2023). The last few years have nonetheless seen exciting progress in the identification and characterization of transcriptional cofactors. Genetic screens and new methods for the multiplexed recruitment of cofactors (Tycko *et al.*, 2020) have revealed large numbers of protein domains with either activating or repressive effects on transcription. The large-scale, unbiased design of these assays has for example enabled the discovery of previously unknown amino acid compositional biases in protein fragments that show co-activator activity and a distinctive role of post-translational SUMOylation of transcriptional repressors (DelRosso *et al.*, 2023). Another step ahead in the study of cofactors is the recent development of inducible degradation methods and the availability of small-molecule inhibitors, which complement gene knock-out approaches by making it possible to study the acute effects of temporary depletion of essential cofactors (de Wit & Nora, 2023). The degradation of Mediator subunits was key to characterizing their cell-type-specific roles in transcriptional regulation (Jaeger *et al.*, 2020) and a role in facilitating E-P interactions (Haarhuis *et al.*, 2022; Ramasamy *et al.*, 2023). Interestingly, depletion and chemical inhibition have shown that chromatin remodeler cofactors, such as the SWI/SNF complex, not only regulate promoter expression by modulating TFBS accessibility in a TF-specific manner (Barisic *et al.*, 2019; Martin *et al.*, 2023), but can also regulate chromosome interactions by modulating accessibility of CTCF binding sites (Barisic *et al.*, 2019).

Not only new approaches such as degron technologies or activator bypass assays, but also potentially methods such as proximity-mediated ligation (BioID) (Göös *et al.*, 2022) that are able to capture more transient interactions, will help to characterize functional interactions between DNA, TFs and cofactors *in vivo*. Complimentary to these, a recent study reconstituted chromatin *in vitro* from its individual components to understand the causal relationships between nucleosome and TF binding (Quililan *et al.*, 2023).

1. INTRODUCTION

1.1.4 Modulation of transcriptional output by enhancer sequence

While it has been described how some of these regulatory factors, i. e. nucleosome remodeling, histone modifications, TF and cofactor binding, can influence an enhancer's activity, it is still poorly understood how these factors work in concert to define an enhancer's activity and especially its specificity for a certain target promoter (i. e. in sum its identity). Moreover, the activity of enhancers will highly depends on other extrinsic variables, such as changes in the levels of TFs present in a given cell type, making the definition of an enhancer's activity also reliant on a cell state.

Recently, massive parallel reporter assays (MPRAs) (Arnold *et al.*, 2013) as well as approaches using mutagenesis or *de novo* synthesis of enhancers (Findlay *et al.*, 2014) have identified the effects of the sequence identity of an enhancer on the quantitative transcriptional output. These assays, combined with new machine learning approaches, have given insights into TFBS usage in enhancer regions and have enabled the design of artificial enhancer sequences that can be employed in a cell-type specific manner (de Almeida *et al.*, 2022; Taskiran *et al.*, 2022).

While such approaches have unlocked the potential to reveal quantitative sequence determinants of intrinsic enhancer activity, it remains unclear if an enhancer can activate any promoter, or rather only a compatible subset of promoters. Answering this fundamental question requires measuring the transcriptional output of large numbers of E-P combinations in parallel. Initial experiments based on transiently expressed combinatorial libraries of enhancer-reporter pairs have suggested that housekeeping and developmentally regulated promoters are activated by distinct sets of enhancers in *D. melanogaster* cells (Haberle *et al.*, 2019). However, in mammalian cells a rather broad compatibility was observed, with quantitative rather than qualitative differences in E-P communication (Bergman *et al.*, 2022; Martinez-Ara *et al.*, 2022; Sahu *et al.*, 2022). This is in line with the earlier finding that once randomly inserted in the mouse genome, ectopic promoter sequences are activated by the (unrelated) surrounding endogenous regulatory landscapes (Ruf *et al.*, 2011; E. Anderson *et al.*, 2014).

Moving the field forward will critically depend on the development of large-scale reporter assays allowing the study of E-P communication in an endogenous chromatin environment, possibly taking advantage of experimental setups where genomic sequence can be engineered bottom-up (Blayney *et al.*, 2022; Brosh *et al.*, 2023; Galupa *et al.*, 2023) and chromatin states and/or recruitment of TFs and cofactors can be artificially modulated (Zhu *et al.*, 2018; Alerasool *et al.*, 2022; Neumayr *et al.*, 2022; Policarpi *et al.*, 2022; Mukund *et al.*, 2023; Naqvi *et al.*, 2023; Noviello *et al.*, 2023). Coupled to increasingly more refined 'explainable' artificial intelligence approaches (Novakovsky *et al.*, 2023), such experiments will allow disentangling regulatory layers and unravel quantitative contributions of enhancer/promoter sequences, TFBS usage and chromatin-mediated effects to E-P communication.

Finally, to what extent these results can be extrapolated to E-P communication in the context of the 3D structure of the genome is unclear. Here, transcriptional output at the promoter is not only modulated by a single enhancer, but e. g. other enhancers can confer a regulatory redundancy (Osterwalder *et al.*, 2018). Furthermore, E-P pairs are often separated by large genomic distances of more than 1 Mb, making it difficult to predict how other (possibly unrelated), intervening *cis*-regulatory elements

perturb the communication of a specific E-P pair. This raises the question of how an enhancer can ultimately contact its target promoter and transfer the information required to activate transcription. If the compatibility between enhancers and promoters as found in MPRA is indeed very broad (Bergman *et al.*, 2022; Martinez-Ara *et al.*, 2022; Sahu *et al.*, 2022), specificity could be encoded on a different level, namely in the physical interactions of enhancers and promoters and the dynamics of these folding processes.

1.1.5 Models of E-P communication

In contrast to promoters that are DNA sequences located in the direct vicinity to the TSS and are by itself sufficient to initiate transcription, as they contain binding sites for TFs and PolIII, enhancers are separated by up to 1 Mb from their target gene, often spanning past other *cis*-regulatory elements. The enhancer therefore has to communicate across large linear genomic distances with the target promoter in order to transfer the information necessary to activate transcription. Indeed, it was recently shown that, given the same regulatory landscape, the expression from the promoter decays with the distance between enhancer and promoter non-linearly (Rinzema *et al.*, 2022; Zuin *et al.*, 2022). Since the identification of enhancers, many models have been formulated that try to explain how distal elements could affect transcription at the promoter. In the following, I will review the most prominent ones and the evidence that argues for them, with a focus on the DNA looping model.

The DNA looping model

Chromosomes can be considered polymer chains that are confined to the nuclear space by the lamina. Due to its polymeric nature, DNA in the nucleus undergoes sub-diffusive motion (i. e. motion that is more constrained than normal diffusion). A mechanism that would involve only passive diffusion would therefore make it highly unlikely for two loci separated by a large linear distance (such as E-P pairs separated by hundreds of kb) to meet within the duration of a cell cycle (Lucas *et al.*, 2014). Furthermore, it would be difficult to achieve the same folding patterns with each re-establishment after mitosis, which is why already early on “DNA looping” (through a mechanism that was then termed “DNA reeling” and later on “DNA loop extrusion”) was suggested as a potential “folding-repair mechanism” (Riggs, 1990).

The advantage of this model is that it explains how enhancer and promoter could be brought into physical proximity which would then enable exchange of regulatory information from enhancer to promoter through actual molecular mechanisms, such as protein-protein interactions or DNA-protein interactions (Fig. 1.2a). Riggs proposed that DNA-binding proteins would move along the DNA in an ATP-dependent manner and that folding could only arise from interactions of parts of the DNA fiber in *cis* (Riggs, 1990).

This model emerged as a more popular model amongst others (which are described below) when more evidence for DNA looping between E-P pairs was found *in vivo*: Through the advent of 3C techniques (described in 1.2.1) and “RNA-trap”, the LCR of the β -globin locus was shown to only contact the β -globin genes in the erythroid lineage where β -globins are expressed (Tolhuis *et al.*, 2002).

1. INTRODUCTION

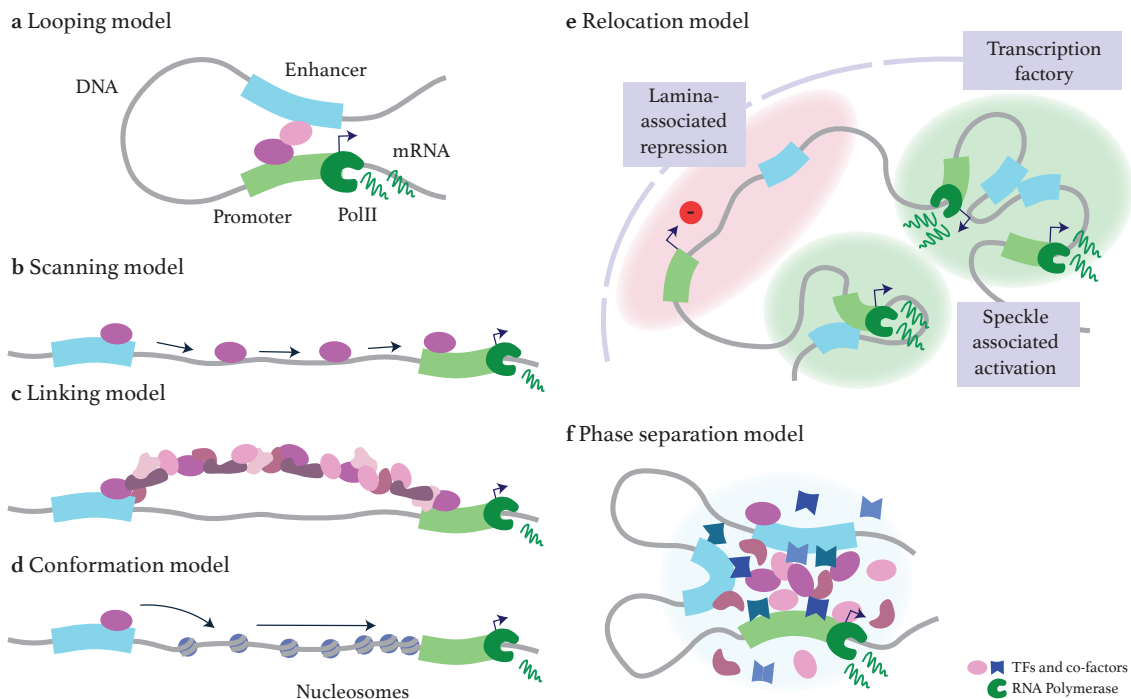


Figure 1.2: Models of enhancer-promoter communication. **a**, The “looping model” proposes that enhancer and promoter will be brought in physical proximity to transfer information. **b**, In the “scanning model”, information from the enhancer is transferred to the promoter by sliding of the bound factors. **c**, The “linking model” suggests that a physical bridge consisting of TFs and cofactors can be built between enhancer and promoter. **d**, The “conformation model” proposes a change in the underlying conformation of chromatin, such as nucleosome positioning, to lead to an information transfer between enhancer and promoter. **e**, The enhancer drives a relocation of the promoter to more active environments, such as near nuclear speckles, in the “relocation model”. **f**, The “phase separation model” suggests that recruitment of TFs and cofactors to the enhancer leads to a local increase in the concentration of TFs and cofactors which in turn will lead to a phase-separated droplet that then also attracts and increases the concentration of TFs and cofactors at the promoter. *Fig. adapted from Popay and Dixon, 2022.*

Soon after, Nasmyth, 2001 proposed condensin as a protein complex that mediates DNA looping in order to explain the resolution of sister-chromatids in mitosis. The cohesin complex was then proposed as a candidate for DNA loop formation in interphase and the idea of “loop extrusion by cohesin” was first put forward (Wendt *et al.*, 2008; Alipour & Marko, 2012; Fudenberg *et al.*, 2016). Since then, loop extrusion by cohesin has been extensively studied both *in vitro* (Davidson *et al.*, 2019) as well as *in vivo* (Nora *et al.*, 2012; S. Rao *et al.*, 2017). I will describe the current state of this model and its implications for genome folding later in this chapter (see 1.3.3).

Alternative E-P communication models

Whilst the “DNA looping model” requires the enhancer to come into close proximity with the promoter, other models have been envisioned where information is transferred through alternative means but physical proximity:

The “scanning model” (also “sliding” or “tracking model”) proposes that an effector bound at the enhancer could slide along the DNA to reach the promoter, where it then activates transcription (Fig.

1.2b). This was first suggested as the means by which the SV40 enhancer acts as a “bidirectional entry site” for PolIII (Moreau *et al.*, 1981) and was also able to explain the functioning of insulator proteins as these would stop the scanning process. Further evidence for enhancer function via this mechanism is the X chromosome inactivation in female mammals. Here, the X-inactive-specific transcript (*Xist*) is transcribed from one X allele and spreads in *cis* along the chromosome to lead to transcriptional inactivation and recruitment of other repressive cofactors (Galupa & Heard, 2015).

In the “linking model”, a protein bridge induced by an oligomerizing scaffold protein would recruit other proteins and would form a contact between the enhancer and the promoter (Bulger & Groudine, 1999) (Fig. 1.2c). It was used to explain the functioning of the *D. melanogaster* protein Chip (and later on also its mammalian homolog LIM domain-binding family protein (Lbd1) (Monahan *et al.*, 2019)) which self-associates into oligomers and is required for enhancer function.

The “conformation model” proposes that a signal at the enhancer is relayed to the promoter via a conformational change to the DNA fiber, possibly by changes to nucleosomal structures or positioning (Dyana & Tjian, 1985) (Fig. 1.2d). More recently, this model has been extended suggesting that e. g. the compaction of repressive chromatin signals, such as H3K27me3 Polycomb sites or H3K9me3 and HP1 heterochromatic regions, can spread and thereby limit accessibility also to surrounding regions (Simon & Kingston, 2013).

More recently, this initial model was extended to encompass the “relocation model” that suggests that enhancers, upon activation, control their positioning within the nucleus thereby bringing their target genes to active chromatin environments (Fig. 1.2e). Indeed, it was shown that the *HoxB* locus relocates to the periphery of its chromosome territory when it is activated during development (Chambeyron & Bickmore, 2004). This model could also explain the association of active genes into what was termed “transcription factories” (Osborne *et al.*, 2004).

This in turn motivated the “phase separation model” which suggests the occurrence of membrane-less organelles in the nucleus consisting of proteins with disordered, but activating domains (and potentially RNA) that interact with enhancers and promoters (Fig. 1.2f). In such a “phase condensate” information transfer from enhancer to promoter would be possible without physical proximity, but would rather be mediated through the increase of activating factors present in the organelle (Sabari *et al.*, 2018; Shrinivas *et al.*, 2019).

Physical proximity

Thus, since the exact molecular mechanisms and factors that mediate E-P communication remain unclear, so too do the physical scales that are the most relevant to study. If activation is mediated by direct TF-cofactor-PolIII interactions, these are expected to take place in the few tens of nanometer range (as a reference, human Mediator/PIC spans $\sim 25 \times 10 \times 10$ nm based on recent cryo electron microscopy (cryo-EM) structures (Rengachari *et al.*, 2021)). If instead communication occurs through macromolecular complexes, phase-separated droplets nucleated by low-complexity protein domains (Hnisz *et al.*, 2017; Chong *et al.*, 2018) or local diffusion of (post-translationally modified) transcrip-

1. INTRODUCTION

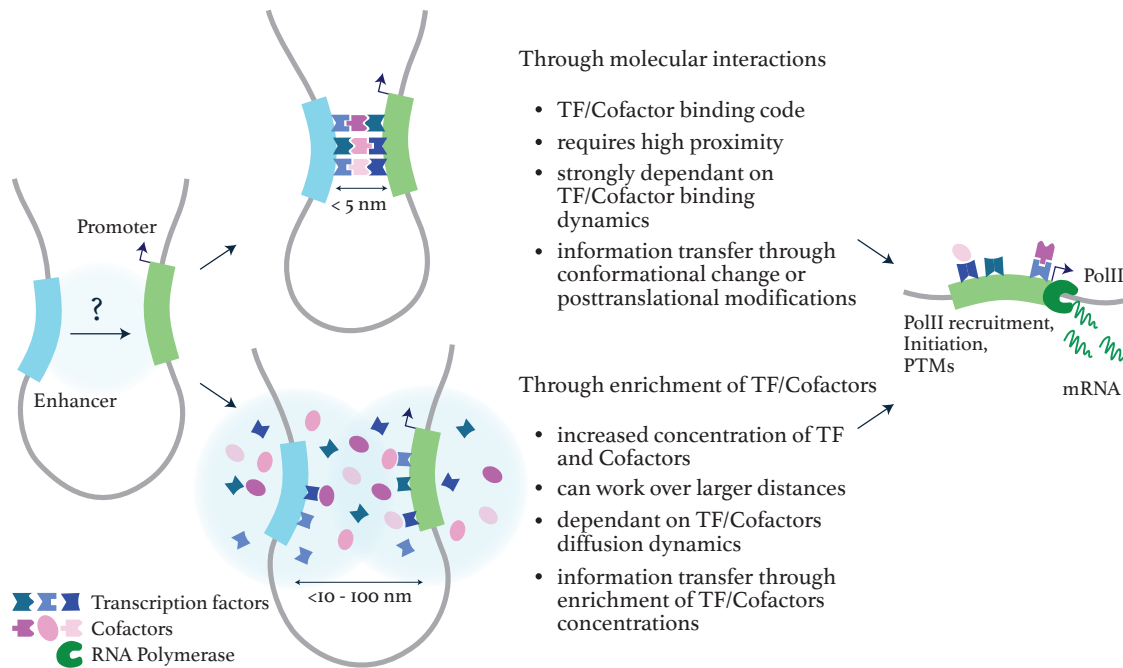


Figure 1.3: Two alternative (and not mutually exclusive) potential mechanisms of information transfer from enhancers to promoters. Information might be passed on to the promoter through direct interactions between transcription factors and effectors; or through local enrichment of transcription factors and cofactors in the vicinity of the promoter. Both mechanisms must result in the transfer of information being used by the promoter to enhance PolII recruitment, post-transcriptional modifications, or events leading to transcriptional initiation or elongation. *Fig. adapted from Mach and Giorgetti, 2023.*

tional co-activators (Karr *et al.*, 2022), it is possible that distances involved are much larger (in the range of hundreds of nm) (Fig. 1.3). Both mechanisms, irrespective of whether they co-occur or are mutually exclusive, must result in transmission of information to the promoter that ultimately leads to PolII licensing into productive transcription (Field & Adelman, 2020).

A better characterization of cofactors will likely enable a more functional, rather than purely physical understanding of E-P contacts. This would allow one to tag cofactors specifically involved in long-range communication with enzymes that promote proximity-mediated chemical modifications (e. g. through proximity-mediated biotinylation followed by mass spectrometry (Göös *et al.*, 2022) or fluorescent complementation assays (Tebo & Gautier, 2019)). This could also account for the temporal aspect of such interactions. One interesting step in this direction was recently provided by proximity-assisted photoactivation (PAPA) (Graham *et al.*, 2022), in which energy transfer between two rhodamine dyes allows the fluorescent detection of tagged molecules separated by a distance of few nanometers at single molecule resolution. Finally, the exciting perspective to visualize physical proximity between genomic locations together with the factors that mediate functional communication might come from spatial sequencing techniques (Nguyen *et al.*, 2020), especially when combined with methods to fluorescently detect chromatin-binding proteins *in situ* (Lu *et al.*, 2022). Recent advances in correlative light and electron microscopy (CLEM) (Müller *et al.*, 2021) also open up the possibility that functional interactions might become observable at nearly atomic resolution in a not-too-distant future.

The notion that proximity is required for conveying regulatory information is also supported by the fact that E-P interaction probabilities measured with cross-linking-based 3C (see 1.2.1) correlate quantitatively with the degree of promoter activity (Kane *et al.*, 2022; Rinzema *et al.*, 2022; Zuin *et al.*, 2022) and that cohesin-mediated CTCF loops (see 1.3.3) can often modulate E-P communication (Huang *et al.*, 2021; Zuin *et al.*, 2022; Chakraborty *et al.*, 2023), although this effect might depend on the actual intrinsic activity of the enhancer (Fulco *et al.*, 2019; Zuin *et al.*, 2022) or compatibility with the promoter.

1.2 Genome architecture and chromosome folding

If the 3D genome can indeed be seen as a “mechanical communication device” (Dekker & Mirny, 2016), then how does it work? How is the genome in mammalian cells organized in 3D and how does this enable communication between loci in *cis*? In the following, I will describe forms of higher order genome folding, the mechanisms that give rise to it and the techniques that are used to study it, with a focus on the sub-megabase scale, where E-P communication is thought to happen.

1.2.1 Methods to study 3D genome architecture

Although the idea of an organization of the genome in 3D was already present more than a century ago (Flemming, 1882), nuclear organization was extensively studied only with the last 20 years, mostly due to the development of a new set of methods called chromosome conformation capture (3C) techniques. Additionally, microscopy methods based on the hybridization of fluorescent, complementary probes to DNA (fluorescence *in situ* hybridization (FISH)), have contributed significantly to our current understanding of the 3D genome. Here, I will review the most prominent of these techniques and will compare their similarities and limitations. For a detailed overview on current methodologies, I will refer the reader to a review by Jerković and Cavalli, 2021.

Chromosome conformation captures techniques

Chromosome conformation capture techniques give an overview of the interactions, a locus in the genome has with other DNA loci. 3C methods rely on cross-linking of nuclei (usually with formaldehyde) and subsequent *in situ* digestion of the DNA with a restriction enzyme. Through proximity-mediated ligation of the ends, hybrid fragments of DNA that were located close in 3D space are created. After de-cross-linking, these fragments can be identified by paired-end NGS (Dekker *et al.*, 2002). While the original 3C technique still relied on PCR for the quantification of contacts and could only be employed to study the interaction between two specific loci, chromosome conformation capture-on-chip (4C) was then developed to assay the interactions of one specific locus with all other loci in the genome (“one vs. many”) (Simonis *et al.*, 2006). Chromosome conformation capture carbon copy (5C) then allowed identification of all pairwise interactions within a region of interest (“many vs. many”) (Dostie *et al.*, 2006) and finally high-throughput chromosome conformation capture (Hi-C)

1. INTRODUCTION

led to the expansion of these set of assays identifying all pairwise interactions genome-wide (“all vs. all”) (Lieberman-Aiden *et al.*, 2009).

These approaches result in two-dimensional (2D) heatmaps that show the contact probability (as defined by the cross-linking radius) for each pair of loci, whereby the resolution, i. e. the bin size of fragment counts, is defined by the cutting frequency of the restriction enzyme used in the experiment. This limitation was recently overcome with the development of micro-C, which uses double-cross-linking and micrococcal nuclease (MNase) digestion instead of restriction enzymes. This leads to a more uniform digestion with fragments the size of the length of nucleosomal DNA thereby allowing to resolve smaller scale structures at the size of the nucleosome (Hsieh *et al.*, 2020; Krietenstein *et al.*, 2020). However, it also further increases the number of reads that need to be sequenced to reach a sufficiently filled heatmap. Since in many cases specific loci are of interest when studying chromosome organization, Hi-C and also micro-C have been further combined with capture approaches: In these, biotinylated baits complementary to the regions of interest are used to enrich the previously generated Hi-C or micro-C library for that specific region (Schoenfelder *et al.*, 2015; Aljahani *et al.*, 2022; Goel *et al.*, 2023). This can either be used to generate a very high resolution saturated map of a specific region or to capture all interactions that e. g. promoter sequences have with other loci in the genome (promoter capture Hi-C (pcHi-C) (Mifsud *et al.*, 2015)).

All of the above-mentioned techniques rely on cross-linking with formaldehyde (or alternatives) to capture the conformation of the 3D genome. However, this might introduce biases due to the nature of cross-linking itself: Transient interactions between loci, e. g. through proteins with a short residence time on DNA might not be captured sufficiently and cross-linking might be biased by the amino acid composition of proteins. Further more, the interaction radius that cross-linking establishes is not easily identified and will vary with the cross-linking and digestion efficiency, but is thought to be in the range of 150-250 nm (L. F. Chen, Lee, *et al.*, 2023). Another drawback of 3C techniques is that due to the readout by short-read paired-end sequencing, only two-way interactions can be identified, although higher order contacts between chromosomal locations are possible.

Many structures initially identified by 3C-based techniques have subsequently been confirmed with alternative approaches that do not rely on cross-linking, such as DamC, which uses DNA methylation to detect chromosomal interactions (Redolfi *et al.*, 2019). Multi-way contacts can be identified with methods such as genome architecture mapping (GAM), where fixed cells are embedded in sucrose, frozen and subsequently cut into thin slices that can then be sequenced (Beagrie *et al.*, 2017), or split-pool recognition of interactions by tag extension (SPRITE) where DNA is cross-linked and fragmented and individual fragments are bar-coded over multiple cycles of splitting and pooling to achieve a unique identifier per fragment (Quinodoz *et al.*, 2018).

These approaches ultimately give an average of all possible contacts across a cell population, as all of these require inputs in the range of million of cells to achieve good complexity of the library. However, some of the techniques were further developed for single cell use, such as single-cell Hi-C (scHi-C) (Flyamer *et al.*, 2017) or single-cell SPRITE (scSPRITE) (Arrastia *et al.*, 2022).

Microscopy-based approaches

While FISH predates the advent of 3C techniques and was previously used to study the 3D conformation of the genome, e. g. for the discovery of chromosome territories (see 1.2.3) (Cremer *et al.*, 1982), it had always been limited by the low throughput. The number of loci that could be studied within a given sample was limited by the number of fluorophores that could be simultaneously visualized by the optical system (at maximum five due to spectral overlap). In recent years, this limitation was overcome with the development of Oligopaint approaches (Beliveau *et al.*, 2012), i. e. oligonucleotide-based FISH probes and protocols, that could be coupled with the use of microfluidics allowing multiple cycles of probe hybridization and washout after imaging. With this approach it was possible to increase the throughput of FISH-based techniques from assaying single loci to assaying entire megabase regions of the genome (Bintu *et al.*, 2018; Nir *et al.*, 2018; Cardozo Gizzi *et al.*, 2019; Mateo *et al.*, 2019). Furthermore, in contrast to 3C-based methods, this approach gives access to actual 3D distances in single cells and their distributions across cells for all assayed loci (Giorgetti & Heard, 2016; Fudenberg & Imakaev, 2017). This showed that indeed conformations of the chromatin fiber vary considerably in single cells, but reproduced contact maps as seen in Hi-C when averaged across the cell population (Bintu *et al.*, 2018).

A second, recent advance in microscopy techniques overcame the issue of the limited resolution in FISH experiments that is due to the Abbé diffraction limit (Abbé, 1873). It states that the minimum resolvable distance between two spots in fluorescence microscopy is limited by the wavelength λ with

$$d = \frac{\lambda}{2 \times NA} \quad (1.1)$$

with NA being the numerical aperture of the optical system (which can reach 1.4-1.6 in modern set-ups). This means that even with the best optical set-up, distances smaller than $\frac{\lambda}{2} \approx 250$ nm (lateral resolution, ≈ 600 nm axial resolution) for green light, cannot be resolved. However, recently developed super-resolution microscopy technologies, namely single-molecule localization microscopy (SMLM) (including stochastic optical reconstruction microscopy (STORM) and photo-activated localization microscopy (PALM) (Betzig *et al.*, 2006; Rust *et al.*, 2006)), structured illumination microscopy (SIM) (Gustafsson *et al.*, 2008) and stimulated emission depletion (STED) microscopy (Klar *et al.*, 2000), have been successfully applied to DNA-FISH experiments achieving a resolution of 20 nm (at best 50 nm in sequential, single-fluorophore DNA-FISH experiments). The development of MINFLUX nanoscopy currently further pushes the resolution limit to the single nanometer range (Balzarotti *et al.*, 2017; Wolff *et al.*, 2023). Whilst it has been used in combination with the DNA-PAINT technology (Ostersehl *et al.*, 2022), it has not yet been applied to trace the conformation of chromatin regions.

Besides light microscopy, chromosome organization has also been extensively studied using electron microscopy (EM) yielding even higher resolutions (Ou *et al.*, 2017). However, EM is limited in that it is currently difficult to identify specific structures as the labeling of loci of interest is not possible. One possible new approach is CLEM which combines the labeling capabilities of light microscopy

1. INTRODUCTION

with the higher resolution of EM (Müller *et al.*, 2021).

One shared characteristic of these techniques is the fixation of the sample, either through cross-linking or cryo-fixation, making it impossible to study chromosome conformation in living cells. Both, FISH as well as 3C techniques, can therefore only provide a snapshot in time of a population average to sample many possible configurations of the chromatin fiber. However, some super-resolution approaches can also be applied to other labeling strategies that can be used in living cells, e. g. to follow specific loci or proteins bound to DNA over time (Nozaki *et al.*, 2017; Deguchi *et al.*, 2023). I refer the reader to the discussion of live-cell imaging approaches in 1.4.1.

1.2.2 “Contacts” in DNA-FISH and 3C methods

What is a “contact” within the organization of the 3D genome? Since both sets of techniques are used to study this question, it is important to remember that these do not necessarily measure the same features: While 3C methods quantify “contacts” that happen at the molecular scale and within the cross-linking and ligation radius, FISH methods give an overview of distributions of spatial distances of a pair of loci. In 3C, thousands of cells are sampled, while throughput limits the amount of cells in FISH to hundreds at best. This leads to a sparse sampling of the distance distribution. FISH can address questions in cell-to-cell variability and is free from binning issues that arise in Hi-C (Fig. 1.4a).

In theory, the relationship of normalized 3C counts between two loci a and b ($3C_{ab}$) and the distribution of 3D distances in FISH $P(r_{ab})$ can be described by the following equation (Rosa *et al.*, 2010), since $3C_{ab}$ is proportional to the number of cells, in which a and b were closer in space than the cross-linking radius R :

$$3C_{ab} \approx \alpha \times \int_{r_{min}}^R 4\pi r_{ab}^2 P(r_{ab}) dr_{ab} \quad (1.2)$$

with α being a proportionality factor, r_{min} the minimum distance between the two loci that can be achieved due to the steric repulsion of the fiber. In brief, what 3C measures is the count of the cumulative distribution function (CDF) of the spatial distances at the cross-linking radius (Fudenberg & Imakaev, 2017) (Fig. 1.4b). However, since 3C mostly samples the part of distributions with small distances, FISH and 3C data from the same experiment can often not be reconciled. This is largely due to the fact that the shape of the underlying distance distributions for two sets of loci (e. g. when comparing between two genetic perturbations) might not be comparable, meaning that the 3C count is not easily compared with the mean/median distance in FISH experiments (Giorgetti & Heard, 2016) (Fig. 1.4c). Indeed, using polymer simulations with dynamic looping, it was shown that loops could lead to an increase in 3C counts, while retaining the same median spatial distances in FISH experiments (Fudenberg & Imakaev, 2017).

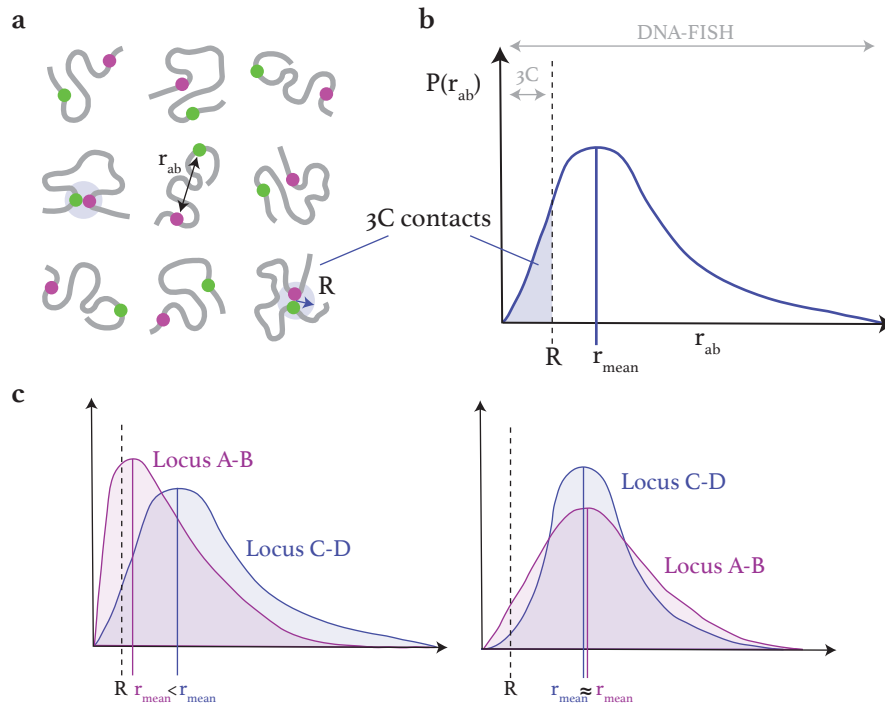


Figure 1.4: Relationship between 3C contacts and spatial distances in DNA-FISH experiments. **a**, Ensemble of configurations that occurs in different cells. The spatial distance r_{ab} between two loci a and b can be read out by DNA-FISH for all of these configurations, however only when a and b are at distances closer than the cross-linking radius R , they will be detected as a contact in 3C methods. **b**, Probability $P(r_{ab})$ distribution for the spatial distances of the two loci a and b . Note that only configurations with distances $r_{ab} < R$ will be sampled in 3C methods (shaded area). **c**, 3C methods and DNA-FISH data only correlate in the case where the underlying distance distribution is comparable for contacts that shall be compared: A wider distribution with the same mean spatial distance r_{mean} will lead to a higher count in contacts in 3C. Fig. adapted from Giorgetti and Heard, 2016 and Fudenberg and Imakaev, 2017.

1.2.3 Hierarchical organization of the 3D genome

The organization of DNA in the nucleus has to fulfill multiple purposes: On the one hand, the ~ 2 m of DNA must be packed into the nuclear space with a diameter of approximately $10 \mu\text{m}$, while certain DNA sequences need to remain accessible for active transcription, DNA replication or repair. Early on, there was evidence that this was realized through a hierarchical organization of specific levels of chromosome organization. While the idea of nuclear organization had persisted since the 19th century (Flemming, 1882), the first evidence of it was only reported in the 1920s by Emil Heitz showing compartmentalization into hetero- and euchromatin using electron microscopy (Heitz, 1928). While nowadays DNA-FISH and variants of Hi-C are the methods of choice to study chromosome organization, historically many discoveries were made using electron microscopy, e. g. identifying nucleosomes as the first important mechanism to fold DNA during interphase (“beads on a string”) (Olins & Olins, 1974). By now, the mechanisms of nucleosome binding, positioning and remodeling have been extensively studied by a multitude of methods in many developmental contexts showing that they mainly function to restrict access to specific DNA sequences in order to regulate binding of TFs (see 1.1.2 and Fig. 1.5a).

1. INTRODUCTION

In this section, I will focus on higher order chromosome structures that can be found at length scales of multiple kilobases up to the organization of the entire genome within the space of the nucleus. With 3C methods, it is possible to investigate all length scales of chromosome organization (from 150 bp to entire chromosomes), which led to the discovery of general principles, such as the scaling of contact probabilities $P(s)$ between any two loci separated by a genomic distance s . $P(s)$ follows a power-law decay with a scaling exponent close to -1 in many species, indicating similar genome organizations throughout the tree of life (Lieberman-Aiden *et al.*, 2009; Sexton *et al.*, 2012; Mizuguchi *et al.*, 2014).

However, across length scales chromosome organization is not homogeneous: On the length scales of entire chromosomes, it was shown using FISH methods that single chromosomes do not intermingle with each other but self-associate, thereby forming chromosome territories (Cremer *et al.*, 1982; Cremer & Cremer, 2001) (Fig. 1.5g). However, the relative position of each chromosome within the nucleus can change with each cell cycle and depends on its transcriptional state (Fig. 1.5f). Generally, domains that are rich in expressed genes and are less compact, are found more often in the center of the nucleus, whereas transcriptionally repressed and more compact domains are associated with the periphery of the nucleus (Bickmore, 2013), although alternative cases have been reported (Falk *et al.*, 2019). With the first characterization of these structures in Hi-C, they were shown to exist at length scales of Mbs to the length of the entire chromosome and have been classified by the eigenvector (i. e. the first component of a principle component analysis (PCA)) into two compartments, namely A (active, open chromatin, demarcated with active histone modifications) and B (inactive, associated with a repressed chromatin state) (Lieberman-Aiden *et al.*, 2009). In Hi-C, A and B compartments show a “checkerboard” pattern indicating that each compartment type preferentially interacts with other compartments of the same type (Fig. 1.5e). More recently, with an increase in resolution in techniques such as micro-C and its derivatives, compartment-like interactions have also been found on a much smaller length scale: Focal interactions between enhancers and promoters as well as structures within regulatory sequences that depend on the activation state of the element, were identified when reaching a resolution of 20 bp (Aljahani *et al.*, 2022; Goel *et al.*, 2023), arguing for a mechanism of E-P communication that involves compartmentalization.

On a sub-megabase scale (10s to 100s of Mb), where E-P communication occurs, TADs were found (Fig. 1.5d). These are domains that show preferential interaction of loci located within them as compared to loci across a TAD boundary (Dixon *et al.*, 2012; Nora *et al.*, 2012; Sexton *et al.*, 2012). Enhancers are often located within the same TAD as their target promoter (Symmons *et al.*, 2014; Spielmann *et al.*, 2018) and E-P regulation only occurs in rare cases across a TAD boundary (Javierre *et al.*, 2016). The mechanisms and functions of TAD formation and the implications for transcriptional regulation are described in more detail in the following section (see 1.2.4). In brief, TADs are thought to arise from nested loop configurations that are mainly established between convergent binding sites for the CTCF through the loop extrusion activity of the cohesin complex (Fig. 1.5b,c). The loop extrusion model and its effects on gene regulation are discussed in 1.3.3.

1.2. GENOME ARCHITECTURE AND CHROMOSOME FOLDING

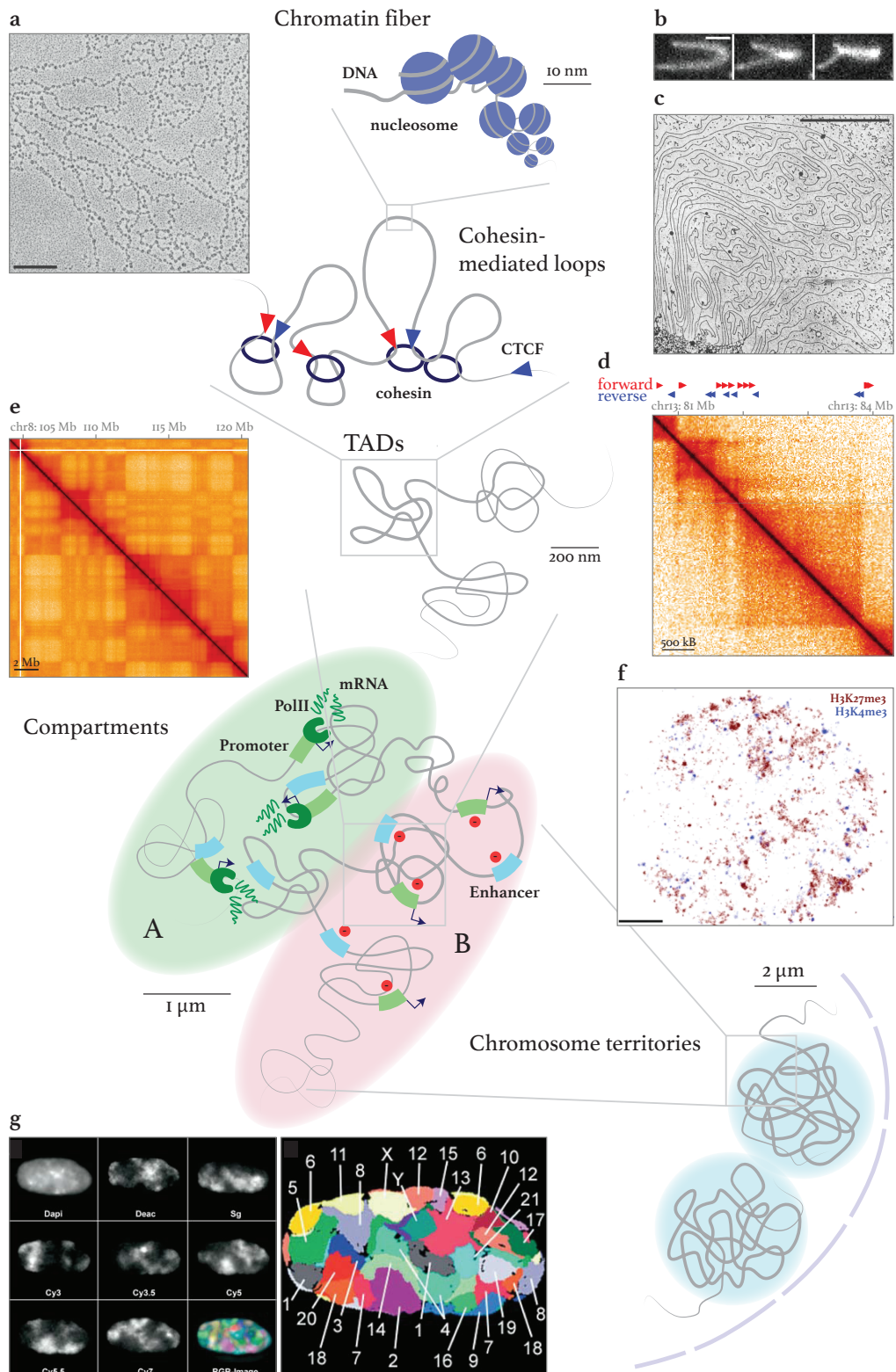


Figure 1.5: Levels of higher order chromosome organization. **a**, Electron micrograph of chromatin from *D. melanogaster* showing nucleosomes seen as "beads on a string" (scale bar 200 nm, image from McKnight and Miller, 1976). **b**, Microscopy images of time series (later time points to the right) showing cohesin extruding DNA *in vitro*. DNA is placed in a lateral buffer flow and components of the cohesin complex as well as regulatory factors and ATP are supplied (scale bar 2 μm , images from Davidson et al., 2019).

1. INTRODUCTION

Figure 1.5: *continued from previous page:* **c**, Electron micrograph of mitotic chromosome showing that in mitosis chromatin is arranged in loops (scale bar 2 μm , *image from Paulson and Laemmli, 1977*). **d**, Hi-C heatmap showing TAD structures whose boundaries coincide with CTCF binding sites. **e**, Hi-C heatmap showing compartments that appear as a “checkerboard” pattern on the megabase scale (*data from Bonev et al., 2017*). **f**, Two-color STORM image showing clusters of histones with active (H3K4me3) or inactive (H3K27me3) histone marks in *D. melanogaster* cells (*images from Cattoni et al., 2017*). **g**, Chromosomes territories in human fibroblasts as shown by DNA-FISH on the left and a model reconstruction on the right (*image from Bolzer et al., 2005*).

However, there are also interactions between specific loci within TADs that are not demarcated by binding of cohesin and CTCF at their boundaries, which are therefore thought to arise or be maintained by independent mechanisms (Batut *et al.*, 2022; Hsieh *et al.*, 2022; L. F. Chen, Long, *et al.*, 2023). These are structures such as E-P loops, sub-TADs and contact domains (Phillips-Cremins *et al.*, 2013; S. S. Rao *et al.*, 2014): While it is not clear yet how these are established and whether their establishment initially requires a loop extrusion activity (Weintraub *et al.*, 2017), contacts between gene bodies remain even in the absence of the interphase loop extruder cohesin pointing towards an alternative mechanism and potentially towards transcription itself for the maintenance of E-P loops (Goel *et al.*, 2023). Since contact domains (including micro-compartments and nano-domains) are often marked by the same active histone marks, compartmentalization through phase separation was proposed as a driver of their formation. A similar mechanism drives heterochromatin formation in *D. melanogaster* and mammals through multi-valent hydrophobic interactions of heterochromatin protein 1 (HP1) (Larson *et al.*, 2017; Strom *et al.*, 2017) and was also proposed to lead to the formation of Polycomb bodies through the positively charged intrinsically disordered domain (IDR) of the Polycomb repressive complex 1 (PRC1) (Tatavosian *et al.*, 2019). The mechanisms that lead to the formation and maintenance of these interactions and cohesin-mediated loops are not exclusive, but might in fact work in concert or depend on each other for proper function.

1.2.4 Topologically associating domains

Topologically associating domains are regions (with a median size of 880 kb) on the sub-megabase scale of genome organization that preferentially show interactions within the domain than with loci on the outside (Dixon *et al.*, 2012; Nora *et al.*, 2012). TADs have been shown to be important for multiple processes on DNA: TAD boundaries correspond to replication domains (Pope *et al.*, 2014; Emerson *et al.*, 2022) as well as units of DNA recombination (Ba *et al.*, 2020) and DNA repair (Arnould *et al.*, 2021) and genes that are co-regulated during development often lie within the same TAD (Nora *et al.*, 2012; Dixon *et al.*, 2016). A comparison of cell types (McArthur & Capra, 2021) and mammalian species showed a high degree of overlap in TAD structures ($\sim 76\%$ of mouse boundaries are present in humans; $\sim 54\%$ of boundaries in human also exist in mice) (Dixon *et al.*, 2012) suggesting a high degree of conservation. TADs were discovered in mammals, but similar structures can also be found in other organisms, such as bacteria (chromosomal interaction domains (CIDs) in *Caulobacter crescentus* (Le *et al.*, 2013)), in yeast (“globules” in *S. pombe* (Mizuguchi *et al.*, 2014)) and in *D. melanogaster* (Sexton *et al.*, 2012). However, the mechanisms that establish TAD-like structures vary across organisms, e. g.

in *D. melanogaster* TADs are correlated more with epigenetic states and do not show enrichment of CTCF or interaction loops at TAD boundaries (Hou *et al.*, 2012).

Although TADs in mammals are a functionally preferred scale of chromosome folding with maximal cell-type conservation and the greatest enrichment of co-regulation of genes within them (Zhan *et al.*, 2017), contact probabilities within TADs are only increased by two-fold as compared to contacts across TAD boundaries (Redolfi *et al.*, 2019; Chang *et al.*, 2020). However, this was recently shown to be sufficient to alternatively regulate E-P pairs within TADs as compared to across TADs due to transcription showing a non-linear response to changes in contact probability (Zuin *et al.*, 2022). TADs can therefore be seen as “regulatory neighborhoods” that serve to confine E-P interactions (da Costa-Nunes & Noordermeer, 2023). Indeed, TAD boundaries are often marked by binding sites for CTCF, which was first identified for its enhancer-blocking activity (Phillips & Corces, 2009). Insulation of enhancers from non-target promoters through TAD boundaries is of great importance during development with structural variations leading to pathologies, e. g. an inversion at the TAD harboring the ephrin type-A receptor 4 (*EPHA4*) gene and its distal enhancer disrupts the TAD structure and thereby inactivates the *EPHA4* gene whilst up-regulating expression at non-target promoters and leading to an abnormal limb development (Lupiáñez *et al.*, 2015). In particular, genes in the *EPHA4* TAD were mis-regulated when the TAD boundary that is demarcated by CTCF was perturbed.

Binding of CTCF to TAD boundaries is also correlated with increased binding of the loop-extruding complex cohesin (Wendt *et al.*, 2008). Approximately 75 to 95% of all TAD boundaries are bound by CTCF (depending on the cell type) (Bonev *et al.*, 2017) pointing towards a role of CTCF and DNA looping by cohesin in the formation of TADs. Indeed, depletion of either CTCF or RAD21, a subunit of the cohesin complex, using an auxin-inducible degron (AID) system, resulted in the loss of TAD structures as seen in Hi-C (Nora *et al.*, 2017; S. Rao *et al.*, 2017). Nipped-B-like protein (SCC2 in mice) (NIPBL), a loading and processivity factor of cohesin, shows a similar phenotype in Hi-C and is therefore also required for TAD formation (Schwarzer *et al.*, 2017). However, upon depletion of the unloading factor of cohesin from DNA, wings apart-like protein (WAPL), TAD boundaries and looping interactions, seen as corner dots in Hi-C, become stronger (Haarhuis *et al.*, 2017). These phenotypes were in good accordance with polymer simulations that reproduced features seen in Hi-C using a bi-directional loop extrusion process (Fudenberg *et al.*, 2016) (for more detail, see 1.3.3 and Fig. 1.6). These perturbations, however, led to a re-consideration of TADs as “regulatory neighborhoods”, since changes in transcription were very mild upon complete loss of TADs via cohesin or CTCF depletion (Nora *et al.*, 2017; S. Rao *et al.*, 2017). Now, it is thought that TAD structures are mainly important for long-range transcriptional regulation (several hundred kb) of certain E-P pairs known to act during specific stages of development or differentiation (Cuartero *et al.*, 2018; Calderon *et al.*, 2022; Rinaldi *et al.*, 2022; Kiefer *et al.*, 2023).

Multiple lines of evidence suggest that in mammals formation of TADs is indeed governed by a distinct mechanism other than segregation of domains based on the activation state as in *Drosophila*: Firstly, TADs can contain multiple chromatin states (Nora *et al.*, 2012; S. S. Rao *et al.*, 2014); secondly, depletion of cohesin may lead to the formation of spontaneous domains, however, these are not

1. INTRODUCTION

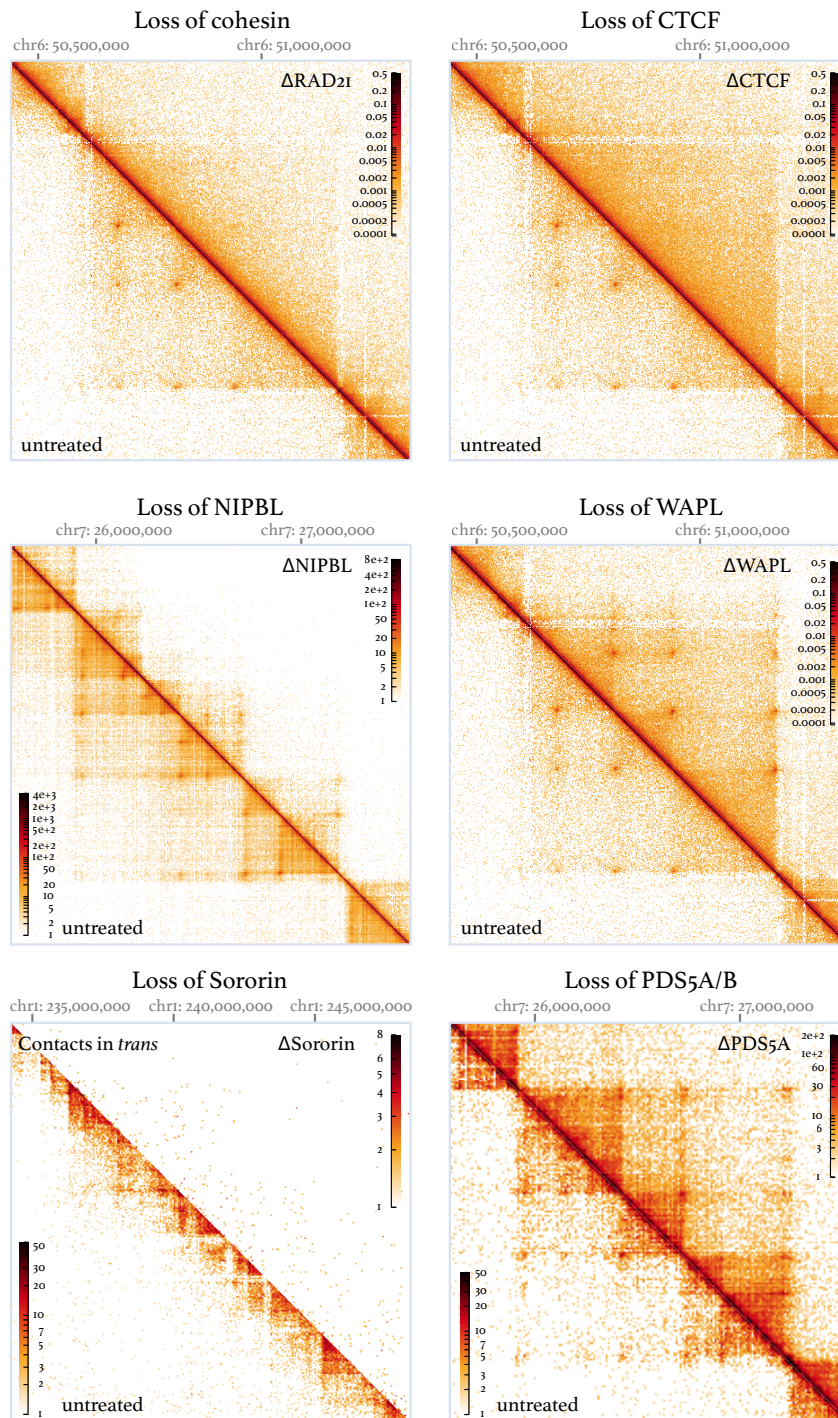


Figure 1.6: TADs and CTCF loops upon depletion of CTCF, cohesin or its regulators. Loss of CTCF leads to loss of boundary structures at TAD boundaries and CTCF binding sites, however increased interaction frequency between distant loci can still be observed due to cohesin's loop activity. Loss of cohesin (Δ RAD21) or NIPBL leads to of TADs and loops, whereas depletion of WAPL or PDS5A/B leads to an increase in loop and boundary strength. *Trans*-contacts are lost upon depletion of Sororin. *Micro-C data from mESC for untreated, Δ CTCF, Δ RAD21 and Δ WAPL from Hsieh et al., 2022, scsHiC data (in HeLa cells) of Δ NIPBL and Δ Sororin from Mitter et al., 2020 and Δ PDS5 (in HAP1 cells) from van Ruiten et al., 2022.*

anchored at precise boundaries, such as at CTCF sites under wild-type (WT) conditions (S. Rao *et al.*, 2017; Schwarzer *et al.*, 2017); thirdly, interactions between active or inactive regions are seen more at the level of compartments, which indeed become stronger upon loss of cohesin (S. Rao *et al.*, 2017); lastly, while TADs and cohesin-mediated loops can be found in zygotic maternal chromatin, there are no compartment interactions (Flyamer *et al.*, 2017). In conclusion, TAD and compartment formation in mammals are two independent and in fact antagonistic mechanisms as proposed by polymer simulations by Nuebler *et al.*, 2018.

Finally, TADs must be interpreted as dynamic structures that only appear in Hi-C contact maps as “statistical frequencies of chromatin interactions within a cell population” (Szabo *et al.*, 2019). ScHi-C, sequential DNA-FISH methods as well as polymer simulations have shown that in single cells conformations of TAD structures and DNA loops can vary significantly and that only the ensemble average of these conformations gives rise to the structures seen in conventional Hi-C (Giorgetti *et al.*, 2014; Flyamer *et al.*, 2017; Bintu *et al.*, 2018). However, the fact that TADs can be seen in ensemble-averaged Hi-C means that these structures occur more frequently than expected by chance, raising the question what the dynamics of TAD formation are.

1.3 Cohesin and the loop extrusion model

Cohesin is one of the mammalian structural maintenance of chromosome (SMC) complexes with its core heterodimer consisting of SMC1 and SMC3 (D. E. Anderson *et al.*, 2002). Together with the other three known SMC complexes, condensin I and II (consisting of SMC2/SMC4) and the SMC5/SMC6 complex, they are involved in a variety of processes that shape mammalian genomes, such DNA repair (Birkenbihl & Subramani, 1992), DNA recombination (Y. Zhang *et al.*, 2022), chromosome condensation and segregation in mitosis (Hirano *et al.*, 1997) and, in the case of cohesin, transcriptional control. SMC complexes exist in all kingdoms of life (Hirano & Mitchison, 1994; Guacci *et al.*, 1997; Michaelis *et al.*, 1997; Losada *et al.*, 1998), suggesting that their functions are ancient, highly conserved processes.

Cohesin itself has multiple functions in mammalian cells: On the one hand, it holds together sister-chromatids after replication until their separation by the mitotic spindle to ensure proper distribution of the sister chromatids onto the daughter cells. On the other hand, it is involved in transcriptional regulation in interphase due to its function in the establishment of TADs through DNA loop extrusion. It is further involved in other DNA-related processes, such as recombination, e. g. by guiding the scanning of immunoglobulin heavy chain (*IGH*) loci by the recombination activating gene (RAG) recombinase during variable, diversity and joining (V(D)J) recombination (C. Guo *et al.*, 2011; Hill *et al.*, 2020). Cohesin is therefore an essential protein complex and even small mutations in cohesin regulators, e. g. NIPBL or establishment of cohesion homolog 2 (ESCO2), will lead to severe pathologies, such as the Cornelia de Lange syndrome (CdLS) or Roberts/SC phocomelia syndrome (RBS), both characterized by severe developmental defects and facial or limb abnormalities (Krantz *et al.*,

1. INTRODUCTION

2004; Vega *et al.*, 2005).

In the following, I will describe the cohesin complex and its role in two main processes, sister-chromatid cohesion and loop extrusion, focusing mainly on its characteristics in mouse or humans. I will further discuss the main regulators of cohesin's function with an emphasis on the impact on loop extrusion dynamics and transcriptional activation.

1.3.1 The cohesin complex

The core unit of the mammalian cohesin consists of the SMC1/3 heterodimer with each chain forming a 50 nm-long anti-parallel coiled-coil domain (D. E. Anderson *et al.*, 2002). The heterodimer is formed through contacts at the hinge domains; the N- and C-terminus of each coiled-coil SMC unit assemble a functioning adenosine 5'-triphosphatase (ATPase) domain. It is connected to the other ATPase domain via the kleisin unit, in the case of cohesin, RAD21 (called SCC1 in mice). Contacts of RAD21 with both SMC units leads to the formation of a ring-like structure (Shi *et al.*, 2020). This trimer is further bound by two HEAT repeated proteins associated with kleisins (HAWK) proteins in two distinct positions at the interface with RAD21; one of the binding sites can be occupied by either stromalin antigen (SA) 1 or 2 (also called STAG1/2) (Sumara *et al.*, 2000) and the other site by either

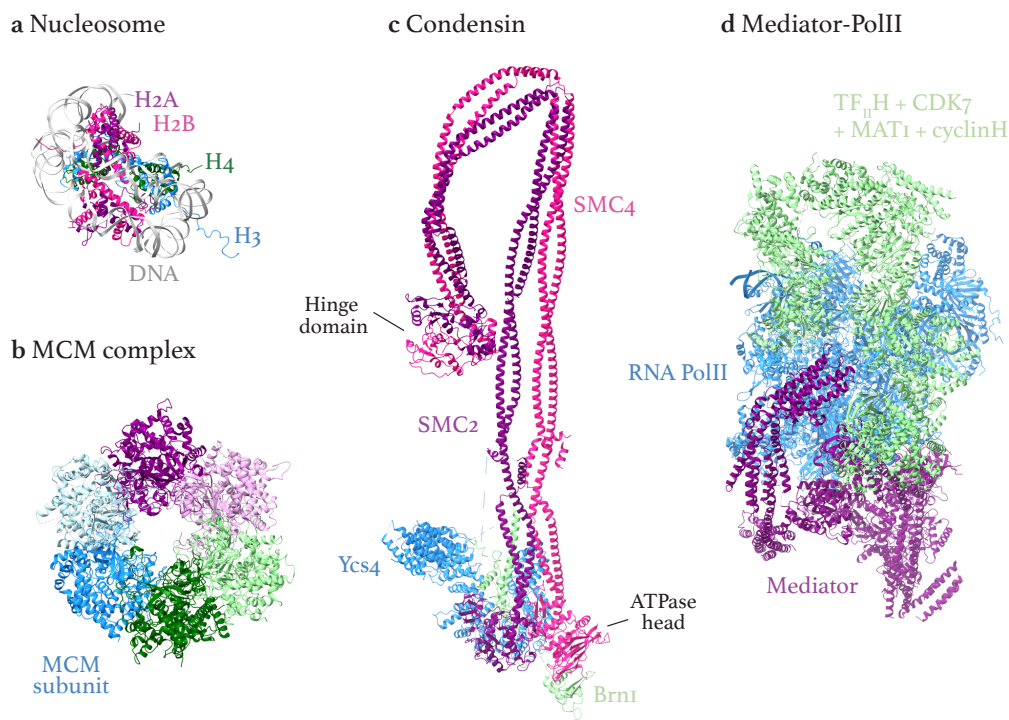
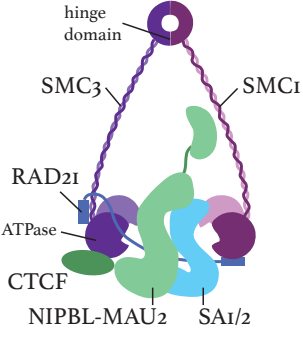
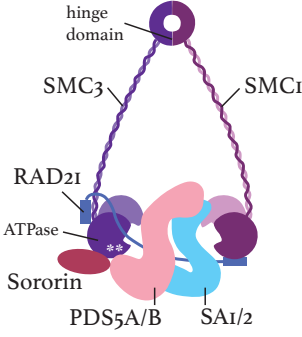


Figure 1.7: Structure of the yeast condensin complex in comparison to the RNA PolIII-PIC complex, the MCM complex and the nucleosome. Condensin is here resolved with the coiled-coil domains; no such structure is available for mammalian cohesin, however it is likely that the mammalian cohesin complex is of similar size as the yeast condensin complex. For a partial structure of the cohesin complex, see Fig. 1.9. All structures to scale. Structures taken from Lee *et al.*, 2020 (condensin, PDB:6YVU), Aibara *et al.*, 2021 (RNA PolIII-PIC, PDB:7NVR), N. Li *et al.*, 2015 (MCM, PDB: 3JA8), Luger *et al.*, 1997 (nucleosome, PDB: 1AOI).

NIPBL or precocious dissociation of sisters 5 protein (PDS5) A or B (called Aprin in mice) (Petela *et al.*, 2018).

Which HAWK protein binds presumably defines the activity of the complex: In its NIPBL bound state, the complex constitutes the loop extrusion holoenzyme, being able to hydrolyze adenosine triphosphate (ATP) with the possibility of recruiting further factors, such as MAU2 (Davidson *et al.*, 2019; Y. Kim *et al.*, 2019). In its PDS5A/B bound state, loop extrusion is stalled and ATP hydrolysis is strongly reduced (Murayama & Uhlmann, 2015; Petela *et al.*, 2018; Davidson *et al.*, 2019). PDS5 recruitment further leads to acetylation of two lysine residues located at the ATPase domain of SMC3 (K105 and K106 in human SMC3) by ESCO1/2 (Rolef Ben-Shahar *et al.*, 2008; Unal *et al.*, 2008) and recruitment of cohesin regulator sororin (Rankin *et al.*, 2005; Schmitz *et al.*, 2007). Sororin binds the complex at the RAD21/SMC3/SA1/2 interface and is responsible for stabilizing the complex on DNA to promote sister-chromatid cohesion (Nishiyama *et al.*, 2010). It is therefore likely that the PDS5A/B-bound cohesin complex stalls loop extrusion and instead promotes sister-chromatid cohesion. The specific function of incorporating either SA1 or SA2 into the complex is less well understood. It was shown that cohesin^{SA1} has a longer residence time (up to several hours) on DNA compared to cohesin^{SA2} (~ 8-20 min) (Kueng *et al.*, 2006; Tedeschi *et al.*, 2013; Wutz *et al.*, 2020). It is therefore possible that cohesin^{SA1} is incorporated into complexes that require stabilization on DNA.

Table 1.1: Cohesin complex composition in loop extrusion or sister-chromatid cohesion. Schemes adapted from Davidson and Peters, 2021.

| Function | “Extrusive” cohesin | “Cohesive” cohesin |
|---|---|---|
| Components |  |  |
| Specific HAWK protein | NIPBL | PDS5 |
| ATPase activity | Yes | No (very much reduced) |
| Residence time | 8-20 min (potentially some cohesin ^{SA1} up to several hours) | Several hours |
| Post-translational modifications | No specific PTM identified | Acetylation of lysine residues in SMC3 ATPase domain |
| Regulators | CTCF, WAPL | Sororin, WAPL, ESCO1/2 |

1. INTRODUCTION

Together, the SMC heterodimer with kleisin unit RAD21 and two HAWK proteins build the complete cohesin complex which then spans ~ 50 nm in height and ~ 40 - 50 nm in width (D. E. Anderson *et al.*, 2002; Shi *et al.*, 2020). Compared to the size of the nucleosome, PolII and transcriptional co-activators such as the Mediator complex (see Fig. 1.7) (Luger *et al.*, 1997; Aibara *et al.*, 2021), its large size allows it to bypass most processes on chromatin and even structures larger than its own size (Pradhan *et al.*, 2022). However, it has been reported that cohesin's translocation on DNA is stalled in some cases by PolII or the minichromosome maintenance (MCM) complex (S. Zhang *et al.*, 2021; Dequeker *et al.*, 2022; Valton *et al.*, 2022; Banigan *et al.*, 2023). It is possible that these effects represent functional interactions with the cohesin complex. Other cofactors that can further interact with the complex to regulate its activity are the aforementioned sister-chromatid cohesion regulator, sororin, but also CTCF and the release factor WAPL. Indeed, these factors do not bind DNA, but they all interact with the same RAD21/SMC3/SA1/2 interface on the complex and their binding is thought to be mutually exclusive (Tedeschi *et al.*, 2013; Hara *et al.*, 2014; Y. Li *et al.*, 2020). The function and interplay of these factors is discussed more in detail in sections 1.3.4 and 1.3.5. Table 1.1 gives an overview of the characteristics of the cohesin complex in its loop-extruding as well as its cohesion performing composition.

1.3.2 SMC complexes during cell cycle progression

Cohesin was first identified for its role in sister-chromatid cohesion where it binds the two DNA strands in *trans* after replication and holds them together until mitosis (Niki *et al.*, 1991; Hirano & Mitchison, 1994). This function is specific to cohesin as it is the only SMC complex known to physically entrap DNA strands in *trans* (Haering *et al.*, 2008). How cohesin is loaded onto DNA to perform cohesion and how this depends on the replication fork, is poorly understood, but two pathways have been proposed based on findings in budding yeast (Xu *et al.*, 2007): The *de novo* pathway argues for *de novo* loading and stabilization of cohesin to DNA which is achieved by NIPBL and ESCO1/2 (Xu *et al.*, 2007). The "conversion pathway" proposes that cohesin that is already bound to DNA can be converted to a "cohesive" complex at sites where converging replication forks terminate (Cameron *et al.*, 2022). Interestingly, it was recently shown that cohesin also topologically entraps DNA at CTCF bound sites (Y. Liu & Dekker, 2022) and that CTCF sites and loop domain boundaries are sites of sister-chromatid cohesion (Mitter *et al.*, 2020). Accordingly, it has been reported that a cohesin mutant with mutations in the hinge domain is unable to perform cohesion and also found less enriched at CTCF sites (Nagasaka *et al.*, 2023). This argues for a model where a conformational change of cohesin (opening of the hinge domain to entrap DNA) can be performed at CTCF sites that will ultimately lead to a transformation of this cohesin complex to a "cohesive" complex. Additionally, it was shown that loop domains determine human replication origins (Emerson *et al.*, 2022) and replication can block loop extrusion (Dequeker *et al.*, 2022), showing that DNA replication, establishment of sister-chromatid cohesion and ongoing loop extrusion are very much connected and the causal relationship between the change in functionality of the cohesin complex is only poorly understood.

With the onset of mitosis, two other SMC complexes, namely condensin I and II, will mitigate the

compaction of chromosomes by loop extrusion. Early on, condensin II will become more stably associated with DNA as its release factor (microencephalin) is inactivated (Houlard *et al.*, 2021). This will lead to the formation of long chromatin loops (Gerlich *et al.*, 2006; Gibcus *et al.*, 2018). Upon nuclear envelope breakdown, the cytoplasmic condensin I will bind DNA and then partition the long chromatin loops into smaller ones of ~ 100 kb size, thereby further condensing the chromosome arms and separating individual chromosomes from each other (Hirota *et al.*, 2004; Ono *et al.*, 2004; Gibcus *et al.*, 2018; Walther *et al.*, 2018). Degradation experiments during mitosis show that the long chromatin loops formed by condensin II are responsible for the axial shortening of the condensed chromosome, while condensin I controls the width of the condensed chromosome fiber (Ono *et al.*, 2003; Shintomi & Hirano, 2011). Since condensins are not focus of this study, I will refer the reader to the review by Hoencamp and Rowland, 2023. Details on the exact mechanism by which loop extrusion is performed by SMC complexes and what are differences in their functioning across complexes is described in 1.3.3.

Cohesion occurs until the onset of prophase where cohesin is unloaded from chromosome arms by the release factor WAPL, after phosphorylation of sororin and the SA2 subunit (known as the prophase pathway) (Losada *et al.*, 1998; Kueng *et al.*, 2006; Nishiyama *et al.*, 2010). Centromeric cohesin remains bound as it is protected by shugoshin 1 (SGO1) that antagonizes WAPL by binding the same interface as sororin or CTCF and preventing phosphorylation by recruitment of protein phosphatase 2A (PP2A) (Salic *et al.*, 2004; Kitajima *et al.*, 2006; Riedel *et al.*, 2006; Hara *et al.*, 2014). Upon attachment to the mitotic spindle in anaphase, the anaphase promoting complex (APC) will degrade securin, thereby allowing separate to cleave the RAD21 subunit of centromeric cohesin (Uhlmann *et al.*, 1999). This will lead to loss of entrapment of the DNA allowing for the correct separation of the sister-chromatids and cytokinesis (Oliveira *et al.*, 2010).

1.3.3 Loop extrusion

That cohesin and condensins might establish TAD structures in interphase and compact chromosomes during mitosis was first explored with the help of theoretical modeling giving rise to what is now commonly referred to as the DNA loop extrusion model (Alipour & Marko, 2012; Fudenberg *et al.*, 2016; Goloborodko *et al.*, 2016). These models proposed that there is a loop-extruding factor (LEF) that would bind to DNA and would “reel” in flanking regions in *cis* through their own motor activity, potentially driven by ATP. This would initially produce small chromosome loops that would increase in size upon progression of the LEF. The interaction between the two DNA loci connected by the LEF would be lost upon dissociation of the LEF from DNA. The original model simulated an one-dimensional lattice that was then used as an input for a 3D polymer simulation where the LEF would extrude symmetrically into both directions until they stochastically dissociate from DNA. The LEFs were assumed to not be able to bypass each other, meaning they would stall upon encountering each other. This could then lead to asymmetric loop extrusion on the unobstructed end. Asymmetric, continued loop extrusion would also happen in the case where the LEF would encounter barrier elements (BEs) (Fig. 1.8a). With these rules, loop extrusion could explain the features seen in Hi-C and

1. INTRODUCTION

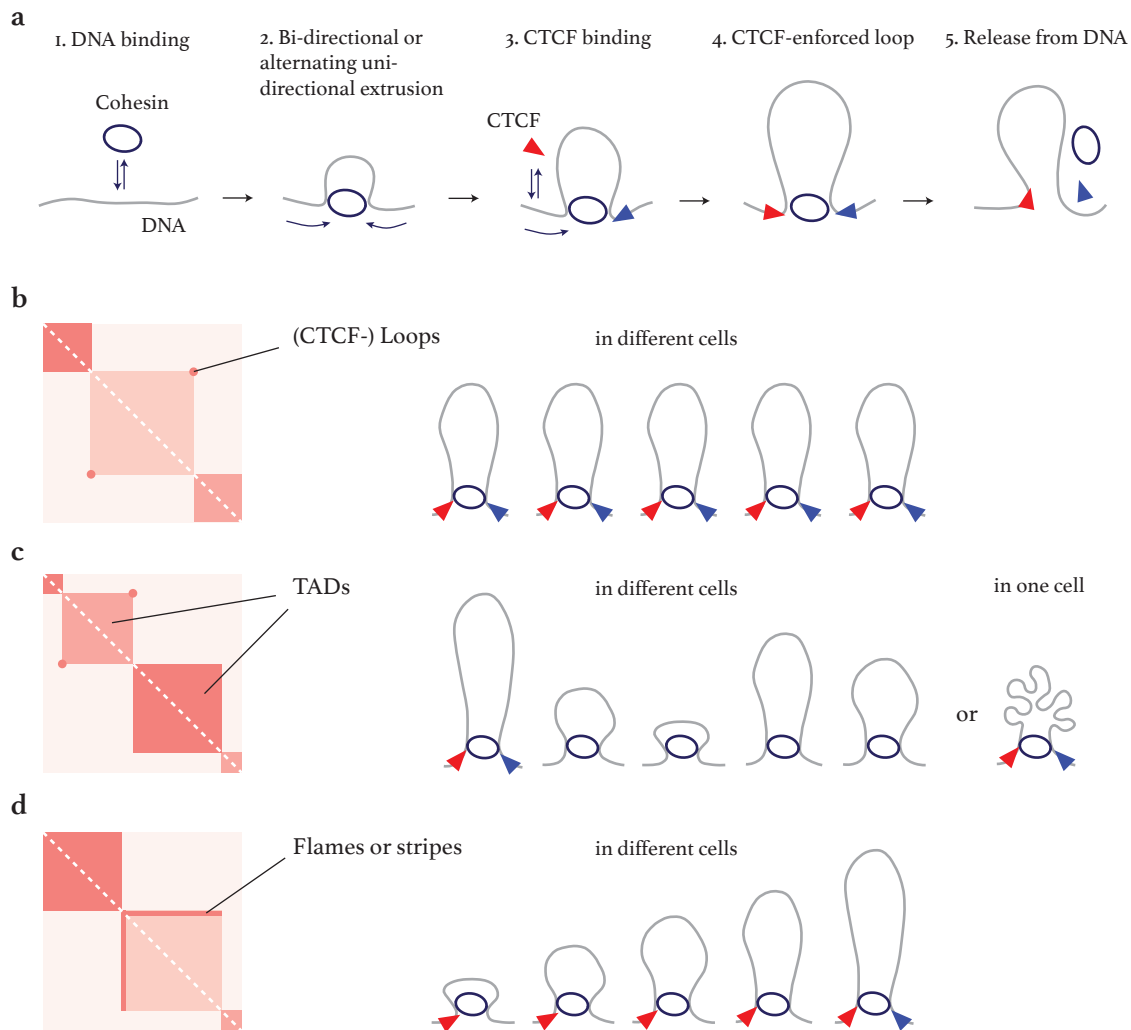


Figure 1.8: The mechanism of DNA loop extrusion. **a**, The cohesin complex binds to DNA and starts to extrude a loop. It does so in a bi-directional manner, either by symmetrical extrusion or by asymmetrical extrusion where the direction of extrusion can be alternated. Upon encountering a CTCF-bound site, it is stalled and a cohesin- and CTCF-mediated loop is enforced. When cohesin or CTCF are released from DNA, the looping interaction is lost. In case CTCF unbinds, cohesin is able to continue to extrude. **b-d**, Structural features in Hi-C and how they arise from looping configurations in single cells. *Schemes in b to d adapted from Davidson and Peters, 2021).*

the formation of TAD boundaries at CTCF binding sites in interphase, when CTCF was considered as an element that could block the LEF (Fudenberg *et al.*, 2016) (Fig. 1.8b-d).

Experimental evidence of loop extrusion

The first evidence that SMC complexes are indeed able to perform loop extrusion stems from experiments on yeast condensins: Although condensins had been hypothesized to have motor function, since they required ATP to compact DNA *in vitro* (Strick *et al.*, 2004), it was only shown recently that condensin can reel in DNA by an *in vitro* single-molecule imaging approach, where fluorescently DNA is attached at two points to a glass side and placed in a buffer flow. With the buffer flow, ATP as well as the purified components of the condensin complex are supplied. These will then bind to

DNA and extrude loops, indicated by the entangling of the DNA in the middle (Ganji *et al.*, 2018). Even though, mitotic chromosome loops have been already visualized using EM (Paulson & Laemmli, 1977), so far, loop extrusion by condensins could not be confirmed *in vivo*.

Cohesin was known for its function in sister-chromatid cohesion (see 1.3.2), but it was then hypothesized as potential LEF for chromosome loops in interphase, since it is not only expressed during S, G2 and M phase, but also in G1 and post-mitotic cells, such as neurons (Wendt *et al.*, 2008; Mönnich *et al.*, 2009). The first evidence that this is indeed the case was provided by experiments using inducible degron system to deplete cohesin and the potential BE, CTCF: ChIP-seq experiments confirmed that cohesin is accumulated at CTCF binding sites (Wendt *et al.*, 2008) and that this accumulation is lost upon loss of CTCF (Nora *et al.*, 2017). Furthermore, Hi-C experiments showed that loss of cohesin led to a loss of CTCF-mediated interactions and TADs (S. Rao *et al.*, 2017). CTCF was shown not to be required for the formation of interactions, but rather for their positioning. Finally, cohesin is also able to perform loop extrusion *in vitro*, as was shown by similar single-molecule imaging experiments as those on condensins (Davidson *et al.*, 2019).

Characteristics of loop extrusion

These discoveries enabled the confirmation of many quantitative predictions of the initial loop extrusion model, such as the extrusion speed of ~ 1 kb/s (measured *in vitro*) (Davidson *et al.*, 2019; Y. Kim *et al.*, 2019) and the average size of loops being approximately 100 kb depending on the cell type (measured by Hi-C) (S. S. Rao *et al.*, 2014; Fudenberg *et al.*, 2017). However, in general experimental Hi-C heatmaps could only be partially reconstituted using polymer modeling with loop extrusion given certain genomic CTCF binding patterns (Fudenberg *et al.*, 2016). This led to the conclusion that the underlying assumptions of the loop extrusion model must be amended or changed (Corsi *et al.*, 2023): Because of single-molecule imaging approaches *in vitro* that observed one-sided loop extrusion of yeast condensins (Ganji *et al.*, 2018; Shaltiel *et al.*, 2022) and both one- and two-sided loop extrusion of human cohesin (Davidson *et al.*, 2019; Kong *et al.*, 2020), it is now in question whether cohesin would extrude symmetrically and in both directions simultaneously *in vivo*. Indeed, polymer simulations using either a mixture of one- and two-sided LEFs or an one-sided LEF that is able to switch directionality, are also able to recapitulate Hi-C heatmaps (Banigan *et al.*, 2020). A switch in directionality was recently shown to happen when cohesin encounters CTCF bound sites (Davidson *et al.*, 2023), indicating that the previously observed symmetrical loop extrusion of cohesin might in fact be due to the switching of a one-sided LEF.

In addition, it is possible that SMC complexes might be able to pass each other whilst extruding (shown for yeast condensin) (E. Kim *et al.*, 2020; Brandão, Ren, *et al.*, 2021), i. e. they might not stop each other, but rather are able to form intermingled loops (“Z-loops”). Even though cohesin might not be paused by other cohesin complexes on DNA, other large complexes, such as PolII (or active transcription) and the MCM complex could be barriers to loop extrusion (S. Zhang *et al.*, 2021; Dequeker *et al.*, 2022). This would in turn also change where cohesin accumulates on DNA and what interactions are preferentially established.

1. INTRODUCTION

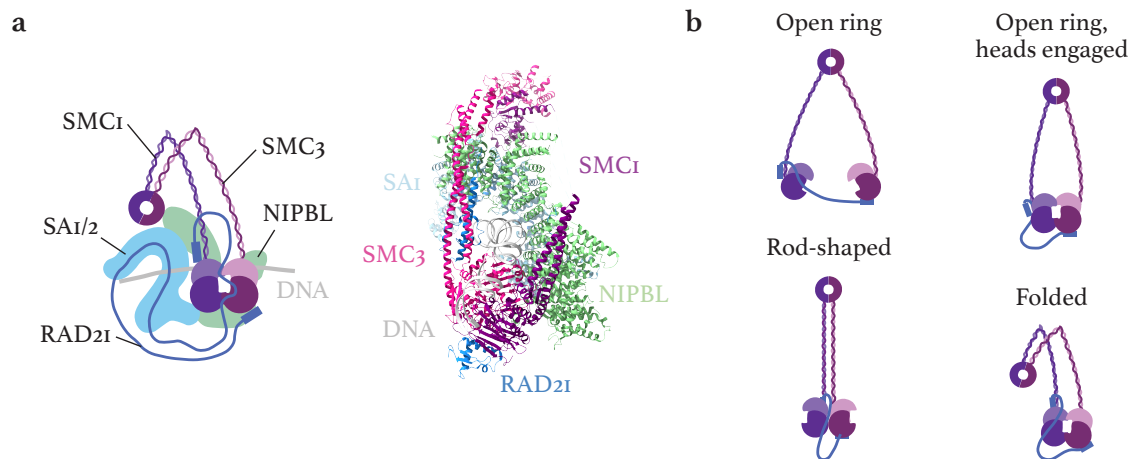


Figure 1.9: A partial structure of the loop-extruding cohesin complex. **a**, Partial structure of the cohesin-NIPBL complex showing that DNA binds the complex at the ATPase domains of SMC1/3. NIPBL builds a channel for DNA at this interface. **b**, Potential configurations of SMC complexes observed in structural studies. These led to the formation of multiple mechanistic models of how cohesin performs loop extrusion. *Structure from Shi et al., 2020, PDB:6WG3, schemes adapted from Shi et al., 2020 and Lee et al., 2020.*

Furthermore, recent studies providing structural insights into the complex and its binding to DNA in combination with single-molecule imaging approaches, have led to a better understanding of how cohesin functions mechanistically: Cohesin's ATPase activity as well as the HAWK protein NIPBL are required for loop extrusion but not necessarily for its binding to DNA (similar to condensin) (X. Wang *et al.*, 2018; Davidson *et al.*, 2019; E. Kim *et al.*, 2020). The same study also showed that cohesin extrudes loops as a monomer and does so without topologically entrapping the DNA. This argues for alternative binding modes of cohesin to DNA for loop extrusion. A recent, partial cryo-EM structure of cohesin bound by NIPBL and DNA suggests that the ATPase domains of SMC1/3 together with NIPBL build a channel for DNA binding (Shi *et al.*, 2020) (Fig. 1.9a). This is in line with the finding that acetylation of the two key residues in the SMC3 ATPase domain lead to reduced ATPase activity (Wilhelm *et al.*, 2015) which was shown to be required for loop extrusion. However, it is not known, whether there are more binding sites within the cohesin complex that can interact with DNA. Although topological entrapment that requires opening of cohesin's ring structure was ruled out as a potential interaction mechanism (Davidson *et al.*, 2019), it is still possible that DNA is pulled through the cohesin ring from the side (pseudo-topological entrapment).

Besides the modes of DNA binding of cohesin, experimental evidence points to a variety of *in vivo* conformations of the cohesin complex (Fig. 1.9b): The SMC complex of *B. subtilis* was shown to exist in two configurations, a rod-shaped and ring-shaped configuration (D. E. Anderson *et al.*, 2002; Huis in 't Veld *et al.*, 2014; Soh *et al.*, 2015). In the rod-shaped cohesin complex, the two coiled-coil domains of the SMC are aligned, where the open-ring structure occurs when the ATPase heads are bound by ATP (Diebold-Durand *et al.*, 2017). Both cohesin and condensin have also been shown to form a bent complex, where the hinge domain is close to the ATPase heads. Disruption of this conformation through mutations in the coiled-coil domains was shown to impair function of the *E. coli* SMC complex (Bürmann *et al.*, 2019).

Mechanistic models of loop expansion

Based on this knowledge, multiple mechanistic models of loop extrusion have been proposed. I will only briefly describe the three most prominent, pointing to the evidence in favor of those models. For more details, I refer the reader to a review by Davidson and Peters, 2021 and Oldenkamp and Rowland, 2022.

The “walking model” proposes that cohesin might “walk” on DNA, like e. g. kinesins walk on microtubules. For this, both ATPase heads would have to bind to DNA independently and ATP hydrolysis would lead to dynamic binding and unbinding cycles of those domains (Fudenberg *et al.*, 2017; Nichols & Corces, 2018). A secondary DNA binding site, e. g. at the hinge domain (called the “safety belt”), which was found to be necessary for loop extrusion by yeast condensin (Kschonsak *et al.*, 2017), would then hold onto the second part of the DNA in *cis* thereby leading to asymmetric loop extrusion. Symmetric loop extrusion could be achieved through pseudo-topological entrapment within the cohesin ring instead of the “safety belt” at the hinge domain. However, unlike kinesins, cohesin lacks a directionality and might also be able to switch between directions. While this “walking model” was initially considered to be the most plausibly, recent evidence points towards two other models:

The “pumping model” (also “zipping model”) comes from the finding that the *B. subtilis* SMC complex changes between rod-shaped and open-ring conformations, when binding ATP (Diebold-Durand *et al.*, 2017). It is therefore possible that in an ATP-engaged open-ring configuration, DNA will bind to the interface built by the ATPase head domains. When ATP is hydrolyzed the rod-shaped configuration will be established again starting from the hinge domain down towards the ATPase head domains. If one part of the DNA is held at the head domain interface, the second part could be “pumped” through the ring by the formation of the rod-shaped configuration (Marko *et al.*, 2019; Bauer *et al.*, 2021; Vazquez Nunez *et al.*, 2021). This would lead to a pseudo-topological entrapment of DNA and requires two separate modes of binding of DNA to the SMC1/3/RAD21 interface.

The “scrunching model” (also “clamping” or “Brownian ratched model”) is based on the observation that cohesin can exist in a bent configuration (Bürmann *et al.*, 2019; Lee *et al.*, 2020; Shi *et al.*, 2020), which could allow a “hand-over” of DNA from the binding site at the hinge domain to the binding site at the ATPase head domains (or vice versa). Hydrolysis of ATP could then drive the handover which could lead to DNA translocation, if one part of the DNA is anchored permanently on cohesin. Symmetrical loop extrusion can be achieved, if the DNA part that is permanently attached can be alternated between both (Higashi *et al.*, 2021).

1.3.4 Barrier elements to extrusion: CTCF

The CCCTC-binding factor (CTCF) protein is a DNA-binding protein consisting of unstructured N- and C-termini that flank 11 zinc fingers (ZFs) (Lobanenkov *et al.*, 1990; Martinez & Miranda, 2010; Ong & Corces, 2014). CTCF itself, but also its binding motif, is highly conserved in many animals and was first described for its “enhancer-blocking activity”, i. e. insulating function (T. H. Kim *et al.*, 2007; Phillips & Corces, 2009). It is important to note that the CTCF binding motif is an

1. INTRODUCTION

asymmetrical binding motif, meaning that CTCF can bind in two possible orientations depending on whether the binding motif is present on the (+) or (-) strand of the DNA. In mammalian cells, $\sim 40,000$ -90,000 binding sites for CTCF can be identified depending on the cell type and also the experimental procedure used (T. H. Kim *et al.*, 2007; Ong & Corces, 2014). 50% of all binding sites are found at promoters, in introns or exons and almost all CTCF bound sites co-localize with a site that is also bound by cohesin in ChIP-seq experiments ($\sim 90\%$) (Rubio *et al.*, 2008; Wendt *et al.*, 2008; Hansen *et al.*, 2017).

CTCF's enhancer-blocking activity stems from the observation that deletion of CTCF binding sites leads to disruption of TAD boundaries and ectopic interactions between enhancers and promoters in previously adjacent TADs (see 1.2.4) (Lupiáñez *et al.*, 2015). Genome-wide comparisons of CTCF binding patterns with Hi-C maps led to the development of the “CTCF convergence rule”: Chromatin loops are preferentially formed between CTCF sites that are found in a convergent orientation (Fig. 1.10a). Inversion of a CTCF site can also lead to disruption of TAD boundaries and merging of TADs (S. S. Rao *et al.*, 2014; Y. Guo *et al.*, 2015; Sanborn *et al.*, 2015; van Bemmelen *et al.*, 2019).

However, not all CTCF motifs found in the genome are boundaries of TADs or cohesin-mediated

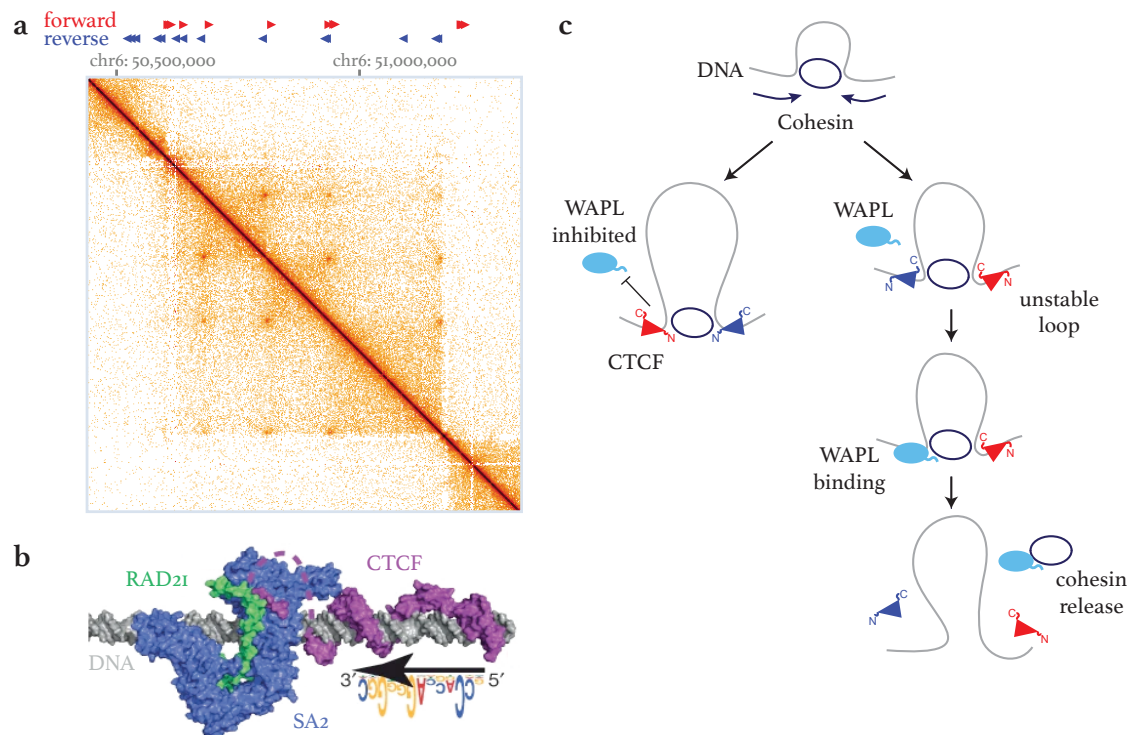


Figure 1.10: The molecular basis of the “CTCF convergence rule”. **a**, Hi-C heatmap showing that CTCF binding sites correspond to the boundaries of TADs with the orientation of the CTCF sites pointing towards the inside of the TAD (convergence). **b**, Structural model proposing a mechanism for how cohesin specifically engages the N-terminus of CTCF through an interaction surface built by SA2/RAD21 (*model from Y. Li et al., 2020*). **c**, Potential mechanism of how cohesin is stabilized on DNA when bound to the N-terminus of CTCF. Interaction of CTCF with cohesin through the SA2/RAD21 surface prevents WAPL from binding to cohesin and releasing it from DNA. *Scheme modified from Davidson and Peters, 2021.*

loops. CTCF bound sites can also be found within TADs and dot-like structures in Hi-C that indicate a CTCF and cohesin-mediated loop, can be established between convergent CTCF sites that skip other convergent sites in between (S. S. Rao *et al.*, 2014). A possible reason for these observations is that CTCF is not always bound at each site (presumably around 50% of the time in mESC) as it also has a much shorter residence time on DNA than cohesin (Cattoglio *et al.*, 2019). Binding could be regulated by other processes on DNA such as CpG methylation. For example, this was shown for the case of allele-specific expression at the H19/insulin-like growth factor 2 (*IGF2*) imprinting control region, where differential methylation of a CTCF binding site leads to CTCF-anchored loops only on the unmethylated allele (Bell & Felsenfeld, 2000; Kurukuti *et al.*, 2006; Splinter *et al.*, 2006). It is further possible that the number of CTCF molecules per cell is limiting so that not all sites can be bound at the same time (Cattoglio *et al.*, 2019; Holzmann *et al.*, 2019), that other DNA-binding proteins compete with CTCF for the binding site (Kaaij *et al.*, 2019) or that post-translational modifications (PTMs) in CTCF could alter its specificity for certain sites (Wutz *et al.*, 2017).

Currently, it is not yet well understood how CTCF interacts with cohesin in order to stop extrusion and how this is achieved in a manner that is compatible with the “CTCF convergence rule”. However, recent structural insights and mutagenesis studies show that the N-terminus of CTCF interacts with the cohesin complex through an interface on SA2 and RAD21 suggesting that the N-terminus of CTCF first encounters cohesin in the convergent orientation (Y. Li *et al.*, 2020; Nora *et al.*, 2020; Pugacheva *et al.*, 2020) (Fig. 1.10b). If this is the interaction that leads to cohesin stalling at CTCF sites, it is expected to be very strong and rapidly forming, as it needs to stop cohesin while passing by and can only do so through interaction at the N-terminus, since the C-terminus does not seem to be required. Mutations that disrupt the CTCF binding interface on SA2/RAD21 only reduce the accumulation of cohesin at CTCF sites, but do not abolish it. However, CTCF-mediated loops are not detected under these conditions (Y. Li *et al.*, 2020). This argues in favor of multiple mechanisms working in concert to stall cohesin at CTCF sites.

A “pause and stabilization” mechanism was proposed that could also explain the “CTCF convergence rule” (Wutz *et al.*, 2017; Y. Li *et al.*, 2020; Nora *et al.*, 2020; Pugacheva *et al.*, 2020; Wutz *et al.*, 2020): Hereby, CTCF would pause cohesin briefly upon encountering it (potentially independent of the binding orientation of CTCF), but cohesin would eventually resume loop extrusion after the pausing period (Nora *et al.*, 2020). This pausing could be achieved through multiple, possible mechanisms, such as by steric hindrance due to a conformational change in the chromatin structure at CTCF bound sites (MacPherson & Sadowski, 2010; Clarkson *et al.*, 2019) or bulky PTMs of CTCF, such as poly-ADP-ribosylation that was shown to contribute to CTCF’s insulating function (Yu *et al.*, 2004).

When then cohesin encounters CTCF in a convergent orientation, i. e. at its N-terminus, the pausing period could be used to stabilize the interaction between cohesin and CTCF thereby leading to a reinforcement of the loop. Different ways of achieving stabilization have been proposed that are not mutually exclusive and could work together: First, protein-protein interactions at the CTCF/SA2/RAD21 interface could lead to stabilization, as indicated by the structural evidence. However, this interaction is quite weak *in vitro* ($\sim 0.6 \mu\text{M}$) (Y. Li *et al.*, 2020). Alternatively, CTCF could modulate how cohesin

1. INTRODUCTION

functions by reducing its capacity to hydrolyze ATP, thereby locking it in place (Wutz *et al.*, 2017; Wutz *et al.*, 2020). This could be achieved, if CTCF recruited PDS5, which displaces NIPBL from the loop-extruding holoenzyme and thereby reduces its ATPase activity. This is supported by Hi-C experiments in PDS5-depleted cells where CTCF loops violate the “CTCF convergence rule” (with 65-92% of convergent loops under WT conditions and 31% in PDS5-depleted cells, with 25% expected, if CTCF orientation was of no consequence) (S. S. Rao *et al.*, 2014; Y. Guo *et al.*, 2015; Wutz *et al.*, 2017). Finally, CTCF might also act to counteract cohesin’s release from DNA by outcompeting the release factor WAPL (Kueng *et al.*, 2006), which was shown to bind the same surface on the cohesin complex, but less strongly than CTCF (Y. Li *et al.*, 2020) (Fig. 1.10c). Indeed, either depleting CTCF entirely or mutating its key interaction surface with cohesin (Y226A and F228A) decreases cohesin’s residence time on DNA. Furthermore, cohesin^{SA1} gets acetylated more often by ESCO1 (which is active throughout the cell cycle) than cohesin^{SA2}, leading to a longer residence time of cohesin^{SA1} and more stable loops (Wutz *et al.*, 2020).

1.3.5 Regulators of cohesin

The following section will briefly review the main function of cohesin’s regulatory proteins (besides CTCF, see 1.3.4), which have already been mentioned in the previous sections for modulating cohesin. This section thereby aims to summarize their functions and effects on 3D genome organization.

Either NIPBL or PDS5 are thought to be integral parts of the cohesin complex, as these are both HAWK proteins and occupy the same interface in the complex. Their binding is therefore mutually exclusive and they are thought to regulate a switch in cohesin’s function: From an actively extruding complex with high ATPase activity (NIPBL) to a complex whose binding to DNA is stabilized and held in place (PDS5) (Petela *et al.*, 2018; van Ruiten *et al.*, 2022).

Sororin and WAPL are not always associated with cohesin, but also function to regulate cohesin’s residence time on DNA. Sororin binds to the surface of the cohesin complex at the SA2/RAD21 interface (as does CTCF) (Y. Li *et al.*, 2020), which is why also their binding is presumed to be mutually exclusive. Binding to this surface is thought to prevent cohesin’s dissociation from DNA by preventing WAPL, the release factor of cohesin, from binding (Schmitz *et al.*, 2007; Nishiyama *et al.*, 2010). Sororin functions during sister-chromatid cohesion; CTCF fulfills a similar function when enforcing an interphase cohesin-extruded loop.

NIPBL: Loading and processivity of cohesin

Nipped-B-like protein (SCC2 in mice), together with MAU2, was first identified as possible ATP-dependent loading factor of cohesin (Ciosk *et al.*, 2000), but recent evidence showed that NIPBL (and ATP) is required for active loop extrusion (Davidson *et al.*, 2019; Y. Kim *et al.*, 2019) and that cohesin can also associate with DNA in a manner that does not require ATP hydrolysis (B. Hu *et al.*, 2011; Petela *et al.*, 2018). It is therefore currently in question, whether NIPBL is indeed a loading factor, or rather functions as a processivity factor, as it stimulates cohesin’s ATPase activity (Çamdere *et al.*, 2015; Davidson *et al.*, 2019). In line with these findings *in vitro*, Hi-C experiments in NIPBL-depleted

cells show a similar phenotype as in cohesin-depleted cells, namely a loss of all looping interactions and TAD boundaries (Schwarzer *et al.*, 2017).

However, NIPBL was found to be sub-stoichiometric compared to other subunits of the cohesin complex (Rhodes *et al.*, 2017), indicating that it is not always associated with it. Its release from the complex (e. g. through displacement by PDS5) was shown to lead to a stalling of loop extrusion (Petela *et al.*, 2018; Davidson *et al.*, 2019). CTCF bound to cohesin seems to facilitate this switch, thereby promoting a halt in loop extrusion at convergent sites (Wutz *et al.*, 2017).

Since NIPBL was also found to accumulate at active promoters in the absence of CTCF, it was hypothesized that promoters function as potential cohesin loading sites (Zuin *et al.*, 2014; Busslinger *et al.*, 2017). However, in line with NIPBL's function as a processivity factor, it is also probable that active transcription functions as a block to loop extrusion, thereby stalling loop-extruding cohesin complexes (i. e. NIPBL bound complexes) (Lengronne *et al.*, 2004; Busslinger *et al.*, 2017). Since stalling by active transcription constitutes a more passive manner than the active regulatory role of CTCF (through direct protein-protein interactions with the cohesin complex), NIPBL is not exchanged for PDS5 when encountering PolII (Banigan *et al.*, 2023).

PDS5: Stabilization of cohesin

PDS5 (precocious dissociation of sisters 5 protein) exists in two isoforms (PDS5A or B) and, by exclusion, promotes cohesin's functions on DNA that require stable and long-lived binding, such as "cohesive" cohesin. In S and G2, it is stably associated with the "cohesive cohesin" complex, also consisting of Sororin and an acetylated SMC3 subunit. These complexes are long-lived on DNA and are unable to perform loop extrusion (Murayama & Uhlmann, 2015; Petela *et al.*, 2018).

In G1, it binds to the complex, when also SMC3 is acetylated by ESCO1 (Chan *et al.*, 2013; Dauban *et al.*, 2020) or cohesin is stabilized by binding of CTCF (Wutz *et al.*, 2017). It counteracts NIPBL as it displaces it from the cohesin complex, thereby preventing the complex from performing loop extrusion (Petela *et al.*, 2018). Loss of PDS5 was shown to lead to larger cohesin-mediated loops that often violate the "CTCF convergence rule", showing that it is required for CTCF's ability to function as a boundary to loop extrusion (Wutz *et al.*, 2017). It was further shown that this role is related to the acetylation of cohesin in G1 by ESCO1 which shortens the length of chromatin loops and interactions anchored at CTCF bound sites. De-acetylation by Hos1 in budding yeast 8 (HDAC8) has the opposite effect, indicating that PDS5 and acetylation regulate cohesin's processivity (and potentially also its velocity) when extruding loops (van Ruiten *et al.*, 2022). The "vermicelli" chromosomes that can be observed in PDS5-depleted cells are therefore distinct from the ones that can be observed in WAPL-depleted cells (see 1.3.5): The higher processivity (or velocity) of cohesin on DNA leads to longer chromatin loops (Fig. 1.6).

Promoting a switch of function in cohesin from a loop extruding complex to a stably bound complex (either in a loop-enforcing role or when mediating sister-chromatid cohesion) might also be an efficient way of re-purposing complexes that are already bound to DNA. It is likely that at the onset of

1. INTRODUCTION

S phase, “extrusive cohesins” can be converted to “cohesive cohesins” which is a possibly explanation for CTCF-bound sites often being sites of sister-chromatid cohesion (Mitter *et al.*, 2020).

WAPL: Release of cohesin from DNA

Wings apart-like protein (WAPL) is the release factor of cohesin from DNA. Depletion of WAPL leads to “vermicelli” chromosomes, i. e. highly condensed chromosomes that look similar to the condensation mediated by condensin II in early mitosis (Paulson & Laemmli, 1977; Tedeschi *et al.*, 2013). Furthermore, loss of WAPL leads to a longer residence time of cohesin on DNA as well as to longer chromatin loops (Kueng *et al.*, 2006).

It binds to the cohesin complex through an interaction with SA1/2 and PDS5, which will then promote cohesin’s dissociation through the formation of an “exit gate” between RAD21 and SMC3 (Huis *et al.*, 2014). If cohesin is acetylated at its SMC3 subunit and bound by Sororin in S and G2 phase, it is protected from WAPL, as one of WAPL’s interaction sites is blocked (Schmitz *et al.*, 2007; Rolef Ben-Shahar *et al.*, 2008; Unal *et al.*, 2008; Nishiyama *et al.*, 2010).

In interphase, CTCF fulfills a similar function as Sororin (as it binds to the same surface) (Y. Li *et al.*, 2020) and thereby defines the lifetime of a loop: Stalled cohesin at CTCF-bound sites cannot be exchanged for freely diffusing cohesin and the residence time of cohesin, defined by WAPL, thereby restricts the lifetime of a loop (Davidson *et al.*, 2019). WAPL is therefore instrumental in regulating the lifetime of cohesin on DNA, which was shown to be in the range of 8-20 min by fluorescence recovery after photobleaching (FRAP) experiments (Kueng *et al.*, 2006; Tedeschi *et al.*, 2013). Longer-lived cohesin^{SA1} complexes have been found that can last up to several hours, however this requires active stabilization on DNA. The long lifetime is dependent on acetylation of SMC3 by ESCO1 and, surprisingly, also on CTCF. Depletion of CTCF leads to a decrease in the residence time of this sub-complex, however, only in the presence of WAPL (Wutz *et al.*, 2020).

Sororin: Promoting sister-chromatid cohesion

Sororin promotes cohesin’s function to hold together the sister-chromatids after replication until separation in mitosis. Sister-chromatid-sensitive Hi-C (scsHiC) experiments in HeLa G2 cells showed significantly reduced contacts in *trans* upon of depletion of Sororin using an AID system, while no change in *cis*-contacts was observed (Fig. 1.6). The opposite was shown for NIPBL depletion in those cells, in line with a role of Sororin in regulating “cohesive cohesin” and NIPBL in regulating “extruding cohesin” (Mitter *et al.*, 2020).

It is recruited to the complex upon acetylation of the SMC3 subunit (Rolef Ben-Shahar *et al.*, 2008) and prevents binding of WAPL to PDS5 and thereby release of cohesin from DNA (Rankin *et al.*, 2005; Schmitz *et al.*, 2007; Nishiyama *et al.*, 2010). However, Sororin has a short residence time on DNA, indicating that constant release and rebinding is required to ensure cohesion of sister-chromatids until mitosis (Ladurner *et al.*, 2016). Here, Sororin is phosphorylated and thereby released from the complex, so that cohesin can be unloaded from DNA by WAPL (Losada *et al.*, 1998; Kueng *et al.*, 2006).

1.4 Dynamics of chromosome folding processes

Irrespective of the exact nature of the biochemical reactions and molecular players mediating changes in chromosome structure and the transfer of regulatory information, much of our current knowledge stems from studies in fixed cells (such as Hi-C and DNA-FISH experiments) that give a snapshot in time of the conformation of chromosomes. Furthermore, findings must often be interpreted as “ensemble averages” as 3C techniques require averaging of results from millions of cells. Studies employing scHi-C, sequential DNA-FISH and polymer modeling have shown that the physical distances of a set of DNA loci vary substantially from one cell to another within the same cell population or tissue (Nagano *et al.*, 2013; Giorgetti *et al.*, 2014; Finn & Misteli, 2019). This suggests that if E-P communication requires physical proximity, it can generally only occur in a subset of cells at any given time. If this proximity is mediated by loop extrusion, cohesin and regulators of its activity should influence, how often these windows of opportunity are created and how long they last in time within individual cells.

To study the development of chromosome structure over time, a different set of techniques is therefore required. In the following, I will introduce polymer modeling and live-cell imaging of chromosomal loci as well as SMLM to study protein dynamics in live cells as methods that have been used to study chromosome folding dynamics. I will then summarize the current understanding of the dynamics of chromosome folding and its interaction partners.

1.4.1 Methods to study dynamics of chromosomes

Classical methods that have been employed to study chromosome dynamics are tagging individual loci within the DNA, e. g. through the integration of bacterial operator array sequences that can be visualized by binding of a repressor fused to a fluorescent protein. However, these approaches were usually limited to be employed in organisms such as yeast, since they require extensive genome engineering. In mammalian cells, chromosome dynamics have been studied through methods such as FRAP and fluorescence correlation spectroscopy (FCS) on labeled histones or other DNA-bound proteins (Cisse *et al.*, 2013; Hansen *et al.*, 2017; Nozaki *et al.*, 2017). Whilst these techniques provided first insights on the time scales of chromosome folding, they also rely on ensemble averages and cannot give information about the heterogeneity in the motion of individual molecules. Complementary to these *in vivo* approaches, *in vitro* approaches have hugely contributed to our understanding of the dynamics of cohesin. Single-molecule imaging of cohesin on DNA first describe dynamical properties, such as the velocity at which cohesin extrudes loops (Davidson *et al.*, 2019; Y. Kim *et al.*, 2019; Golfier *et al.*, 2020). However, these measurements were performed on naked DNA, and not on chromatin in the context of a crowded nuclear environment.

With the advances in genome engineering technologies in mammalian cells and the development of new labeling and microscopy strategies, new ways of studying chromosome dynamics have been pioneered that I will summarize in the following. In addition to that, theoretical modeling of dynamics of chromosome folding have recently led to a better understanding of the polymeric nature of

1. INTRODUCTION

chromosomes and what dynamical mechanisms are probable to explain the phenotypes seen in ensemble-averaged experimental data, such as Hi-C.

Live-cell imaging approaches to study the dynamics of DNA

The first technique that allowed SPT of a chromosomal locus *in vivo* used an artificial array of Lac operators (LacO) that were integrated into the genome of yeast. This was then visualized by the binding of a fluorescently-tagged LacI repressor (LacI) (here with an enhanced green fluorescent protein (eGFP)) (Robinett *et al.*, 1996; Marshall *et al.*, 1997). This showed for the first time that DNA undergoes sub-diffusive behavior *in vivo* and can be influenced by encounters with other nuclear structures. By now, alternative fluorescent repressor operator systems (FROs) have been developed, such as the Tet operator (TetO)/Tet repressor (TetR) system (Masui *et al.*, 2011; Lucas *et al.*, 2014; Giorgetti *et al.*, 2016) or the cumate operator (CuO) bound by the CymR repressor (Mullick *et al.*, 2006; Alexander *et al.*, 2019). However, using these operator array-based systems requires genome engineering to integrate the arrays. Until the development of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing, tagging a specific locus with rather large arrays proved difficult. The development of smaller arrays bound by brighter fluorophores that require less signal amplification also improved the feasibility of this approach in mammalian cells. The major advantage of this approach is the high and also constant signal-to-noise ratio (SNR) that can be achieved, making it the favorable system for studying chromosome dynamics over long periods of time (Shaban *et al.*, 2020).

Alternatively, the ANCHOR/ParB labeling system can be used (Saad *et al.*, 2014). This system also requires integration of a small sequence in the genome. However, it does not rely on signal amplification through multiple binding events to an array of the DNA sequence. This means that the targeted sequence is much smaller and therefore more easy to integrate. The signal amplification is achieved by accumulation of the fluorescently-tagged ParB protein at the integration site. This system was shown not to disturb other processes on DNA such as transcription at nearby loci (Germier *et al.*, 2017).

Also, approaches that do not require genome engineering were developed: TALE (Ma *et al.*, 2017) and CRISPR/dCas9 approaches (B. Chen *et al.*, 2013; Clow *et al.*, 2022) target complementary guide sequences to the locus of interest, analogous to their use in genome editing. Since these are catalytically inactive and fluorescently tagged versions of these proteins, they will only bind to the locus and thereby indicate its position. The drawback of these approaches is the low and variable SNR that highly depends on transfection efficiency. Furthermore, efficient labeling was only achieved at repetitive loci that could be recognized by the same guides. Table 1.2 summarizes techniques to study chromosome dynamics with its advantages and drawbacks (van Staaldin *et al.*, 2023).

All of these techniques confirmed the sub-diffusive behavior of DNA (Zidovska *et al.*, 2013; Germier *et al.*, 2017; Gu *et al.*, 2018), although some studies also reported seeing super-diffusive motion in some cases (Levi *et al.*, 2005; Chuang *et al.*, 2006). Coupling these strategies to methods to visualize nascent RNA allowed to follow E-P interactions and transcription at the same time (Germier *et al.*,

Table 1.2: Techniques for visualizing DNA in living cells. Table adapted from Bystricky, 2015; van Staalduinen et al., 2023.

| Advantages | Drawbacks | References |
|---|---|---|
| Fluorescent repressor operator system (FROS) | | |
| Signal amplification possible with SunTag (Peng <i>et al.</i> , 2022). | | |
| <ul style="list-style-type: none"> - multiple orthogonal arrays can be used - 96 repeats give a sufficient signal strength - very consistent SNR across cell populations | <ul style="list-style-type: none"> - genome engineering required (challenging in mammalian cells) - potential chromatin disruption and heterochromatin formation - interferes with transcription | <ul style="list-style-type: none"> - Alexander <i>et al.</i>, 2019 - H. Chen <i>et al.</i>, 2018 - Masui <i>et al.</i>, 2011 |
| Transcription activator-like effectors (TALE) | | |
| Signal amplification possible with quantum dots (Ma <i>et al.</i> , 2017). | | |
| <ul style="list-style-type: none"> - no genome editing required - highly versatile: in any cell type - labeling of multiple loci simultaneously | <ul style="list-style-type: none"> - restricted to repetitive sequences - for single loci only in combination with quantum dots - design of TALEs very labor intensive | <ul style="list-style-type: none"> - Ma <i>et al.</i>, 2017 - Popp <i>et al.</i>, 2021 |
| Clustered regularly interspaced short palindromic repeats (CRISPR)/dCas9 | | |
| Signal amplification used is RNA aptamers (Shechner <i>et al.</i> , 2015), SunTag (Tanenbaum <i>et al.</i> , 2014), Casilio (Cheng <i>et al.</i> , 2016) and ArrayG (Gustavsson <i>et al.</i> , 2022). | | |
| <ul style="list-style-type: none"> - no genome editing required - highly versatile: in any cell type - labeling of multiple loci simultaneously | <ul style="list-style-type: none"> - low SNR - can block transcription - low homogeneity of transfection efficiency | <ul style="list-style-type: none"> - Clow <i>et al.</i>, 2022 - Shechner <i>et al.</i>, 2015 |
| ANCHOR (ParB/INT) | | |
| No signal amplification described. | | |
| <ul style="list-style-type: none"> - only small sequence (1 kb) insertion - multiple ANCHOR systems can be used simultaneously - can be used close to enhancers/intragenic regions | <ul style="list-style-type: none"> - genome editing required - not all ANCHOR systems work to the same extent | <ul style="list-style-type: none"> - Gabriele <i>et al.</i>, 2022 - Germier <i>et al.</i>, 2017 |

2017). TF dynamics and the response of a promoter were studied in yeast by combining, DNA and RNA labeling with SMLM of the TF (see also 1.4.1) (Donovan *et al.*, 2019). Finally these approaches have been used in combination with new innovative strategies to perturb the chromosome dynamics, i. e. in this case by “moving” a locus within a living cell through magnetic forces (Keizer *et al.*, 2022).

1. INTRODUCTION

Single-molecule localization microscopy: Following protein dynamics on DNA

Single-molecule localization microscopy complements techniques, such as FRAP and FCS, that are traditionally used to study dynamics of proteins interacting with DNA and infer kinetics at the molecular scale. In FRAP, a fluorescently labeled protein population is photo-bleached in a small (often diffraction-limited) area of the nucleus (Axelrod *et al.*, 1976). The recovery of fluorescence in that area is then observed over time, giving insights on how fast photo-bleached protein in the area is exchanged for protein from the surrounding. The shape of the recovery curve is fit to a model to infer the underlying dynamics. This was for example used to study the dynamics of PolII and to study chromatin-binding kinetics of CTCF and cohesin (Darzacq *et al.*, 2007; Hansen *et al.*, 2017). In FCS, a point-detector is used to measure how emissions from a diffraction-limited area that has been illuminated with a focused laser, vary over time (Magde *et al.*, 1972). Based on the fluctuations, a model can be fit that infers the number of measured molecules in that area. This allows to infer diffusion kinetics. e. g. also of chromatin-binding events (Plachta *et al.*, 2011). However, both these techniques are highly model-dependent and only give an estimate of the dynamics of the ensemble average, although it is likely that binding and diffusion of proteins interacting with DNA is highly variable and occurs over many different temporal scales (Mazza *et al.*, 2012). In contrast to these techniques, SMLM can provide information about the heterogeneity in the dynamics as it allows for tracking of individual molecules over time in their spatial context within the nucleus. It has been successfully used to study clustering of PolII (Cisse *et al.*, 2013), the search dynamics of CTCF (Hansen, 2020) and condensates formation of PRC1 (Zhen *et al.*, 2016).

To be able to visualize a single protein, SMLM overcomes the limitations classical arising from the diffraction limit in light microscopy (Abbé, 1873): The single proteins need to be distinguished from each other and from the background noise. Since single proteins are objects smaller than the diffraction limit (a few nanometer (nm)) and therefore show a point-spread-function (PSF), i. e. a Airy function, as a response to the imaging system, SMLM requires a high SNR and sparse detections to reduce the probability that two molecules are separated by a distance smaller than can be resolved in light microscopy. *In vivo* this is usually achieved using sparse localization microscopy (R. E. Thompson *et al.*, 2002). This can be achieved through labeling only a small subset of the molecules of interest or using sparse excitation (which is also used in super-resolution techniques as PALM and STORM) (Betzig *et al.*, 2006; Rust *et al.*, 2006). Recently, tracking of single fluorophores with a resolution of ~ 1 nm was achieved using the MINFLUX technology: Here, a donut-shaped excitation laser is positioned in the vicinity of a fluorophore and the emission of the fluorophore is used to reposition the excitation laser, until the fluorophore is at its center (so that the position of the fluorophore is defined) and no emission can be detected anymore (Balzarotti *et al.*, 2017). The detection is not camera-based and relies on far fewer photons to be detected. However, throughput is low as only one molecule can be followed at a time. This method was recently also used to study kinesin stepping *in vivo* (Deguchi *et al.*, 2023; Wolff *et al.*, 2023).

A sparse labeling for SMLM is usually achieved by tagging the endogenous proteins of interest either with photo-activatable or photo-convertible fluorescent proteins (such as Dendra or EosFP)

(Wiedenmann *et al.*, 2004; Gurskaya *et al.*, 2006) or with self-labeling peptides, such as the SnapTag or HaloTag, that can subsequently be covalently bound by a ligand containing a bright organic fluorophores (Gautier *et al.*, 2008; Los *et al.*, 2008). Both approaches require stochastic labeling, i. e. either the integrated tag is only expressed on a subset of molecules via translational read-through or the molecules within the nucleus are labeled with a limiting amount of ligand (H. Liu *et al.*, 2018; Nora *et al.*, 2020). Compared to sparse excitation, this approach has the advantage that much longer trajectories can be acquired.

Since in the ideal case, the excitation of a single fluorophore should be detected as a single spot in the image (resolved from its neighbors), SMLM requires a high excitation power to collect enough photons from the single emitter. This limits the number of localizations until the emitter is permanently photo-bleached. Therefore, every SMLM experiment is an optimization problem that tries to maximize different aspects of the experiment: localization precision, temporal resolution and trajectory length, all depending on photon budget and the quantum yield of the fluorophore. Depending on the biological question, one may choose to compromise e. g. the localization precision, defined by the SNR and the number of photons collected (Bobroff, 1986; R. E. Thompson *et al.*, 2002), for the length of the trajectory and the sample health. For a more detailed description on parameters that are involved in designing a SMLM experiment, I refer the reader to the review by Boka *et al.*, 2021.

Polymer modeling of chromosome folding

Many characteristics of chromosome folding can be understood by viewing chromosomes as long polymer chains (i. e. consisting of millions of monomers) that are confined within the space of the nucleus. In contrast to that, DNA-binding proteins are relatively small objects that can only affect their immediate vicinity (i. e. $\sim 1\text{-}5$ nm) through protein-protein or protein-DNA interactions. In order for these interactions to take effect and e. g. mediate the exchange of information between an enhancer and promoter, the two loci of DNA need to be already in very close proximity (in the nanometer range, see 1.1.5). It is therefore important to understand, how DNA moves in the nucleus and encounters of loci are achieved within the 3D nuclear space. Polymer simulations have been instrumental in identifying probable mechanisms that can explain these processes:

Due to the polymeric nature of DNA, the contact probability between two loci that are closely located in the linear sequence of the genome is higher than of loci that are further apart. Dekker and Mirny, 2016 estimated that if one were to take a melt of polymers corresponding to the size of chromosomes confined to a spherical space the size of a nucleus, contacts between two loci separated by 1 Mb would be very rare with an average distance of $1.4\ \mu\text{m}$ and a measured Hi-C contact probability of $P(1\text{Mb}) \approx 10^{-3} - 10^{-4}$. This would mean that these loci would only interact in a few out of 10,000 cells. However, due to the dynamics of chromosomes, where a monomer in the chain is held back in its motion by the chain (which e. g. can be modeled as a Rouse polymer), a locus within DNA explores a smaller volume with a certain characteristic size (estimated 150-300 nm in 100 s). This would facilitate multiple different contacts in the local vicinity of the locus and would prevent exploration of distant areas. Such a sub-diffusive behavior has also been observed in live-cell imaging experiments (see 1.4.1)

1. INTRODUCTION

(Bronstein *et al.*, 2009; Lucas *et al.*, 2014; Gu *et al.*, 2018).

However, this does not explain, how certain loci distant in the linear genome can interact more often with each other. Only assuming polymer dynamics, this would mean that an enhancer that is positioned more than 1 μm from its promoter in physical space would likely never interact within the length of a cell cycle. Models could show that compartmentalization of DNA at different length scales can reduce the search time, e. g. organization into chromosome territories can facilitate interactions within the chromosomes and thereby reduce the time it takes for two loci to interact (Cremer & Cremer, 2001; Branco & Pombo, 2006; Rosa & Everaers, 2008).

On a smaller (sub-megabase scale), where loci are likely to interact given chromosome dynamics and compartmentalization, other forces are needed to regulate, which contacts are facilitated and which are prevented: Here, theoretical modeling preceded even empirical evidence from biological experiments with the formation of the loop extrusion model (see 1.3.3) (Fudenberg *et al.*, 2016). Such a mechanistic ensemble approach has the advantage that new mechanisms that alter chromosome conformations can be tested, resulting in an ensemble of conformations that can be compared to experimental data (e. g. Hi-C maps) (M. V. Imakaev *et al.*, 2015). It has been used to explore multiple possible mechanisms to explain chromosome folding patterns, such as a block copolymer model that takes into account local epigenetic information to explain mammalian interphase Hi-C maps (Jost *et al.*, 2014) or centromere clustering to explain the conformation of yeast chromosomes in interphase (Wong *et al.*, 2012). Alternative approaches are the study of well-known structural ensemble models, such as the random walk, the self-avoiding walk (SAW), equilibrium globules, the fractal globule or a melt of polymer rings (Grosberg *et al.*, 1988; van den Engh *et al.*, 1992; Grosberg *et al.*, 1995; Lieberman-Aiden *et al.*, 2009; Vettorel *et al.*, 2009). Comparison of their characteristics with experimental data, can inform on possible mechanisms that give rise to the seen chromosome structure without having to observe their dynamics. Finally, polymer modeling can also be data-driven: These models use the averaged Hi-C contact maps or DNA-FISH data as input to build spatial models of chromosomes, either by informing on potential ensembles of single-cell conformations (Giorgetti *et al.*, 2014) or by building a consensus structure (M. Hu *et al.*, 2013).

In conclusion, polymer modeling can inform on a variety of characteristics of chromosome folding and, together with mathematical models, has pushed forward our understanding of different aspects of chromosome organization and the processes that facilitate it. For the sake of brevity, I will not discuss mathematical models of E-P communication, but I refer the reader to recent studies that explored hypotheses on how transcriptional bursts are generated (Lammers *et al.*, 2020) and how this connects with E-P contact probabilities (Xiao *et al.*, 2021; Zuin *et al.*, 2022). Further, they have allowed us to quantify how regulatory information is processed by E-P models into cell states (Tkačik & Gregor, 2021).

1.4.2 Time scales of chromosomal interactions

Until recently, the time scales of chromosomal interactions, especially at the sub-megabase scales, where E-P communication happens, were incompletely understood. However, since TADs are established by the loop-extrusion activity of cohesin, the dynamics of chromosome structure inside a TAD is expected to be linked to the kinetics of loop extrusion and of CTCF binding.

Based on the topics discussed in the previous sections, I will give an overview of our current knowledge of chromosome dynamics: Insights from scHi-C and DNA-FISH studies suggested that the conformation of the chromatin fiber within a TAD might vary considerably over the timescale of minutes in living cells (Flyamer *et al.*, 2017; Bintu *et al.*, 2018). Analysis of the motion of genomic locations (Masui *et al.*, 2011; Germier *et al.*, 2017; Gu *et al.*, 2018; Khanna *et al.*, 2019) or chromatin-bound proteins (Nozaki *et al.*, 2017) initially suggested that chromatin motion is sub-diffusive and that promoters or enhancers explore the surrounding nuclear space in a way that might depend on their activity (Germier *et al.*, 2017; Gu *et al.*, 2018). Only very recently a live-cell imaging study provided insights into how CTCF and cohesin control the looping dynamics of chromosomal sequences located close to TAD boundaries. Analysis of the dynamics of pairs of genomic loci in mESC revealed that cohesin-mediated loops between pairs of convergent CTCF sites last ~ 10 -30 minutes on average (Gabriele *et al.*, 2022), which is a relatively small fraction of the cell cycle duration. Together with other recent studies of chromatin motion (Keizer *et al.*, 2022; Brückner *et al.*, 2023), these new results suggested that windows of opportunities for regulatory exchange might be transient events occurring as enhancers and promoters explore the nuclear space in the context of a highly dynamic chromatin fiber. However, it is still unclear, what happens within a TAD: How is the interaction between two loci influenced by cohesin and CTCF and what are interaction frequencies and duration for linear separations of the size of a typical loop?

Finally, a series of recent exciting studies based on quantitative microscopy of reconstituted protein complexes demonstrated that mammalian cohesin indeed extrudes loops of DNA and can be blocked by CTCF, and even measured rates of extrusion *in vitro* (~ 1 kb/s) (Davidson *et al.*, 2019; Y. Kim *et al.*, 2019; Goller *et al.*, 2020). Single-molecule tracking and analysis of protein mobility *in vivo* (Cattoglio *et al.*, 2019; Wutz *et al.*, 2020) additionally suggested that cohesin resides on DNA for 8-20 minutes on average. Although cohesin's binding kinetics to DNA have been studied by FRAP and SMLM, little is known about the dynamics of the cohesin in living cells when bound to DNA.

1.5 Aim of this thesis

In order to better study how changes in chromosome structure affect transcriptional regulation, we require an understanding of the dynamics of chromosome folding processes as a great variability across cells has been reported in single cells. Therefore, the goal of this thesis is a description of the dynamics of chromosome folding, in particular at the sub-megabase scale, and of its regulators, i. e.

1. INTRODUCTION

cohesin and CTCF. With this thesis, I would like to address questions, such as: How does chromosome conformation within a TAD evolve in time? How often do genomic sequences meet each other within the span of a cell cycle and how long do their interactions last? How does this depend on factors involved in loop extrusion, such as cohesin and CTCF? Finally, what are the dynamics of cohesin on chromatin and how does it depend on the process it mediates?

To this end, I employed a live-cell imaging approach that allowed me to image tens of random chromosomal insertions of the TetO array in mESCs in the presence and absence of factors involved in loop extrusion, such as RAD21, CTCF and WAPL. This revealed global effects on chromosome folding, such as an effect of cohesin in constraining chromosome motion. Further, I asked, how interactions of chromosomal loci *in cis* are affected by cohesin by studying the dynamics of an ectopic 150 kb-sized CTCF loop. Through imaging the temporal dynamics of this loop, I could show that the interaction frequency and duration are modified by cohesin as well as CTCF and that interactions between two loci *in cis* last on average only around 6 minutes. My measurements in combination with polymer simulations also allowed a quantitative estimate of the frequency and duration of the molecular looped state with a CTCF- and cohesin-mediated loop lasting on average 5-15 min. This work is described in chapter 2.

The notion that cohesin restricts chromosome motion and thereby facilitates longer and more frequent contacts between chromosomal loci *in cis* can be an effect of its role in sister-chromatid cohesion or in loop extrusion, respectively, since both these functions enable the formation of chromosome loops between two strands of DNA. Therefore I asked: Do dynamics of cohesin change depending on its function (i. e. sister-chromatid cohesion or loop extrusion) and how does this affect the motion of the bound chromatin fiber? I set out to measure the characteristics of cohesin's motion on DNA by performing SPT of cohesin bound to DNA. Simultaneous imaging of DNA loci in the context of degron systems for regulators of cohesin's function in either sister-chromatid cohesion (Sororin) or loop extrusion (NIPBL) allowed me to study how cohesin's motion depends on the underlying chromatin fiber motion and how this is altered by depleting either of the aforementioned factors. Here, both functions of cohesin are shown contribute to its role in constraining chromosome motion. Furthermore, I could show that the motion of cohesin on DNA is slower than the motion on DNA, which indicates that the observed motion mostly describes the motion of loop-enforcing or "cohesive" cohesin. This work is described in chapter 3.

Chapter 2

Cohesin and CTCF control the dynamics of chromosome folding

Pia Mach^{1,2*}, Pavel I. Kos^{1*}, Yinxiu Zhan^{1*}, Julie Cramard¹, Simon Gaudin^{1,3}, Jana Tünnermann^{1,2}, Edoardo Marchi⁴, Jan Eglinger¹, Jessica Zuin¹, Mariya Kryzhanovska¹, Sebastien Smallwood¹, Laurent Gelman¹, Gregory Roth¹, Elphège P. Nora⁵, Guido Tiana⁴, Luca Giorgetti¹

¹ Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

² University of Basel, Basel, Switzerland

³ École Normale Supérieure de Lyon and Université Claude Bernard Lyon I, Université de Lyon, Lyon, France

⁴ Università degli Studi di Milano and INFN, Milan, Italy.

⁵ Cardiovascular Research Institute, University of California San Francisco; Department of Biochemistry and Biophysics, University of California San Francisco; San Francisco, CA, USA

Contribution

The contents of this chapter have been published in *Nature Genetics* on the 5th December 2022 (Mach *et al.*, 2022). This project was a collaboration with Pavel Kos and Yinxiu Zhan and the other authors listed in the manuscript.

My contributions included the conception of this study and the writing of the manuscript together with Pavel Kos, Yinxiu Zhan and Luca Giorgetti as well as the performance of most of the experiments and their interpretation. Namely, I designed and performed all genome engineering, with the technical help of Simon Gaudin, Julie Cramard and Mariya Kryzhanovska. I carried out the live-cell imaging experiments with help from Jana Tünnermann and Simon Gaudin and performed the Hi-C and 4C experiments.

* Pavel Kos, who performed and analyzed the polymer simulations, and Yinxiu Zhan, who performed image and trajectory analysis, contributed equally to this study.

2.1 Abstract

In mammals, interactions between sequences within topologically associating domains enable control of gene expression across large genomic distances. Yet it is unknown how frequently such contacts occur, how long they last and how they depend on the dynamics of chromosome folding and loop extrusion activity of cohesin. By imaging chromosomal locations at high spatial and temporal resolution in living cells, we show that interactions within topologically associating domains are transient and occur frequently during the course of a cell cycle. Interactions become more frequent and longer in the presence of convergent CTCF sites, resulting in suppression of variability in chromosome folding across time. Supported by physical models of chromosome dynamics, our data suggest that CTCF-anchored loops last around 10 min. Our results show that long-range transcriptional regulation might rely on transient physical proximity, and that cohesin and CTCF stabilize highly dynamic chromosome structures, facilitating selected subsets of chromosomal interactions.

2.2 Introduction

In mammalian cells, interactions between chromosomal sequences play important roles in fundamental processes such as DNA replication (Marchal *et al.*, 2019), repair (Arnould *et al.*, 2021) and transcriptional regulation by distal enhancers (Zuin *et al.*, 2022). 3C methods, which measure physical proximity between genomic sequences in fixed cells, revealed that chromosomal contacts are organized into sub-megabase domains of preferential interactions known as TADs (Dixon *et al.*, 2012; Nora *et al.*, 2012) whose boundaries can functionally insulate regulatory sequences (Zuin *et al.*, 2022). TADs mainly arise from nested interactions between convergently oriented binding sites of the DNA-binding protein CTCF, which are established as chromatin-bound CTCF arrests the loop extrusion activity of the cohesin complex (Fudenberg *et al.*, 2016; Y. Li *et al.*, 2020; Nora *et al.*, 2020; Davidson *et al.*, 2023).

Determining the timing and duration of chromosomal interactions within TADs and their relationship with CTCF and cohesin is key to understanding how enhancers communicate with promoters (Schoenfelder & Fraser, 2019; McCord *et al.*, 2020). Single-cell analyses of chromosome structure in fixed cells (Amano *et al.*, 2009; Nora *et al.*, 2012; Giorgetti *et al.*, 2014; Finn *et al.*, 2019), chromosome tracing experiments (Bintu *et al.*, 2018; Mateo *et al.*, 2019; M. Liu *et al.*, 2020; Huang *et al.*, 2021), *in vitro* (Davidson *et al.*, 2019; Davidson *et al.*, 2023; H. Zhang *et al.*, 2023) and live-cell (Hansen *et al.*, 2017) measurements of CTCF and cohesin dynamics, and polymer simulations (Barbieri *et al.*, 2012; Giorgetti *et al.*, 2014; Fudenberg *et al.*, 2016), as well as live-cell imaging of chromosomal locations and nascent RNA (H. Chen *et al.*, 2018; Alexander *et al.*, 2019), all suggested that TADs and CTCF loops are dynamic structures whose temporal evolution might be governed by the kinetics of loop extrusion (Hansen *et al.*, 2018). Recent live-cell measurements of a CTCF loop connecting two opposite TAD boundaries in mESCs provided direct evidence that this is the case, and revealed that

cohesin-mediated loops between CTCF sites located 500 kilobases (kb) away last 10-30 min (Gabriele *et al.*, 2022). However, it is still unclear if contacts between sequences separated by genomic distances where enhancers and promoters interact within the same TAD occur on the timescale of seconds, minutes or hours. We also have little knowledge on whether and how rates and durations of such contacts are modulated by loop extrusion. We finally do not know if cohesin increases chromosome mobility and thus favors the encounters between genomic sequences by reeling them into loops, or if instead it provides constraints that decrease mobility and prolong the duration of such encounters. Both scenarios have been suggested to be possible theoretically (Lawrimore *et al.*, 2016; Nuebler *et al.*, 2018), but it is unclear which effect dominates in living cells.

Here we use live-cell fluorescence microscopy to measure chromosome dynamics and its dependence on cohesin and CTCF in mESCs. By combining two live-cell imaging strategies with polymer simulations, we reveal that loops extruded by cohesin constrain global chromosome motion, while also increasing the temporal frequencies and durations of physical encounters between sequences inside the same TAD. Convergent CTCF sites substantially stabilize contacts through cohesin-mediated CTCF-anchored loops that last around 5-15 min on average. Our results support the notion that chromosome structure within single TADs is highly dynamic during the span of a cell cycle and thus that long-range transcriptional regulation might rely on transient physical proximity between genomic sequences. They also reveal how contact dynamics and the temporal variability in chromosome folding are modulated by cohesin and CTCF in single living cells and provide a quantitative framework for understanding the role of folding dynamics in fundamental biological processes.

2.3 Results

2.3.1 Cohesin decreases chromosome mobility independently of CTCF

To study how cohesin and CTCF influence the global dynamics of the chromatin fiber independently of local chromatin state and structural differences, we examined the dynamic properties of large numbers of random genomic locations in living cells. We generated clonal mESC lines carrying multiple random integrations of an array of ~ 140 repeats of the bacterial TetO sequence using piggyBac transposition (Cadiñanos & Bradley, 2007). These can be visualized upon binding of TetR fused to the red fluorescent protein tdTomato. To compare the motion of genomic locations that either block or allow the loop extrusion activity of cohesin, the TetO array was adjacent to three CTCF motifs ($3 \times \text{CTCF}$) that could be removed by Cre-assisted recombination (Fig. 2.1a). Motifs were selected based on high CTCF enrichment in ChIP-seq and each was confirmed to be bound by CTCF in nearly 100% of alleles at any time in mESCs using dual-enzyme single-molecule footprinting (Krebs *et al.*, 2017) (R. Grand and D. Schübeler, personal communication), thus providing a close experimental representative of an “impermeable” loop extrusion barrier.

$3 \times \text{CTCF}$ -TetO sequences were introduced in mESCs that stably expressed OsTir1 and where the

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

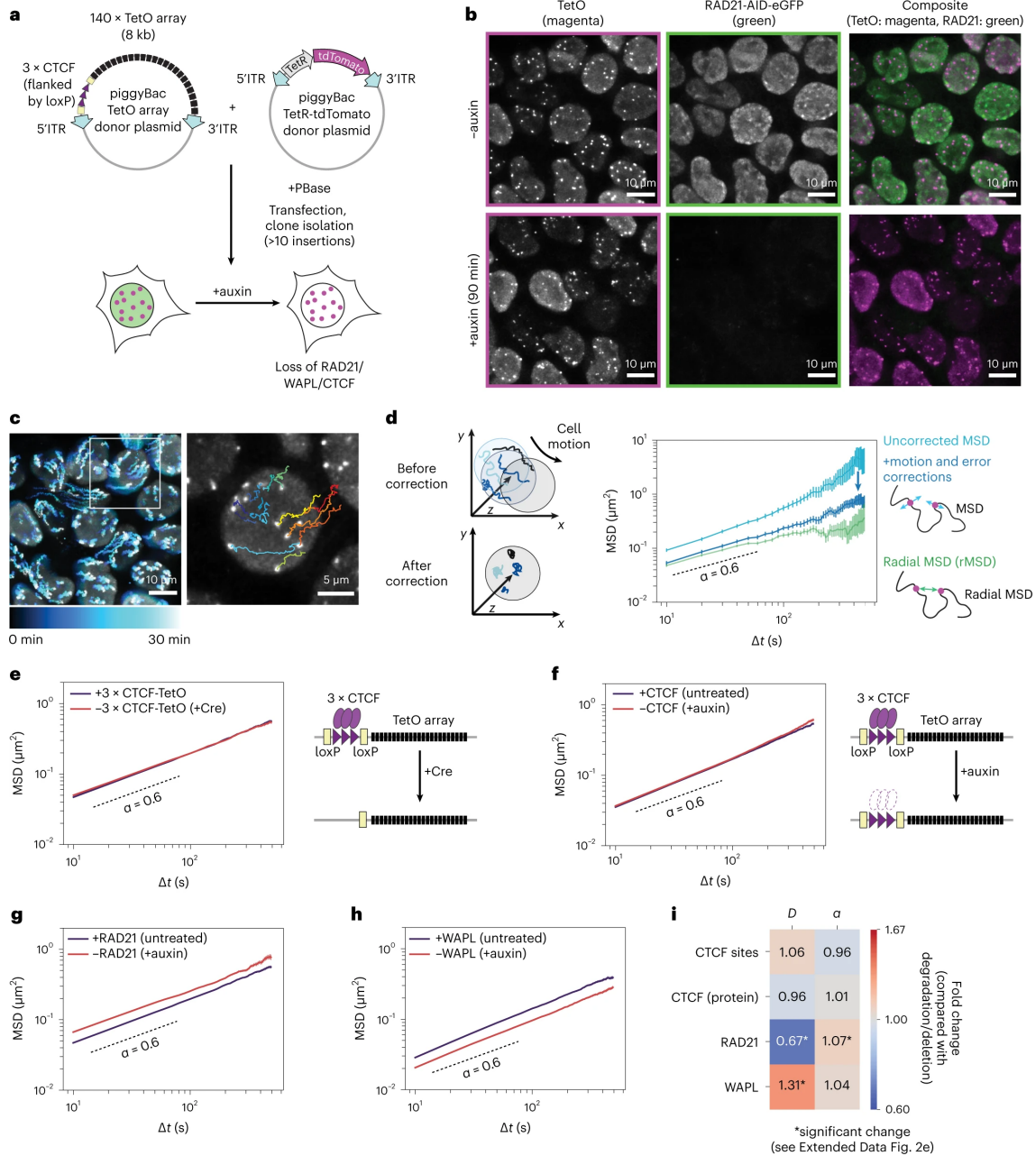


Figure 2.1: Cohesin slows down chromosome dynamics in living cells. **a**, Clonal mESC lines containing random TetO arrays flanked by $3 \times$ CTCF motifs and expressing TetR-tetTomato. Constructs were integrated using piggyBac transposition in mESCs allowing auxin-inducible degradation of GFP-tagged RAD21, WAPL or CTCF. ITR, inverted terminal repeats. **b**, Representative images of RAD21-AID-eGFP cells containing $3 \times$ CTCF-TetO imaged before or after 90 min of auxin treatment (exposure time eGFP and tetTomato: 50 ms, deconvolved, maximum intensity projection, bicubic interpolation, $n = 3$ replicates). **c**, Left, time series of TetR-tetTomato signal over 30 min (maximum intensity projection, time interval $dt = 10$ s, color-coded for intensity changes over time). Right, magnification with overlay of TetR-tetTomato signal with reconstructed trajectories of individual TetO arrays. **d**, Left, cell motion is approximated as the average roto-translational motion of TetO signals within the same nucleus. Right, MSD averaged over trajectories within one nucleus ($\text{mean} \pm \text{s.e.m.}$) before (cyan, $n = 77$) and after (blue, $n = 77$) cell motion and localization error correction. Green, radial MSD of pairs of operator arrays within the same nucleus ($\text{mean} \pm \text{s.e.m.}$, $n = 491$ pairs)

Figure 2.1: *continued from previous page: e*, Left, MSD (mean \pm s.e.m.) in mESC lines before (blue, 310 cells, 13,537 trajectories) or after (red, 271 cells, 11,082 trajectories) Cre-mediated removal of 3 \times CTCF sites. Three replicates per cell line and three lines per condition were analyzed and merged here and in all following MSD graphs. *P*-values (two-sided Student's t-test) for all panels shown in Extended Data Fig. E2e. Right, schematic representation of Cre-mediated removal of CTCF sites. **f**, Left, same as in e but in mESC lines with 3 \times CTCF-TetO arrays, before (blue, 323 cells, 9,829 trajectories) or after (red, 365 cells, 12,495 trajectories) CTCF degradation (6 h of auxin treatment). Right, schematic representation of auxin-induced CTCF degradation. **g**, MSD (mean \pm s.e.m.) of 3 \times CTCF-TetO insertions before (blue, 310 cells, 13,537 trajectories) or after (red, 240 cells, 8,788 trajectories) RAD21 degradation (90 min of auxin). **h**, MSD (mean \pm s.e.m.) of 3 \times CTCF-TetO before (blue, 336 cells, 6,687 trajectories) or after (red, 350 cells, 6,717 trajectories) WAPL degradation (24 h of auxin). **i**, Fold changes in generalized diffusion coefficients (*D*) and scaling exponents (α) in untreated cells compared with cells where degradation of CTCF, RAD21 and WAPL or removal of CTCF motifs (3 \times CTCF) occurred.

endogenous *Rad21*, *Wapl* or *Ctcf* genes were targeted with an AID peptide fused to eGFP (Nora *et al.*, 2017; N. Q. Liu *et al.*, 2021). This resulted in several mESC clones (three per degron condition) with different sets of genomic insertions of the 3 \times CTCF-TetO cassette, where over 95% of any of the AID-tagged proteins could be rapidly depleted upon addition of auxin (Fig. 2.1b and Extended Data Fig. E1a). This allowed us to study chromosome dynamics following acute depletion of factors affecting cohesin-mediated chromosome structure (Extended Data Fig. E1b) at previously reported time points (90 min for RAD21 (S. Rao *et al.*, 2017), 6 h for CTCF (Nora *et al.*, 2017) and 24 h for WAPL (N. Q. Liu *et al.*, 2021)) that minimize secondary effects such as defects in cell-cycle progression (Extended Data Fig. E1c).

Mapping TetO insertion sites revealed 10-20 insertions per cell line, with on average 1-2 heterozygous insertions per chromosome without any strong bias towards active or inactive chromatin (Extended Data Fig. E1d,e). Insertions were on average 10 kb away from the nearest endogenous CTCF binding sites (Extended Data Fig. E1f). 4C confirmed that insertion of 3 \times CTCF-TetO cassettes often led to the formation of ectopic interactions with endogenous CTCF sites, which were lost upon removal of 3 \times CTCF sites or depletion of RAD21 (Extended Data Fig. E1g,h).

To measure the dynamics of 3 \times CTCF-TetO insertions, we acquired 3D movies (one z-stack of 10 μ m every 10 s for 30 min) using highly inclined and laminated optical sheet microscopy (Tokunaga *et al.*, 2008) (Fig. 2.1b and Supplementary Video S1). This resulted in \sim 270 cells per condition with over 8,000 trajectories from three clonal lines imaged with 3-4 biological replicates per condition. Detection and localization of TetO arrays as subdiffraction fluorescent signals (Eichenberger *et al.*, 2021) enabled reconstruction of trajectories of individual genomic insertions (Fig. 2.1c and Methods). We then studied their mean squared displacement (MSD) as a function of time after correcting each trajectory for the confounding effect of cell movement, which we inferred from the collective displacement of all insertions in each nucleus (Fig. 2.1d, Extended Data Fig. E2a and Methods). Independently of the degron background, in untreated cells, genomic locations underwent on average a subdiffusive motion whose anomalous exponent (\sim 0.6) and generalized diffusion coefficients (*D*) (\sim 1.2 \times 10⁻² μ m² s^{- α}) were in line with previous studies of specific genomic loci (Gu *et al.*, 2018; Khanna *et al.*, 2019) (Fig. 2.1d and Extended Data Fig. E2b). The MSD of radial distances (radial MSD) between insertions within the same nuclei showed the same scaling although statistics were less robust for

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

long time intervals due to the shorter trajectories that could be built based on pairwise distances (Fig. 2.1d). Interestingly, removal of $3 \times$ CTCF sites (Extended Data Fig. E1g) or degradation of CTCF (6 h auxin treatment) did not have a significant impact on MSD averaged over all genomic locations nor on its distribution across trajectories and cells (Fig. 2.1e,f, results for single clones in Extended Data Fig. E2c, P -values in Extended Data Fig. E2e).

By contrast, acute depletion of RAD21 (90 min auxin treatment) led to a significant increase in mobility both in the presence (Fig. 2.1g) and absence of $3 \times$ CTCF sites (Extended Data Fig. E2d), with only a very minor impact on anomalous exponents (Extended Data Fig. E2b,e, P -values in Extended Data Fig. E2e). In the presence of wild-type levels of RAD21, generalized diffusion coefficients were on average $\sim 30\%$ lower than in depleted cells, where RAD21 levels were low enough to prevent formation of cohesin-mediated structures (compare with Extended Data Fig. E1b). This outcome was consistent across three clonal cell lines with different TetO insertion sites and the small differences in the magnitude of the effect were likely due to location-dependent effects (Extended Data Fig. E2f). Importantly, the effect was specific for RAD21 degradation as we did not observe any changes in MSD behavior in control cell lines expressing OsTir1 but no AID-tag (Extended Data Fig. E2g). In addition, depletion of WAPL (24 h auxin treatment), which results in higher levels of DNA-bound cohesin (N. Q. Liu *et al.*, 2021), caused a substantial decrease in chromosome mobility (Fig. 2.1h and Extended Data Fig. E2e). Together, these results indicate that increasing levels of DNA-bound cohesin decrease chromosome mobility, with only very minor effects (if any) mediated by the presence of even strong CTCF motifs (Fig. 2.1i).

2.3.2 Loop extrusion can explain reduced chromosome dynamics

We next used polymer simulations to determine if loop extrusion alone could explain the observed global reduction in chromosome dynamics in the presence of cohesin and minimal effects from CTCF. We simulated the dynamics of a polymer with excluded volume, with or without loop extrusion and extrusion barriers whose linear arrangement and orientation were sampled from endogenous CTCF sites (Fig. 2.2a and Extended Data Fig. E3a). To mimic random insertion of $3 \times$ CTCF sites, we also simulated the same polymers with additional loop extrusion barriers separated by 800 kb which were inserted at random positions in the polymer (magnified area in Fig. 2.2a). To emphasize their potential effects on chromosome dynamics, all barriers in the simulations were impermeable to loop extruders. Every monomer represented 8 kb of chromatin, corresponding to the genomic size of the TetO array. Simulation steps were approximated to real-time units by matching the time needed for a monomer to move by its own diameter with the time required by the TetO array to move by its estimated mean physical size (Methods). We sampled an extremely large range of extruder residence times and loading rates (4 orders of magnitude each) centered around a residence time of ~ 30 min and extruder densities of ~ 20 per Mb (in line with previous measurements (Cattoglio *et al.*, 2019; Holzmann *et al.*, 2019)), and using two extrusion speeds corresponding to *in vivo* and *in vitro* estimates ($\sim 0.1 \text{ kb s}^{-1}$ and $\sim 1 \text{ kb s}^{-1}$, respectively) (Davidson *et al.*, 2019) (Fig. 2.2b and Extended Data Fig. E3b).

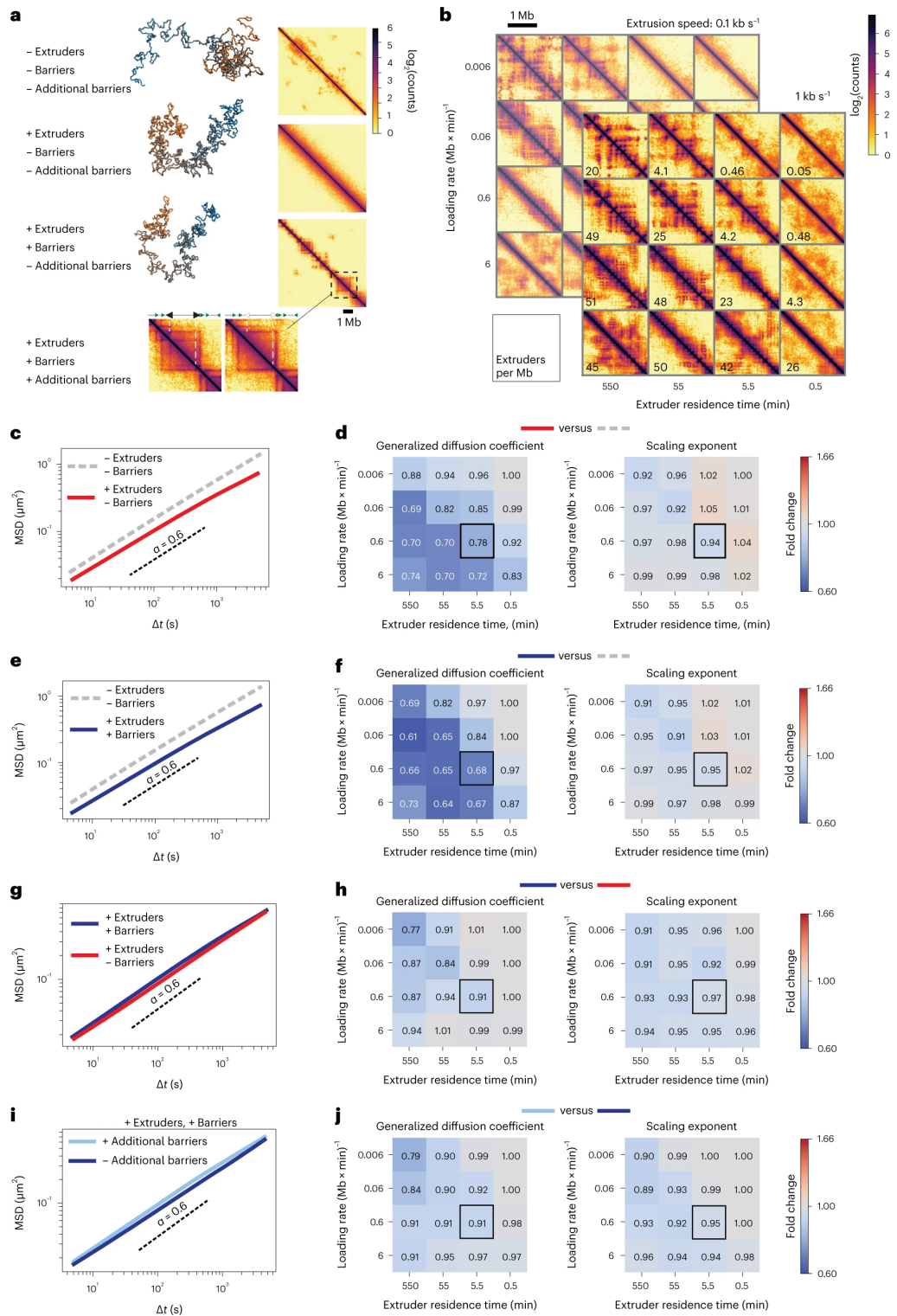


Figure 2.2: Loop extrusion generally slows down polymer motion. **a**, Representative snapshots of conformations and simulated contact maps for a polymer model with excluded volume and increasingly complex models with loop extruders, extrusion barriers sampled from CTCF motifs within 9 Mb on chromosome 15 (Chr15:7–16 Mb) and additional randomly distributed extrusion barriers.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

Figure 2.2: *continued from previous page:* For the system with additional barriers, the contact map is presented aside with magnification of the contact map of the system without additional barriers to highlight the differences. **b**, Simulated contact maps (with loop extrusion and extrusion barriers) for polymers with two extrusion speeds (1 kb s^{-1} and 0.1 kb s^{-1}) and different combinations of extruder loading rates and residence times. The resulting linear densities of extruders (number per Mb) are shown in the bottom left corner of each contact map. **c**, Effect of extruders. MSDs of polymers with (red line) or without (gray dashed line) loop extruders in the absence of extrusion barriers (loading rate $0.6 (\text{Mb} \times \text{min})^{-1}$ and residence time 5.5 min, corresponds to black square in panel d). Black dashed curve represents $\alpha = 0.6$ as an eye guide. **d**, Effect of extruders. Ratios of generalized diffusion coefficients and anomalous exponents between the two conditions shown in panel c. Black square, set of parameters whose corresponding MSDs are shown in panel c. **e**, MSDs of polymers with (blue line) or without (gray dashed line) both extruders and barriers. Same parameters as in panel c. **f**, Same as panel d for cases illustrated in panel e. **g**, MSDs of polymers with loop extruders in the presence (blue) or absence (red) of extrusion barriers. Same parameters as in panels c and e. **h**, Same as panels d and f but for cases illustrated in panel g. **i**, MSDs of polymers either with (light blue) or without (red) additional randomly inserted extrusion barriers. Same parameters as in panels c, e, g. **j**, Same as panels d, f and h but for cases illustrated in panel i.

In the absence of loop extrusion, the polymer underwent subdiffusive behavior with anomalous exponent of ~ 0.6 (Fig. 2.2c), as expected from simple polymers with excluded volume (Dünweg & Kremer, 1993; Tamm & Polovnikov, 2017; Salari *et al.*, 2022) (see Supplementary Information) and compatible with our experimental results on random TetO insertions (Fig. 2.1e). Strikingly, in line with experimentally measured effects of RAD21 (Fig. 2.1g), introduction of loop extrusion led to lower generalized diffusion coefficients and minor effects on anomalous exponents, independently of loading rate and residence time (Fig. 2.2c,d), extrusion speed (Extended Data Fig. E3c) or the presence of extrusion barriers (Fig. 2.2e,f). Interestingly, for extruder residence times of 5.5–11 min and unloading rates corresponding to extruder linear densities of ~ 20 per Mb, the predicted decrease in generalized diffusion coefficients was in quantitative agreement with the experimentally observed value of $\sim 30\%$ (Fig. 2.2d,f; extruder densities as in Fig. 2.2b; compare with Fig. 2.1g). Also, consistently with WAPL depletion experiments (Fig. 2.1h), increasing extruder residence times systematically resulted in larger reductions in generalized diffusion coefficients (Fig. 2.2d,f and Extended Data Fig. E3c).

Importantly, addition of barriers in the presence of loop extrusion led to substantially smaller changes in polymer dynamics compared with the effect of loop extrusion itself even when probed directly on the barriers (Fig. 2.2g,h and Extended Data Figs. E3c and E4a,b), in agreement with our experimental finding that CTCF degradation had no strong effect on MSDs of TetO insertions (Fig. 2.1e,f and Extended Data Fig. E2c,e). Similarly, insertion of additional barriers had little impact on MSD (Fig. 2.2i,j), thus recapitulating the negligible effect of removal of $3 \times$ CTCF sites (Fig. 2.1e). Polymer simulations thus strongly support the notion that the observed decrease in chromosome mobility and lack of effects from CTCF is a macroscopic manifestation of the physical constraints imposed by cohesin in living cells.

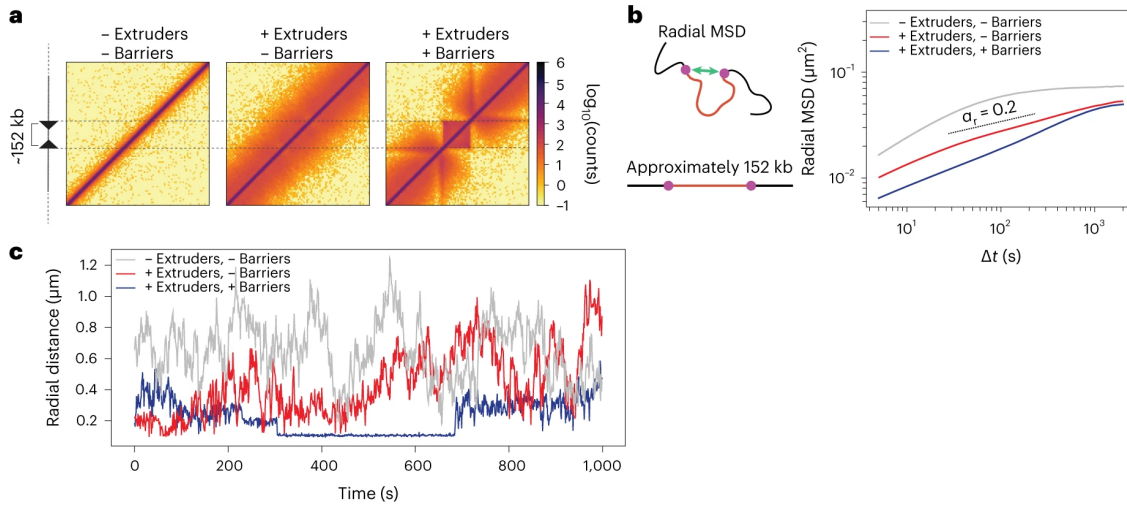


Figure 2.3: Convergent CTCF sites further constrain polymer dynamics. **a**, Simulated contact maps of a region spanning the equivalent of 800 kb for a polymer chain without loop extrusion, with loop extruders and with convergent extrusion barriers separated by the equivalent of 152 kb. **b**, Radial MSD of the two monomers separated by the equivalent of 152 kb in the three conditions from panel **a**. Dashed line is an exponent of 0.2 as a guide to the eye (α_r indicates the slope of radial MSDs). Loop extrusion parameters as in Fig. 2.2c. **c**, Representative examples of distances between the two monomers in simulations with or without loop extrusion and extrusion barriers. The flat stretch in the trajectory with extrusion and barriers corresponds to a loop anchored by the two barriers.

2.3.3 Cohesin and CTCF constrain the dynamics of sequences in *cis*

We next asked how cohesin and CTCF impact the reciprocal motion of two genomic sequences located on the same DNA molecule. To this aim, we simulated the dynamics of a polymer carrying two convergent impermeable extrusion barriers mimicking strong CTCF motifs separated by ~ 150 kb (Fig. 2.3a). This is comparable to median distances between convergent CTCF sites within TADs genome-wide in mESCs (141 kb, Methods) and also to the estimated average separation between enhancers and promoters in human cells (~ 160 kb) (Jung *et al.*, 2019). Simulations performed with extrusion parameters recapitulating the dynamic effects of RAD21 depletion (black square in Fig. 2.2f) predicted that radial MSDs should be lowest in the presence of loop extrusion and barriers (Fig. 2.3b) due to the formation of transient loops anchored by the barriers (Fig. 2.3c). Similar to MSDs (Fig. 2.2), radial MSDs should increase upon removal of extrusion barriers and become maximal when loop extrusion is also removed (Fig. 2.3b).

Importantly, simulations also predicted that scaling exponents of radial MSD curves should be considerably smaller (~ 0.2) than those we previously observed for TetO arrays separated by several Mb or located on different chromosomes (~ 0.6 , Fig. 2.1d). This is because correlations in the motion of two monomers are stronger when they are located closer along the polymer. Indeed, simulations predicted that scaling exponents fitted from radial MSD curves at short times should increase with increasing genomic distance and approach 0.6 for loci separated by several Mb (consistent with radial MSDs of randomly inserted TetO arrays) (Extended Data Fig. E5a) before saturating to stationary values at longer times. This holds true also without loop extrusion (theoretical analysis in

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

Supplementary Information and simulations in Extended Data Fig. E5b).

To test these predictions, we turned to a live-cell imaging approach allowing us to measure the radial dynamics of two sequences located within the same TAD, in the presence and absence of cohesin and/or strong CTCF sites. We engineered mESCs carrying targeted integrations of two orthogonal operator arrays: $\sim 140 \times$ TetO and $120 \times$ LacO separated by 150 kb (Fig. 2.4a), which could be visualized upon binding of TetR-tdTomato and a weak DNA-binding variant of LacI^{**} fused to eGFP (LacI^{**}-eGFP) (Dubarry *et al.*, 2011). To minimize confounding effects from additional regulatory sequences such as active genes or enhancers, we targeted the arrays into a 560 kb “neutral” TAD on chromosome 15 where we previously removed internal CTCF sites (Zuin *et al.*, 2022) (Fig. 2.4a). The two operator arrays were directly adjacent to excisable $3 \times$ CTCF site cassettes arranged in a convergent orientation (Fig. 2.4a). Cell lines were verified by Nanopore Cas9-targeted sequencing (nCATS) (Gilpatrick *et al.*, 2020) to contain a single copy of each targeting cassette (Extended Data Fig. E5c). We additionally targeted the endogenous *Rad21* locus with a C-terminal HaloTag-FKBP fusion allowing the inducible degradation of RAD21 upon treatment with dTAG-13 (Nabet *et al.*, 2018) as confirmed by severely decreased protein levels ($>95\%$ after 2 h treatment, Extended Data Fig. E5d).

Capture-C with tiled oligonucleotides revealed that integration of operator arrays themselves did not lead to detectable changes in chromosome structure (Extended Data Fig. E5e). Convergent $3 \times$ CTCF sites, however, led to the formation of a new CTCF-mediated interaction within the TAD ($2.8 \times$ increase in contact probability after correcting the confounding contribution of the wild-type allele) (Fig. 2.4b), which was lost upon RAD21 depletion along with all other CTCF-mediated interactions across the locus (Extended Data Fig. E5f).

We imaged cells for 3 h every 30 s in three dimensions (Fig. 2.4c and Supplementary Video S2), either in the presence or absence of RAD21, and measured distances between the two arrays over time (Fig. 2.4d, $n = 3-7$ biological replicates for each condition, on average 220 cells per condition, Supplementary Table S1 and Methods). Doublet signals corresponding to replicated alleles occurred in a very minor fraction (3%, Methods) of trajectories, compatible with the late-replication profile of the “neutral” TAD and the cell-cycle distribution (Extended Data Figs. E5g and E6a). In these cases, only trajectories that were initially closest across channels were considered. After correction of chromatic aberrations (Methods and Extended Data Fig. E6b), we estimated our experimental uncertainty on radial distances to be ~ 130 nm by measuring pairwise distances in control cells where multiple TetO insertions were simultaneously bound by both TetR-tdTomato and TetR-eGFP (Extended Data Fig. E6c-e).

In agreement with model predictions for locations separated by 150 kb, radial MSDs of the two arrays showed scaling exponents close to 0.2, much smaller than those observed with randomly inserted TetO arrays (Extended Data Fig. E7a,b). Also in line with model predictions (Extended Data Fig. E5a), presence of RAD21 and $3 \times$ CTCF sites led to the most constrained radial mobility, whereas RAD21 degradation and deletion of CTCF sites resulted in the least constrained motion (Extended Data Fig. E7c). These measurements thus verified the model prediction that genomic sequences located at short

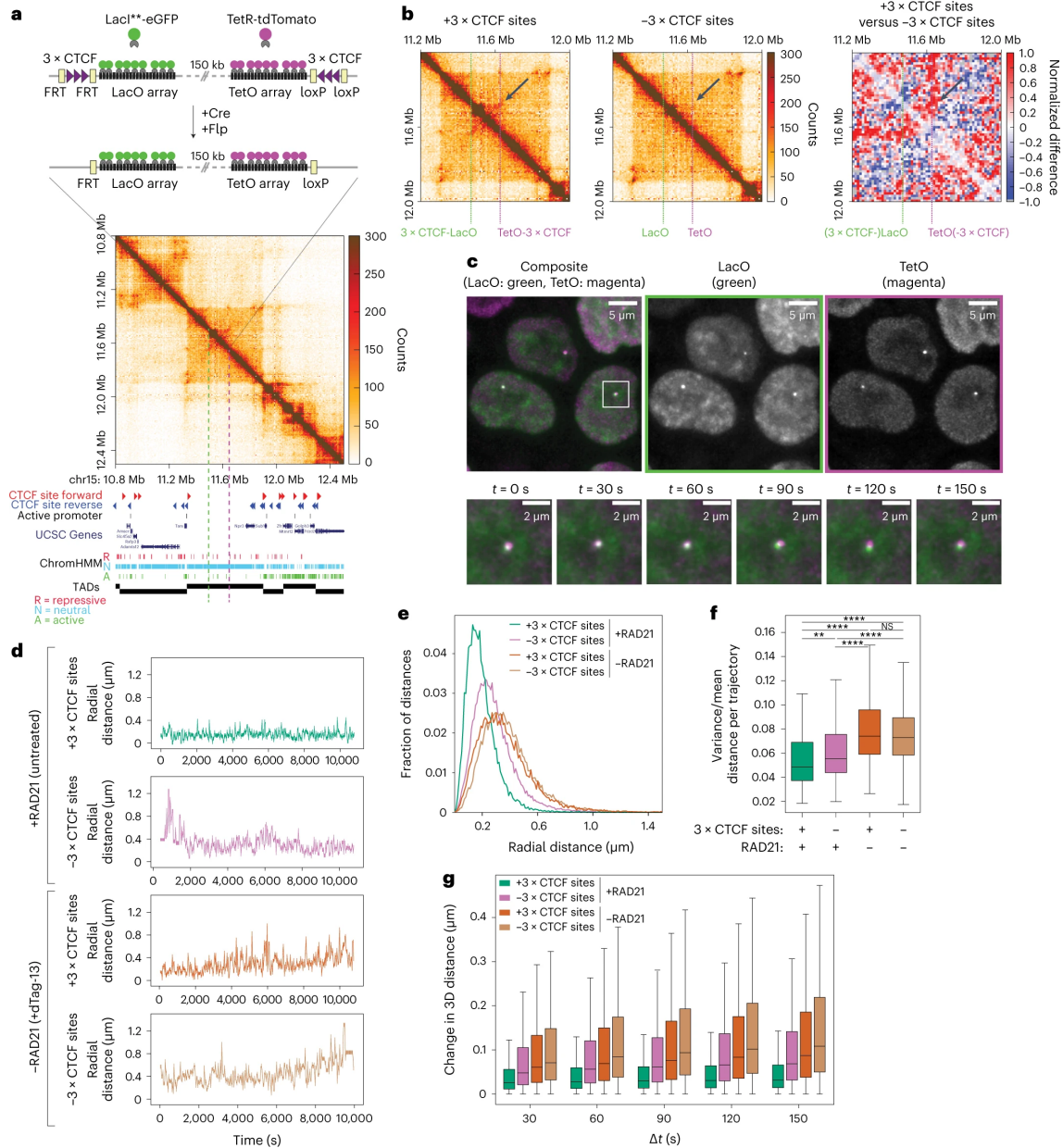


Figure 2.4: Cohesin and CTCF reduce variability in chromosome folding dynamics. **a**, Top, insertion of TetO and LacO arrays separated by 150 kb within a “neutral” TAD on chromosome 15 in mESCs. Flanking $3 \times$ CTCF sites can be excised by Cre and Flp recombinases. Arrays are visualized by binding of $\text{LacI}^{**}\text{-eGFP}$ and TetR-tdTomato , respectively. Bottom, tiled Capture-C map (6.4 kb resolution) and genomic datasets in mESCs in a region in 2.6 Mb surrounding the engineered TAD. Capture-C was performed in cells where arrays were flanked by $3 \times$ CTCF sites. Dashed lines, positions of LacO and TetO insertions. **b**, Capture-C maps in mESC lines with (left) or without (middle) $3 \times$ CTCF sites flanking TetO and LacO arrays, and differential map (right, $+3 \times$ CTCF versus $-3 \times$ CTCF, Methods) highlighting interactions formed between convergent $3 \times$ CTCF sites (arrows).

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

Figure 2.4: *continued from previous page:* **c**, Top, representative fluorescence microscopy images of mESCs with 3×CTCF-LacO and TetO-3×CTCF insertions. Bottom, magnified view with time series overlay of LacI^{**}-eGFP and TetR-tdTomato signals (exposure time 50 ms, deconvolved, maximum intensity projection, bicubic interpolation). **d**, Representative trajectories of TetO-LacO radial distances with or without convergent 3×CTCF sites, either before or after degradation of RAD21 (2 h of dTag-13) (dt = 30 s). **e**, Distribution of TetO-LacO radial distances in the four experimental conditions (+3×CTCF sites/+RAD21: n = 152 cells, 4 pooled replicates; -3×CTCF sites/+RAD21: n = 214 cells, 4 pooled replicates; +3×CTCF sites/-RAD21: n = 248 cells, 7 pooled replicates; -3×CTCF sites/-RAD21: n = 277 cells, 6 pooled replicates). **f**, Distributions of variance over mean within single trajectories across the four experimental conditions (no. of cells as in panel e). Boxes, lower and upper quartiles (Q1 and Q3, respectively). Whiskers denote 1.5× interquartile region (IQR) below Q1 and above Q3. *P*-values are calculated using two-sided Kolmogorov–Smirnov test. NS, not significant; ** = $p < 0.01$, *** = $p < 0.0001$. Exact *P*-values can be found in Supplementary Table S2. Outliers are not shown. **g**, Distribution of jump step size (changes in TetO-LacO radial distance) across increasing time intervals for the four experimental conditions (no. of cells as in panel e). Boxes, lower and upper quartiles (Q1 and Q3, respectively). Whiskers, 1.5×IQR below Q1 and above Q3. Outliers are not shown.

distances (150 kb) experience stronger physical constraints than sequences located at larger genomic distances (Gabriele *et al.*, 2022) (Fig. 2.1 and Extended Data Fig. E7b), and that loop extrusion provides constraints that are further reinforced by convergent CTCF sites.

Consistently with their more constrained radial MSD behavior, we finally observed that distances between TetO and LacO signals were smallest in the presence of convergent CTCF sites and cohesin. In these conditions, distances between TetO and LacO arrays tended to remain close to the ~130 nm experimental uncertainty with only occasional fluctuations toward larger values in the course of the 3 h of imaging (Fig. 2.4d,e). Removal of 3×CTCF sites led to increased radial distances and variability within single trajectories, which were further increased upon degradation of RAD21, irrespective of the presence or absence of CTCF sites (Fig. 2.4d,e and Supplementary Video S3). Thus, constraints imposed by extruding cohesin and convergent CTCF sites reduce not only average physical distances between sequences but also their variability in time (Fig. 2.4f), also supported by analysis of distance changes (jumps) as a function of time (Fig. 2.4g).

Finally, to test whether the effects of cohesin on chromosome motion would be different in the presence of active transcription at nearby locations, we measured looping dynamics this time in a parental mESC line before the removal of resistance cassettes. In this line, both the TetO and LacO arrays were immediately flanked by mouse *Pgk1* promoters (Adra *et al.*, 1987) driving the transcription of resistance genes (Extended Data Fig. E6f,g). In line with previous studies (Germier *et al.*, 2017; Nozaki *et al.*, 2017), we found that active transcription led to slightly decreased radial MSD. Cohesin depletion resulted in similar amounts of increased radial mobility irrespective of the presence or absence of active promoters (Extended Data Fig. E7c).

2.3.4 Chromosomal contacts are transient

We next set off to quantify changes in distances over time and determine whether despite the experimental uncertainty on 3D distances (Extended Data Fig. E6e) we could observe transitions between two states: a “proximal” state with small radial distances (presumably including cohesin-mediated

loops between convergent CTCF sites), and a generic “distal” state with larger spatial distances corresponding to other configurations of the chromatin fiber. This was motivated by the expectation that any polymer with site-specific attractive interactions, such as those mediated by cohesin at convergent CTCF sites, should in principle result in two-state thermodynamic behavior. We thus fitted a two-state hidden Markov model (HMM) on the ensemble of trajectories obtained in cells where both convergent $3\times$ CTCF sites and RAD21 were present (Fig. 2.5a,b). Interestingly, distances in the proximal state inferred by HMM largely overlapped with those detected on perfectly colocalizing signals in control experiments where TetR-eGFP and TetR-tdTomato were bound to the same set of randomly inserted TetO arrays (149 versus 130 nm on average, respectively) (Fig. 2.5b and Extended Data Fig. E6e). The proximal state thus corresponds to configurations of the chromatin fiber where the two arrays were in very close physical proximity, also including (but not restricted to) cohesin-mediated loops between CTCF sites. For simplicity, we refer to the proximal state interchangeably as “contact”, without implying a direct molecular interaction between the two DNA fibers. Radial distances in the distal state (288 nm on average) instead were similar to those measured in cells where both CTCF sites had been removed (291 nm) (Extended Data Fig. E8a,b). Thus, the distal state largely overlapped with chromosome conformations where specific cohesin-mediated CTCF loops were lost.

We next fitted the HMM to all experimental conditions while keeping the same proximal state as in cells with $3\times$ CTCF sites and RAD21 (Fig. 2.5a). This showed that in the presence of RAD21, the LacO and TetO arrays spent $\sim 78\%$ of the time in contact (that is, in the proximal state) when $3\times$ CTCF sites were present. This was $2.3\times$ higher than the 33% of time they spent in contact when the $3\times$ CTCF sites were removed (Fig. 2.5c), in agreement with the corresponding 2.8-fold difference in contact probability inferred from Capture-C (Fig. 2.4b). The fraction of time spent in contact decreased markedly upon depletion of RAD21 to $\sim 23\%$ in the presence of $3\times$ CTCF sites and 11% in the absence (Fig. 2.5c). Both the average duration of contacts and their rate of formation were maximal in the presence of RAD21 and $3\times$ CTCF sites, where they lasted around 16 min and reformed every 5 min on average (Fig. 2.5d,e). Contacts became substantially shorter (6 min) and rarer (one every 10 min) when $3\times$ CTCF sites were removed, and even more so upon RAD21 depletion (lasting 2 min and occurring every 22 min on average). Interestingly, these results were not affected by the presence of actively transcribed promoters in the immediately flanking regions (Extended Data Fig. E8b–f), in line with the lack of changes in contact probability measured in Capture-C (Extended Data Fig. E6g). Thus, both cohesin and CTCF impact both the duration and the probability of formation of chromosomal contact events between loci separated by 150 kb within an “empty” TAD.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

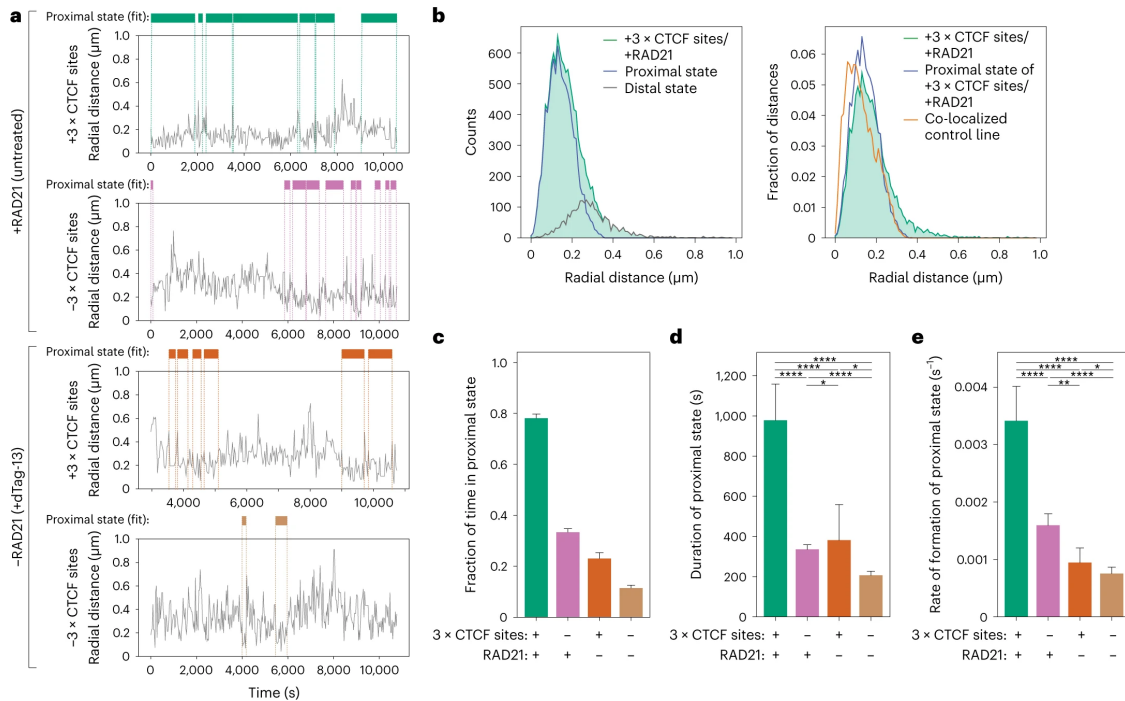


Figure 2.5: Cohesin and CTCF control contact dynamics inside a TAD. **a**, Representative trajectories of radial distance (gray) and occurrences of the proximal state called by HMM (colored bars). The HMM was fitted on data with convergent $3 \times \text{CTCF}$ sites and RAD21 (top left) to find the proximal state which was then imposed on the other three samples. **b**, Left, radial distance distribution in cells with convergent $3 \times \text{CTCF}$ sites and RAD21 overlaid with those of proximal and distal states called by HMM on the same sample. Right, same as in the left panel but normalized and with the additional display of the distance distribution from a control cell line where TetO and LacO signals perfectly co-localize. **c**, Fraction of time spent in the proximal state called by HMM in the four experimental conditions (no. of replicates as indicated in Fig. 2.4e). Shown are averages across experimental conditions; error bars represent bootstrapped ($n = 10,000$) standard deviations. **d**, Average durations of proximal states (mean \pm 95% confidence interval (CI), $n = 680$ ($-3 \times \text{CTCF}/+\text{RAD21}$); $n = 287$ ($+3 \times \text{CTCF}/+\text{RAD21}$); $n = 268$ ($-3 \times \text{CTCF}/-\text{RAD21}$); $n = 114$ ($+3 \times \text{CTCF}/-\text{RAD21}$)). P -values (two-sided Kolmogorov–Smirnov): * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Exact P -values can be found in Supplementary Table S2. **e**, Average rates of contact formation—time elapsed between the end of a proximal state and the beginning of the next (mean \pm 95% CI, $n = 726$ ($-3 \times \text{CTCF}/+\text{RAD21}$); $n = 323$ ($+3 \times \text{CTCF}/+\text{RAD21}$); $n = 268$ ($-3 \times \text{CTCF}/-\text{RAD21}$); $n = 138$ ($+3 \times \text{CTCF}/-\text{RAD21}$)). P -values as in panel d.

To understand if these results could be rationalized in terms of loop extrusion, we compared them with polymer simulations with convergent impermeable loop extrusion barriers separated by ~ 150 kb. Simulations were performed using loop extrusion parameters spanning a finer-grained 25-fold range around experimentally realistic values that reproduced the dynamic effect of RAD21 degradation (compare with Fig. 2.2d, black square) and with both *in vitro* and *in vivo* estimates of extrusion speeds (Cattoglio *et al.*, 2019; Davidson *et al.*, 2019). In a large region of the parameter space, distances between convergent barriers were bimodally distributed, supporting the expectation that the polymer can be approximated as a two-state system (Extended Data Fig. E9a). To allow direct comparison with experimental distance-based HMM states, we applied random errors matching experimental uncertainty levels to radial distances generated by the models (Extended Data Fig. E9b). We called proximal and distal states using the same HMM strategy as with experimental data. Importantly, for a large number of parameter combinations, distances in the proximal state largely overlapped with the

corresponding distribution observed experimentally in the presence of convergent CTCF sites and cohesin (Extended Data Fig. E9c and Supplementary Fig. S3a).

We then compared the distance, duration and fraction of time spent in the proximal state with those experimentally observed in the presence of RAD21 with or without $3 \times$ CTCF sites. We found that their similarity was maximal for extruder densities ranging from 8 to 32 per Mb (Supplementary Fig. S3e) and residence times of 2.8–11 min, with extrusion speeds of both 0.1 and 1 kb s^{-1} , all of which were in the range of previous estimations of experimental values (Hansen *et al.*, 2017; Cattoglio *et al.*, 2019; Davidson *et al.*, 2019; Wutz *et al.*, 2020) (Fig. 2.6a and Extended Data Fig. E9d). Considering the five best-matching scenarios (red- and yellow-marked values in Fig. 2.6a), the two locations spent 45–55% of the time in the proximal state with an average contact duration of around 10–17 min, which reduced to 18% and 8 min in the absence of extrusion barriers (Fig. 2.6b,c). Similar to the effects observed experimentally upon depletion of RAD21, decreasing extruder densities (for example, by decreasing loading rates) led to decreased fractions of time and shorter durations of the proximal state (Fig. 2.6d,e, shown for the best case, general trends in Supplementary Fig. S3d). Thus, the duration and the fraction of time spent in the proximal state, and most importantly how these quantities change upon removing cohesin and/or CTCF sites, can be understood in terms of a simple loop extrusion model.

The HMM-based proximal state likely provides an overestimation of the duration of underlying CTCF-CTCF loops mediated by stalled cohesins, since it also contains a fraction of CTCF-independent proximity events that cannot be distinguished from loops. To estimate the duration and times the two loci spent in a cohesin-mediated CTCF-CTCF looped conformation, we quantified occurrences in the simulated polymer where the two monomers formed the base of an extruded loop (Fig. 2.6f and Methods). As expected, these events were rarer and shorter than contacts detected by HMM on polymer simulations (Fig. 2.6b,c), with two monomers spending ~ 20 – 31% of time at a loop base for 5–15 min on average in the presence of extrusion barriers (Fig. 2.6g,h). Finally, transient cohesin-dependent loops that are not stabilized by CTCF sites should occur much more rarely (1–3% of the time) and lasted less than a minute on average (Fig. 2.6g,h). Comparison of polymer simulations with HMM states thus suggests that the dynamics of chromosome contacts detected at a range of 150 nm are generated by faster and rarer cohesin-mediated CTCF loops (Fig. 2.6i).

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

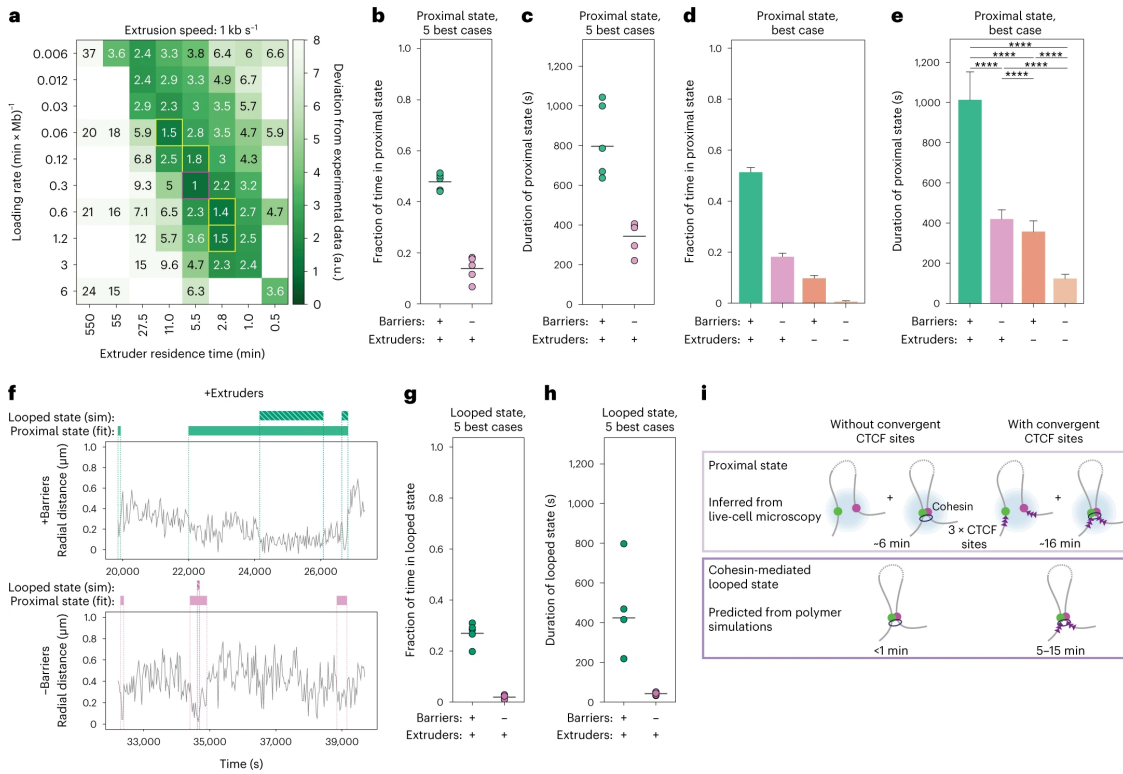


Figure 2.6: Estimation of frequency and duration of cohesin-mediated CTCF loops. **a**, Levels of agreement between simulations and experimental data as a function of loop extrusion parameters (here shown with extrusion speed 1 kb s⁻¹). The score represents the deviations of the distance, duration and fraction of time spent in the proximal state with those experimentally observed in the presence of RAD21 with or without 3 × CTCF sites (Methods). Magenta square, parameter set maximizing the agreement with experimental values. Yellow squares, four additional second-best parameter sets. **b**, Fraction of time spent in the proximal state called by HMM on simulations with the five best-matching parameters (magenta and yellow squares in panel a) for +Extruder case, Methods). **c**, Average duration (mean ± 95% CI) of proximal state called by HMM on simulations with the five best-matching parameters. **d**, Fraction of time spent in the proximal state called by HMM on simulations (over n = 15,880 time points) for the best-matching parameter set in the presence of extruders (+) or low levels (-) of extruders, either with or without extrusion barriers. Shown are averages across experimental conditions; error bars represent bootstrapped (n = 10,000) standard deviations. **e**, Average duration of the proximal state (mean ± 95% CI, over n = 15,880 time points) either in the presence of extruders (+) or low levels of extruders (-), either with or without extrusion barriers. Two-sided Kolmogorov–Smirnov *P*-values can be found in Supplementary Table S2. **f**, Representative trajectories of radial distances (gray), contact states called by HMM (full bar) and looped states in the underlying polymer conformations (striped bars) from +Extruders/+Barriers (top) and +Extruders/-Barriers simulations (bottom) with best-matching parameters (magenta square in panel a). **g**, Fraction of time spent in the looped state based on simulations with the five best-matching parameters. **h**, Average duration of the looped state based on simulations with the five best-matching parameters (mean ± 95% CI). **i**, Scheme summarizing the durations of proximal and looped states in the presence and absence of 3 × CTCF sites. a.u., arbitrary unit; sim, simulation.

2.4 Discussion

Our study provides quantitative measurements of chromosome folding dynamics in living cells and reveals how they are controlled by cohesin and CTCF. Two experimental strategies allow us to minimize biological variation from specific regulatory and structural genomic contexts and enable direct comparison with polymer models. By studying large numbers of random genomic locations, we average over local differences in chromosome mobility and reveal the global dynamic effects of cohesin. By visualizing and manipulating two locations within a “neutral” genomic environment, we unravel how cohesin and CTCF impact chromosome looping within a single TAD. We show that although higher extrusion speeds could in principle result in increased chromosome motion (Extended Data Fig. E9e,f), physiological extrusion rates rather generate transient constraints that decrease chromosome dynamics, in line with previous measurements of histone mobility (Nozaki *et al.*, 2017). Similar to previous reports (Szabo *et al.*, 2020), we observe that constraints introduced by cohesin reduce spatial distances between genomic sequences in *cis* and increase the chances that they interact. We now, however, reveal that this entails an increase in both the rate of formation and the duration of contacts. Convergent high-affinity CTCF motifs lead to higher contact frequencies and substantially longer contact durations, somewhat similar to the effect of insulator elements in *Drosophila* (H. Chen *et al.*, 2018). Comparison with polymer simulations reveals that in mESCs this can be understood in terms of stalling of loop-extruding cohesins. This observation also suggests that asymmetries in contact patterns established by CTCF motifs genome-wide might also lead to temporal asymmetries in physical interactions, notably between regulatory sequences. We additionally observe that constraints introduced by cohesin and CTCF sites lead to reduced temporal variability in physical distances, arguing that loop extrusion increases the reproducibility of chromosome folding at selected genomic sites.

Our study also provides estimates of the frequency and duration of chromosomal contacts at genomic-length scales that represent enhancer–promoter communication genome-wide. In our study, contacts are defined by physical distances (~ 150 nm) that might be comparable to those where signals arise in 3C methods (McCord *et al.*, 2020). For sequences separated by 150 kb, such contacts assemble and disassemble over minutes. This provides many opportunities in a single cell-cycle for regulatory sequences in a TAD to contact each other, and suggests that long-range regulation by distal enhancers might rely on transient interactions. We note that despite accurate correction of chromatic aberrations, shorter-range and thus potentially faster proximity events remain inaccessible in our experimental set-up (Brandão, Gabriele, *et al.*, 2021). Estimates based on comparison with polymer simulations further suggest that cohesin-mediated interactions between convergent CTCF sites might last around 5–15 min on average and at least for sequences located 150 kb apart occur around 27% of the time. This is in good agreement with recent estimates of the duration of a 500 kb loop in mESCs (10–30 min on average) (Gabriele *et al.*, 2022), which, however, occurs more rarely (3.5–6% of the time). This is in line with the predictions from polymer simulations that increasing the genomic distance between convergent CTCF sites should substantially decrease the frequency of CTCF-mediated interactions,

but not their duration (Extended Data Fig. E10). Taken together, our data establish firm quantitative bases for understanding the dynamics of chromosome folding within TADs and provide temporal constraints for mechanistic models of chromosome structure and its impact on fundamental biological processes such as long-range transcriptional regulation.

2.5 Methods

Culture of embryonic stem cells

All cell lines are based on E14 mouse embryonic stem cells mESCs. E14 CTCF-AID-eGFP (clone EN52.9.1) were published in Nora *et al.*, 2017. E14 WAPL-AID-eGFP and E14 RAD21-AID-eGFP were published in N. Q. Liu *et al.*, 2021. The latter were kindly provided by Elzo de Wit (Netherlands Cancer Institute). All cell lines for the dual array imaging approach are based on the double-CTCF knockout cell line described in Zuin *et al.*, 2022. Cells were cultured on gelatin-coated culture plates in Glasgow Minimum Essential Medium (Sigma-Aldrich, G5154) supplemented with 15% foetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μ M β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U/ml leukemia inhibitory factor (Miltenyi Biotec, premium grade) in 8% CO₂ at 37 °C. Cells were tested for mycoplasma contamination regularly and no contamination was detected. After genome engineering and for Hi-C, Capture-Hi-C, 4C, Western Blot and imaging experiments, cells were cultured in standard E14 medium supplemented with 2i (1 μ M MEK inhibitor PDO35901 (Axon, 1408) and 3 μ M GSK3 inhibitor CHIR 99021 (Axon, 1386)). For live-cell imaging experiments, cells were cultured in Fluorobrite Dulbecco's Modified Eagle Medium (DMEM) (Gibco, A1896701) supplemented with 15% foetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μ M β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U/ml leukemia inhibitory factor (Miltenyi Biotec, premium grade) and with 2i inhibitors (1 μ M MEK inhibitor PDO35901 (Axon, 1408) and 3 μ M GSK3 inhibitor CHIR 99021 (Axon, 1386)).

Generation of mESC lines carrying random integrations of TetO arrays

To generate clonal cell lines carrying random integrations of the TetO array in the degron cell lines (E14 Rad-AID-eGFP, E14 CTCF-AID-eGFP and E14 WAPL-AID-eGFP), 0.5x10⁶ cells were transfected with 2 μ g PB-3xCTCF-TetO vector, 200 ng PB-TetR-tdTomato and 200 ng pBroad3_hyPBase_IRES_tagRFpt (Redolfi *et al.*, 2019) with Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's recommendations. Cells were cultured in standard E14 medium for 5 days and subsequently sorted by fluorescence activated cell sorting (FACS) for fluorescent emission at 581 nm (tdTomato) on 96 well-plate to isolate clonal lines. Sorted cells were kept for 2 days in standard E14 medium supplemented by 100 μ g/ μ l primorcin (InvivoGen, ant-pm-1) and 10 μ M ROCK inhibitor (STEMCELL Technologies, Y-27632). 10 days after sorting the plates were duplicated by detaching

with accutase (Sigma Aldrich, A6964) and re-seeding in full E14 culture medium. 1/3 of the cells were replated onto Corning High-Content Imaging Glass Bottom Microplates (96-well, Corning, 4580). 2 days after reseeding, clonal lines were screened by microscopy for >10 insertions of TetO/cell and a good SNR. Selected clones were expanded and genotyped by PCR for the absence of random integration of the PiggyBase itself. Primers used for genotyping are listed in Supplementary Table S3.

Generation of dual array (TetO-LacO) mESC line

Integration of the TetO array into the genomic locus on chr15:11,647,372: The vector containing the gRNA sequence was available from a previous study (PX459-chr15-gRNA/Cas9 (Zuin *et al.*, 2022)). The gRNA sequence can be found in Supplementary Table S3. E14 mESC already containing a double-knockout for CTCF sites (clone D6 in Zuin *et al.*, 2022) were transfected with the targeting vector pMK-3xCTCF-TetO-Rox-PuroR-Rox and the gRNA vector PX459-chr15-gRNA/Cas9 using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024 KT). 2×10^6 cells were nucleofected with 1 μ g TetO targeting vector and 1 μ g of PX459-chr15-gRNA/Cas9 as described above and treated with 1 μ g/ml of puromycin (InvivoGen, ant-pr-1) 48h after transfection for 3 days to select cells for insertion of the TetO cassette. Cells were then cultured in standard E14 medium for additional 7 days and subsequently sorted by FACS on 96 well-plate as described above to isolate clonal lines. 10 days after sorting the plates were duplicated by detaching with accutase (Sigma Aldrich, A6964) and re-seeding in full E14 culture medium. Genomic DNA was extracted on plate by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl and 1 mg/ml proteinase K (Macherey-Nagel, 740506) and 0.05 mg/ml RNase A (Thermo Fisher Scientific, EN0531) and subsequent isopropanol precipitation. Individual cell lines were analyzed by genotyping PCR to determine heterozygous insertion of the TetO cassette. Cell lines showing the corrected genotype were selected and expanded. Primers used for genotyping are listed in Supplementary Table S3. Targeted nanopore sequencing with Cas9-guided adapter ligation (Gilpatrick *et al.*, 2020) (as described below) was performed on expanded clones to confirm single-copy insertion of the TetO cassette. Clone 2G5 was used for further engineering.

Integration of the LacO array into the genomic locus on chr15:11,496,908: The gRNA sequence for the CRISPR/Cas9 knock-in of the LacO cassette was designed using the online tool https://eu.idt.dna.com/site/order/designtool/index/CRISPR_SEQUENCE and purchased from Microsynth AG. The gRNA sequence can be found in Supplementary Table S3. The gRNA sequence was cloned into the PX330 plasmid (Addgene, #58778) using the BsaI restriction site. The clonal line carrying the TetO cassette (clone 2G5) was transfected with the targeting vector pUC19-ITR-NeoR-ITR-3xCTCF-LacO and the gRNA vector pX330-chr15_LacO_gRNA/Cas9 using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024 KT) as described for the Tir1 integration. 48 hours after transfection, 250 μ g/ml of G418 (InvivoGen, ant-gn-1) was added to the medium for 3 days to select cells for insertion of the LacO cassette. Cells were sorted and genotyped as described for the TetO integration. Primers used for genotyping are listed in Supplementary Table S3. Cell lines showing the corrected genotype were selected and expanded. Expanded clones were transiently transfected with 200 ng PB-TetR-tdTomato and 200 ng

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

PB-LacI^{**}-eGFP using Lipofectamine3000 according to the manufacturer's instructions (Thermo Fisher Scientific, L3000008) and 2 days after transfection validated for heterozygous insertion of the LacO cassette on the same allele as the TetO by microscopy. Targeted nanopore sequencing with Cas9-guided adapter ligation (as described below) was performed on correct clones to confirm single-copy insertion of the LacO cassette. Clone 1F11 was used for further engineering.

To visualize the operator arrays in live-cell imaging and remove the puromycin resistance gene used for selection during integration, 0.5×10^6 E14 TetO-LacO cells (clone 1F11) were transfected with 200 ng PB-TetR-tdTomato, 200 ng PB-LacI^{**}-eGFP and 200 ng pBroad3_hyPBBase_IRES_tagRFpt (Redolfi *et al.*, 2019) with Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. 7 days after transfection the cells were sorted (as described previously) for fluorescent emission at 507 nm (eGFP) and 581 nm (tdTomato). Sorted cells were cultured and genotyped as described for the random TetO integration. Primers used for genotyping are listed in Supplementary Table S3. Cell lines showing the corrected genotyping pattern were selected and expanded and a good and comparable SNR was selected for by microscopy. Clones 1B4 (+PuroR) and 2C10 (-PuroR) were used for further engineering.

Live-cell imaging

35 mm glass-bottom dishes (Mattek, P35G-1.5-14-C) were coated with 1-2 $\mu\text{g}/\text{ml}$ Laminin (Sigma-Aldrich, L2020) in PBS at 37°C overnight. Cells (1×10^6) were seeded in Fluorobrite medium (as described above) 24 h before imaging. For targeted degradation of RAD21, WAPL or CTCF in the degron cell lines, the medium was exchanged to medium containing 500 μM auxin (Sigma-Aldrich, I5148-2G) at the respective time required for complete degradation of the protein target prior to imaging (RAD21: 90 min, WAPL: 24 h, CTCF: 6 h). For targeted depletion of RAD21 using the FKBP degron system (dual array cell lines), cells were cultured in Fluorobrite medium containing 500 nM dTAG-13 (Sigma-Aldrich, SML2601-1MG) 2 h prior to imaging.

For fixed cell measurements to estimate the localization error, 1×10^6 cells were seeded onto Mattek dishes and incubated for 24 h at 37 °C, 8%CO₂. The medium was removed and the cells were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15710) in PBS for 10 min at RT. The cells were washed three times in PBS and Fluorobrite medium was added to the Mattek dish to achieve comparable background fluorescence levels.

Cells were imaged with a Nikon Eclipse Ti-E inverted widefield microscope equipped with a Total Internal Reflection Microscopy iLAS2 module (Roper Scientific), a Perfect Focus System (Nikon) and motorized Z-Piezo stage (ASI) using a CFI APO TIRF 100 x 1.49 NA oil immersion objective (Nikon). The microscope was operating in highly inclined and laminated optical sheet (HILO) mode (Tokunaga *et al.*, 2008). Excitation sources were a 488 nm, 200 mW Toptica iBEAM SMART laser and a 561 nm 200 mW Coherent Sapphire laser. Images were collected on two precisely aligned back-illuminated Evolve 512 Delta EMCCD cameras with a pixel size of 16 μm x 16 μm (Photometrics). Cells were maintained at 37 °C and 8% CO₂ using an enclosed microscope environmental control setup (The BOX and The CUBE, Life Science Instruments). Before the acquisition of movies for the dual-array

set-up, TetraSpeck™ Microspheres, 0.1 μm beads (Thermo Fisher Scientific, T7279) were imaged to allow for correction of chromatic aberrations during image processing and analysis. Movies for measurement of random TetO integrations in degron cell lines were acquired every 10 s (exposure time: 50 ms) in 34 z-planes (10 μm stack, $\text{dz} = 300 \text{ nm}$) with the Visiview software (Visiview 4.4.0.12, Visitron). Images for measurement of cell lines with the dual array set-up were acquired every 30 s, with an exposure time of 50 ms, respectively, each in a sequential mode with 21 z-planes (6 μm stack, $\text{dz} = 300 \text{ nm}$). For the measurement of the time it takes the operator arrays to displace by its own size, images were acquired continuously on a single focal plane over 10 s every 0.1 s with exposure times of 50 ms.

Image processing

Raw images were deconvolved using the Huygens Remote Manager and a classical maximum likelihood estimation algorithm with a theoretical point-spread-function. The initial signal-to-noise ratios were estimated from the images and images were deconvolved until one of the following stopping criteria was reached: The maximum number of iterations was performed (for random integrations: 20 cycles, for tdTomato and eGFP; in dual-color set-up: 15 cycles for tdTomato signal, for GFP signal: 5 cycles) or a quality change criterion below 0.001 was returned. Representative image series shown in the main figures were deconvolved as described above, adjusted to display the same brightness and contrast and interpolated using a bicubic interpolation. Movies were corrected for bleaching over time using an exponential fit. The 2D projection of intensity changes over time was created using the Temporal Color Code in Fiji v. 2.0. (https://github.com/fiji/fiji/blob/master/plugins/Scripts/Image/Hyperstacks/Temporal-Color_Code.ijm).

Spot detection and localization of multi operator data

Our field of view typically contains approx. 25 mESC nuclei. Despite the fact that our mESC lines are clonal, background nuclear fluorescence intensities in each cell can vary substantially. This poses challenges to conventional threshold-dependent algorithms for spot detection and localisation which perform unevenly across cells with different background intensities. To overcome these limitations, we implemented a two steps procedure for 3D spot detection and localisation. To detect spots, we used deepBlink v. 0.1.1 (Eichenberger *et al.*, 2021), a convolutional neural network-based spot detection and localization algorithm in 2D, which has been shown to be able to deal with different background intensities and to detect spots in a threshold independent manner. To enhance our detection efficiency, we employed custom models trained on a combination of the following datasets: smFISH and SunTag datasets provided by deepBlink and in-house manually curated live cell imaging images. To detect 3D spots, we applied deepBlink to all z-stacks separately followed by linkage of the spots across z-stacks using Trackpy (Allan *et al.*, 2021). The precise 3D coordinates of the spots are then determined using 3D gaussian fitting using a voxel of size 6x6x4 pixels centered at the spot in the brightest z stack. deepBlink models can be found at https://github.com/zhanyinx/SPT_analysis/tree/main/models. The parameters and models used for each cell line can be found in Supplementary Table S4. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

Tracking and cell motion correction of multi operator data

3D spots coordinates are fed into TrackMate for tracking using linear assignment problem (LAP) tracker. Each track is assigned to manually annotated cell masks (from max z-projection of frame 93) using a custom script (https://github.com/zhanyinx/SPT_analysis/blob/main/source/spot_detection_tracking/assign_cellids.py), which uses the majority rule. Motion correction is then performed using a roto-translation model. Specifically, for each pair of consecutive time frames, a set of matching spots in every cell is determined by solving the linear assignment problem using the euclidean distance between spots as a measure of distance. Only spots that match across two consecutive frames are then used to estimate the roto-translation model that is then applied to correct for nuclear motion (6 matching spots on average across all time frames, trajectories and movies, with a minimum of 4 spots per pair of time frames). All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

MSD analysis of multi operator data

Tracks with less than 10 spots are filtered out for followed up analysis. To calculate the MSD, we first calculate the time-averaged MSD for each trajectory. We then calculate the ensemble average (across trajectories) MSD by pooling all replicates. The ensemble average is done in log space. We corrected the localization error effect on the MSD curve by estimating the standard deviation of the error distribution using fixed images as described in Kepten *et al.*, 2013. To calculate the scaling (α) and the generalized diffusion coefficient (D) of each MSD curve, we fitted the ensemble average of the log-time average MSD between 10-100s. To test the significance of differences between conditions, we fitted α and diffusion coefficient for each cell. The *P*-value is calculated using Student t-test (two-sided). Since we are always comparing two conditions whose cell cycle profiles are similar, we ignore the effect of sister chromatids. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/. The specific Fiji and relative plug-ins can be found at https://github.com/giorgetti/Mach_et_al_chromosome_dynamics/tree/master/Fiji.

Chromatic aberration correction of dual color data

To correct for chromatic aberration we took 3D image stacks of TetraSpeck™ Microspheres, 0.1 μm beads (Thermo Fisher Scientific, T7279) adsorbed on MatTek dishes in 1xPBS at the beginning of every imaging session and used them to correct the corresponding set of movies. After detecting signals from single beads in each channel using deepBlink and determining their 3D location by Gaussian fitting, we first identified spots that are shared across channels by solving the linear assignment problem (LAP) using the euclidean distance between spots. We then used the common set of bead signals to compute a 3D roto-translation that we finally applied to xyz positions. This procedure corrects for x, y and z aberrations simultaneously. The same transformations accurately corrected chromatic aberrations in actual experiments in double-labeled mESC (see “Control TetO” in Extended Data Fig. E6G), with the exception of a small residual systematic shift (approx. 40 nm) along the z axis (see TetO-LacO case in Extended Data Fig. E6G), which is likely due to 3D image anisotropies that cannot be measured using “2D” bead images.

Tracking and MSD analysis of dual color data

To increase the ability to detect longer tracks, we used an in-house script to stitch multiple tracks belonging to the same cell (https://github.com/zhanyinx/SPT_analysis/blob/main/source/dual_channel_analysis/utils.py, stitch function). In short, if two tracks from the same cell overlap more than 50% in time, the shortest one is filtered out. We called cell masks using CellPose (Stringer *et al.*, 2021) on the max z-projection of the middle frame of the movie using the GFP channel and used these masks to define cell identity. For tracks with overlaps lower than 50%, the overlapping part of the tracks are randomly removed from one of the two tracks. The resulting tracks are stitched if the distance across the time gap is smaller than 1.6 μm . To match tracks across channels, we used the following measure to calculate the distance between tracks across channels:

$$\frac{\langle \sum_{i=1}^3 (x_{1i}(t) - x_{2i}(t))^2 \rangle_{t \in T_1 \cap T_2}}{\sqrt{\text{len}(t \in T_1 \cap T_2)}} \quad (2.1)$$

Where x_1 are the coordinates from channel 1 and x_2 are the coordinates from channel 2, T_1 contains all the time frames from channel 1 and T_2 contains all the time frames from channel 2, and len is a function that returns the length of an array. We solved the linear assignment problem using the distance measure above to match tracks across channels. Tracks with average distances across channels higher than 1 μm are filtered out. Matched tracks with lower than 25 time points are filtered out. For each matched pair of tracks, we calculate the pairwise distance using the euclidean distance in 3D. We define noisy pairwise distance using the ratio of the pairwise distance in 3D and 2D. In particular, we defined noisy the top 5% of this ratio and filtered them out. To calculate the radial MSD, we first calculate the time-averaged radial MSD for each pairwise distance “trajectory”. We then calculate the ensemble average (across trajectories) of the log of time-averaged radial MSD. We corrected for the radial localisation uncertainty by estimating the standard deviation of the error distribution using fixed images as described in Kepten *et al.*, 2013. To calculate the scaling (α) and the generalized diffusion coefficient (D) of each MSD curve, we fitted the ensemble average time average MSD between 30-300 s. Since we are always comparing two conditions whose cell cycle profiles are similar, we ignore the effect of sister chromatids. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

Estimation of experimental uncertainty on radial distance

To estimate our uncertainty in detecting distances across channels, we used a cell line with multiple integration of TetO arrays that can be tagged with TetR-eGFP and TetR-tdTomato. Spot detection is done as for our dual color lines. We corrected for chromatic aberration using TetraSpeck™ Microspheres, 0.1 μm beads (Thermo Fisher Scientific, T7279) and then matched spots across channels by solving the linear assignment problem using `scipy.optimize.linear_sum_assignment` function with the euclidean distance between spots as a measure of distance. Spots across channels with distances higher than a threshold are filtered out to avoid mismatches. We used a threshold of 300 nm for matching the spots registration. We applied a second round of chromatic aberration correction using

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

the set of registered points themselves. The resolution limit (uncertainty) is then estimated as the average distance between registered spots which corresponds to $130 \text{ nm} \pm 70 \text{ nm}$.

HMM for detection of the proximal state

To detect the proximal state in a threshold independent manner, we used a HMM with two hidden states (proximal and distal). We used a Gaussian model for the emission probabilities. Only distance trajectories with less than 20% missing values at any time point are kept. Missing values are filled with the first preceding time point with distance value. In order to more reliably detect the proximal state, we used all the trajectories from the experimental condition with both cohesin and CTCF to train an HMM. We then re-train an HMM model per experimental condition by using the proximal state (Gaussian mean and standard deviation) from the experimental condition with both cohesin and CTCF. Finally, we applied the experimental condition specific-HMM to every trajectory to estimate the contact duration and rate of contact formation for all the experimental conditions. The HMM model training can be found as a jupyter notebook (https://github.com/zhanyinx/SPT_analysis/blob/main/notebooks/HMM_experimental_data.ipynb). We modified the `hmmlearn` library to allow fixing proximal state during HMM training. The modified `hmmlearn` library can be found at <https://github.com/zhanyinx/hmmlearn>.

Simulations

Polymer simulations were performed using LAMMPS (A. P. Thompson *et al.*, 2022). We chose Langevin dynamics with the NVT thermostat. Arbitrary units were set such that thermal energy $k_B T = 1$, where k_B – the Boltzmann constant and T – temperature, corresponding to 300 K. For every set of parameters, we performed ten independent runs. A run consists in an equilibration part of 107 simulation steps and a production part of 108 simulation steps. For subsequent analysis and calculation of contact maps we recorded the data every 104 simulation steps. In simulations for Fig. 2.2, the chain length was 1125 beads. In simulations for Fig. 2.5 and 6, the chain length was 1000 beads. We used PyMOL software (v. 2.3.3) to represent snapshots of polymer chain in Fig. 2.2a. Examples of initial conformations and simulation parameters can be find at https://github.com/giorgettilab/Mach_et_al_chromosome_dynamics, in the polymer simulations section.

To simulate the loop extrusion process, we developed and embedded in LAMMPS a new package called “USER-LE”. Loop extrusion model contains extruders and barriers on the polymer. An extruder is represented as an additional sliding bond, which extrudes the loop in a two-sided manner. It can be loaded to the polymer between (i) and (i+2) beads with a certain probability only when the bead (i+1) is unoccupied by another extruder and is not a barrier. Each extruder can be unloaded from polymer with a certain probability. Every bead could be occupied by only one extruder. Extruders cannot pass through each other. When extruders meet each other on the polymer, they stall until one of them is released. Every extruder attempts to make an extruding step every N simulation steps.

In addition to “neutral” polymer beads, there are 3 types of barriers blocking loops coming from the left, from the right and from any direction. These barriers mimic CTCF sites, for which one can define a probability for the loop extruder to go through (the same probability for all barriers). To launch loop

extrusion, one should define three fixes with LAMMPS syntax: loading, unloading and loop extrusion. Loading: frequency in number of steps to try to load extruders, types of beads, max distance to create, type of the bond (extruder) to be created, probability to create, seed for pseudorandom generator of numbers, new type of the first beads and new type for the second bead. Unloading: frequency in number of steps to try to unload extruders, type of the bond (extruder), min distance to release bond, probability to release bond, seed for pseudorandom number generator. Loop extrusion: frequency in number of steps to try to move extruders, neutral polymer type, left barrier type, right barrier type, probability to go through the barrier, type of the bond (extruder), and type of two-sided barrier (optional).

Statistics and Reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. No randomization was performed as the study did not require sample allocation into different groups. Live-cell imaging experiments were performed in 3–7 biological replicates and all replicates showed consistent results. For Capture-C, Hi-C, 4C, piggyBac insertion site mapping and Nanopore sequencing with Cas9-guided adapter ligation, one biological replicate was performed. For flow cytometry measurements two biological replicates were performed. Western blot analysis and genotyping PCR with subsequent agarose gel electrophoresis were performed with 1–2 biological and 2 technical replicates. Blinding was not possible for data collection in live-cell imaging experiments, as data acquisition required identification of the sample for further processing. Data analysis for live-cell imaging, Capture-C, Hi-C, 4C and piggyBac insertion site mapping were performed in a blinded manner. Blinding was not necessary for the other experiments since the results are quantitative and did not require subjective judgment or interpretation. Whenever Student's t-test was used, we formally verified the normality of distributions but assumed variance equality.

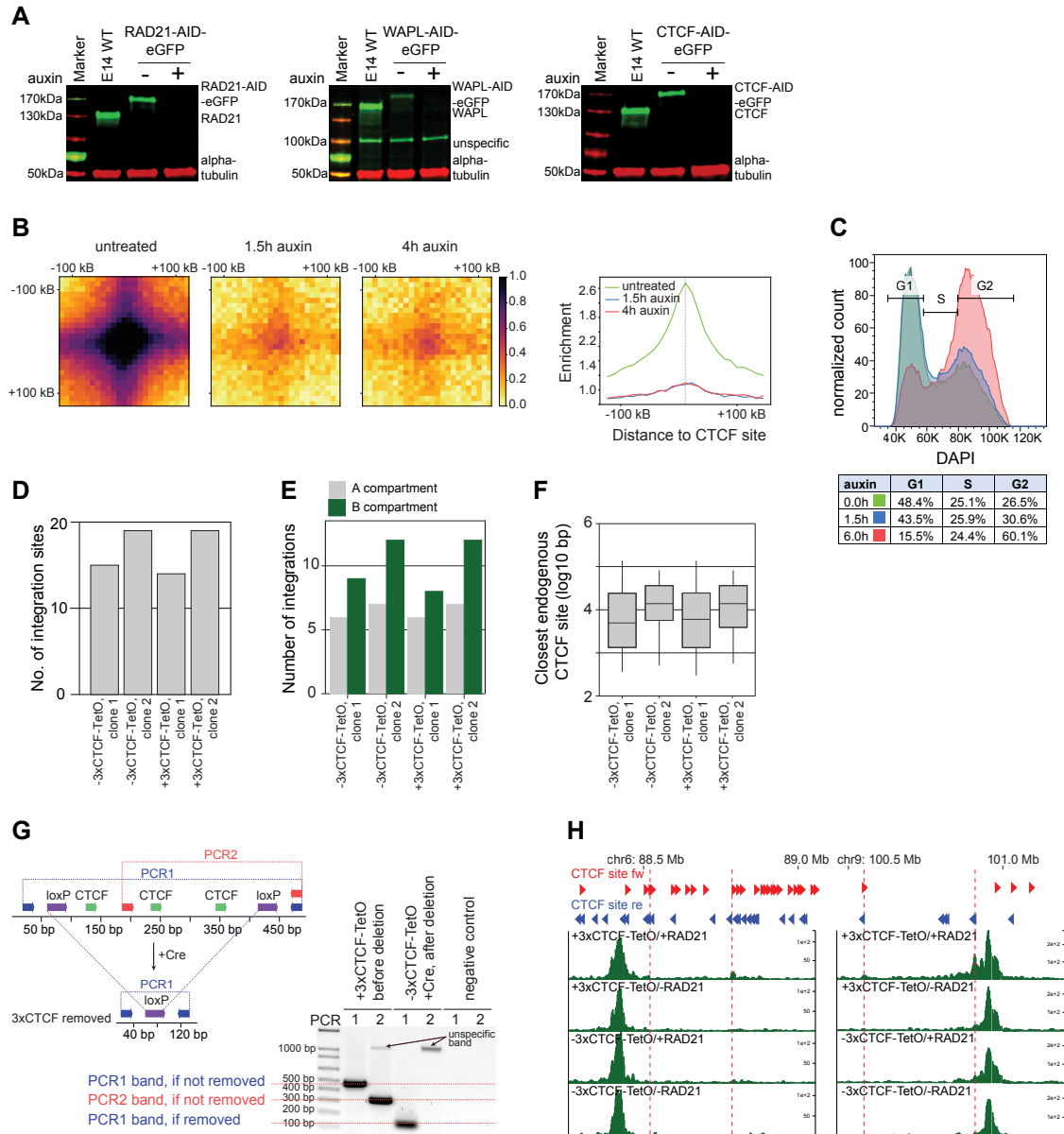
Data availability

All Capture-C, Hi-C, 4C and integration site mapping sequencing fastq files generated in this study have been uploaded to the Gene Expression Omnibus (GEO) under accession GSE197238. The following public database was used: BSgenome.Mmusculus.UCSC.mm9 (<https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Mmusculus.UCSC.mm9.html>). The trajectories from imaging data can be found at <https://doi.org/10.5281/zenodo.6627715>. Source data are provided with this paper.

Code availability

Custom codes generated in this study are available at: https://github.com/zhanyinx/SPT_analysis/ (image analysis); https://github.com/giorgettilab/Mach_et_al_chromosome_dynamics/ (4C, Hi-C, nanopore, simulation analysis); https://github.com/polly-code/lammps_le (repository with loop extrusion module for the LAMMPS); and <https://github.com/zhanyinx/hmmlearn> (the modified version of hmmlearn).

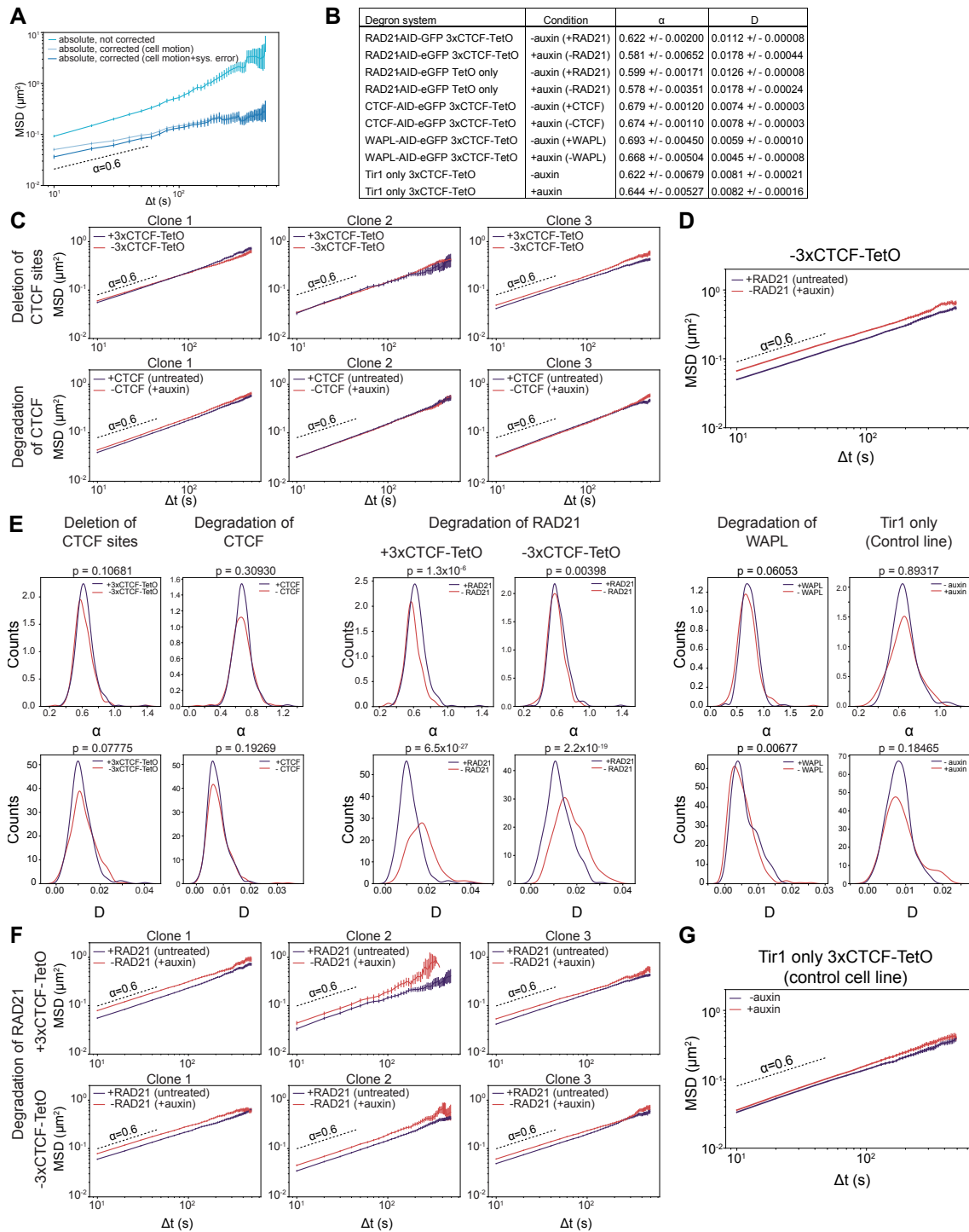
2.6 Extended Data



Extended Data Figure E1: Chromosome structure is altered upon degradation of factors involved in loop extrusion. **A**, Western Blots showing degradation of RAD21, WAPL and CTCF upon 1.5 h, 24 h and 6 h, respectively. Loading control: α -tubulin, $n = 1-2$ replicates for each cell line. **B**, Left: Average enrichment in Hi-C read counts at CTCF sites based on Hi-C data in RAD21-AID-eGFP cells either untreated (left), treated for 1.5 h (middle) or 4 h (right) with auxin. Right: Differences in enrichment at CTCF peaks. Peaks were called on Hi-C data from untreated cells. **C**, Flow cytometry analysis of fixed cells stained with DAPI showing cell-cycle stage distributions of RAD21-AID-eGFP mESC cultured with serum, LIF and 2i, either before (green) or after 1.5 h (blue) and 6 h (red) auxin treatment. **D**, Integration site numbers in two clones of RAD21-AID-eGFP lines with and without 3xCTCF sites. **E**, Distribution of integration sites from lines shown in panel D that belong to A and B compartments called on distance-normalized Hi-C map (same as panel B).

Extended Data Figure E1: *continued from previous page:* **F**, Integration sites distances from the closest endogenous CTCF site. Boxplot: lower and upper quartiles (Q1 and Q3, respectively); whiskers: 1.5x interquartile region (IQR) below Q1 and above Q3. n = 15 and 19 insertions for -3xCTCF-TetO clones 1 and 2, respectively, n = 14 and 19 insertions for +3xCTCF-TetO clones 1 and 2, respectively. **G**, Example of genotyping PCR upon removal of 3xCTCF sites in a RAD21-AID-eGFP +3xCTCF-TetO clonal line. PCR1 amplifies the entire 3xCTCF cassette and product size changes from 470 bp to 147 bp if the cassettes are successfully removed. PCR2 amplifies half of the 3xCTCF cassette and no product is expected if 3xCTCF cassettes were removed from all insertion sites; otherwise a PCR band of 303 bp is expected. **H**, Representative 4C profiles from insertions on chromosomes 6 and 9 using TetO as a viewpoint showing that 3xCTCF-TetOs lead to the formation of ectopic contacts (dashed red lines) with nearby endogenous CTCF sites in the presence of RAD21. Contacts are lost upon deletion of 3xCTCF cassette (-3xCTCF-TetO) and upon degradation of RAD21 (-RAD21).

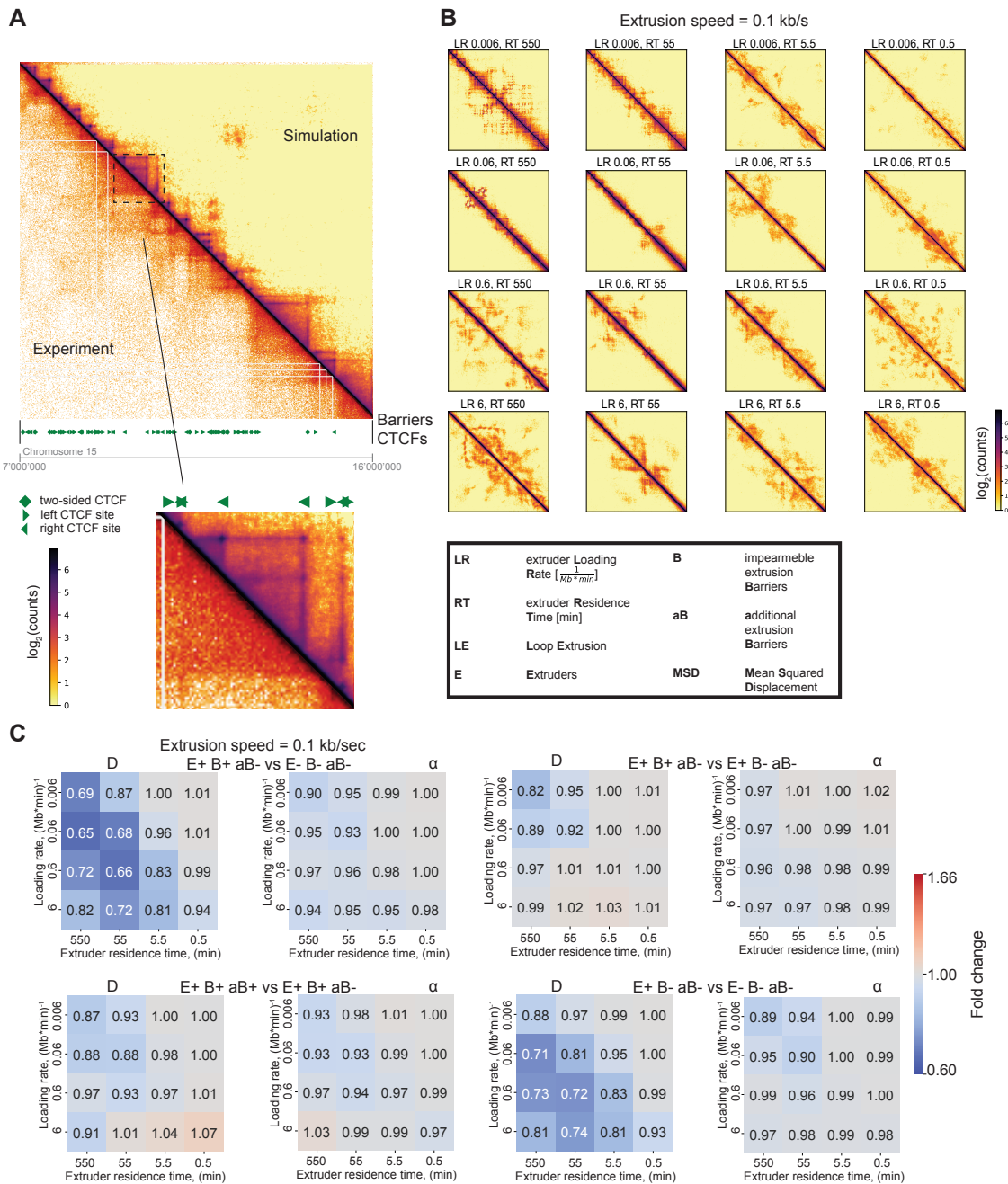
2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING



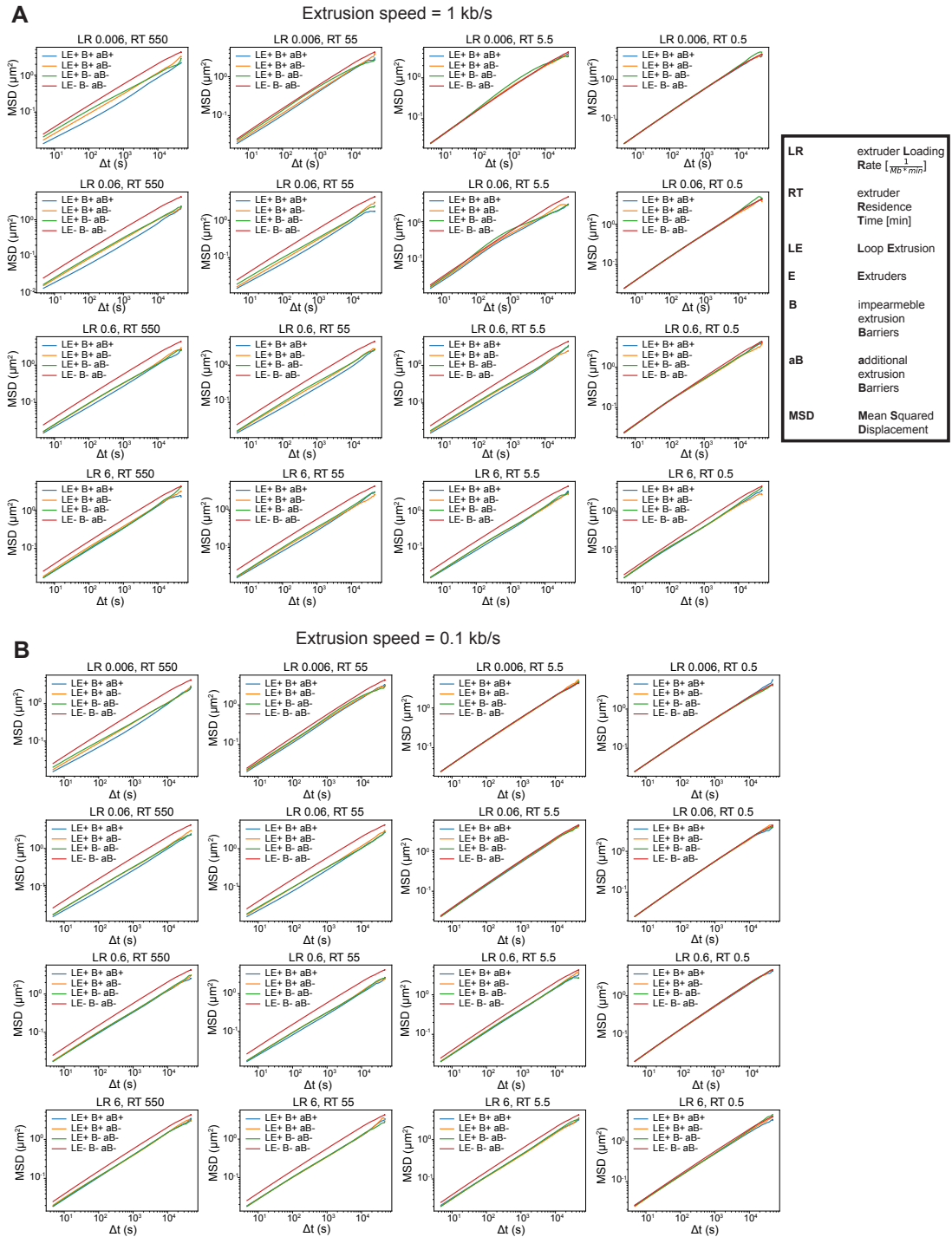
Extended Data Figure E2: Chromosome dynamics is modulated by degradation of factors involved in loop extrusion. **A**, MSD of trajectories from TetO insertions within the same cell (MSD, mean \pm s.e.m., $n = 45$ tracks) before (cyan) and after applying cell motion (light blue, $n = 45$ tracks) and localisation error correction (dark blue, $n = 45$ tracks). **B**, Scaling exponents (α) and generalized diffusion coefficients (D) across all conditions and cell lines were fitted by pooling all three biological replicates. Shown are the numbers for the best fit \pm error of the fit. **C**, MSD (mean \pm s.e.m.) plots for a single clonal cell line (biological replicate) when looking at removal of 3xCTCF sites (top row) next to the array or degrading all CTCF (bottom row).

Extended Data Figure E2: *continued from previous page:* **D**, MSD (mean \pm s.e.m.) in the cell lines ($n = 3$ replicates per clonal cell line, three cell lines) where the 3xCTCF cassette was excised. Shown are the MSDs for cells either depleted of RAD21 for 90 min (red, 266 cells, 9,020 trajectories analyzed) or not (blue, 271 cells, 11,082 trajectories analyzed). Global depletion of RAD21 increases mobility. *P*-values in panel **E**, Distributions of α and D fitted based on single trajectory MSD and significance test for differences in generalized diffusion coefficients (D) and scaling exponents (α). The *p*-value is calculated using Student *t*-test (two-sided) (see Methods). **F**, Same as in **C** for a single clonal cell line (biological replicate) with integrations with 3xCTCF-TetO (top row) or without 3xCTCF-TetO (bottom row) when degrading RAD21. Global depletion of RAD21 increases mobility. **G**, Same as in **D** in the cell lines that contain integrations of 3xCTCF-TetO and the Tir1 protein, but do not contain any AID-tag for targeted degradation. MSDs for cells either treated with auxin for 90 min (red, 97 cells, 2,155 trajectories analyzed) or not (blue, 111 cells, 3,711 trajectories analyzed). No significant changes were detected. *P*-values in panel **E**.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

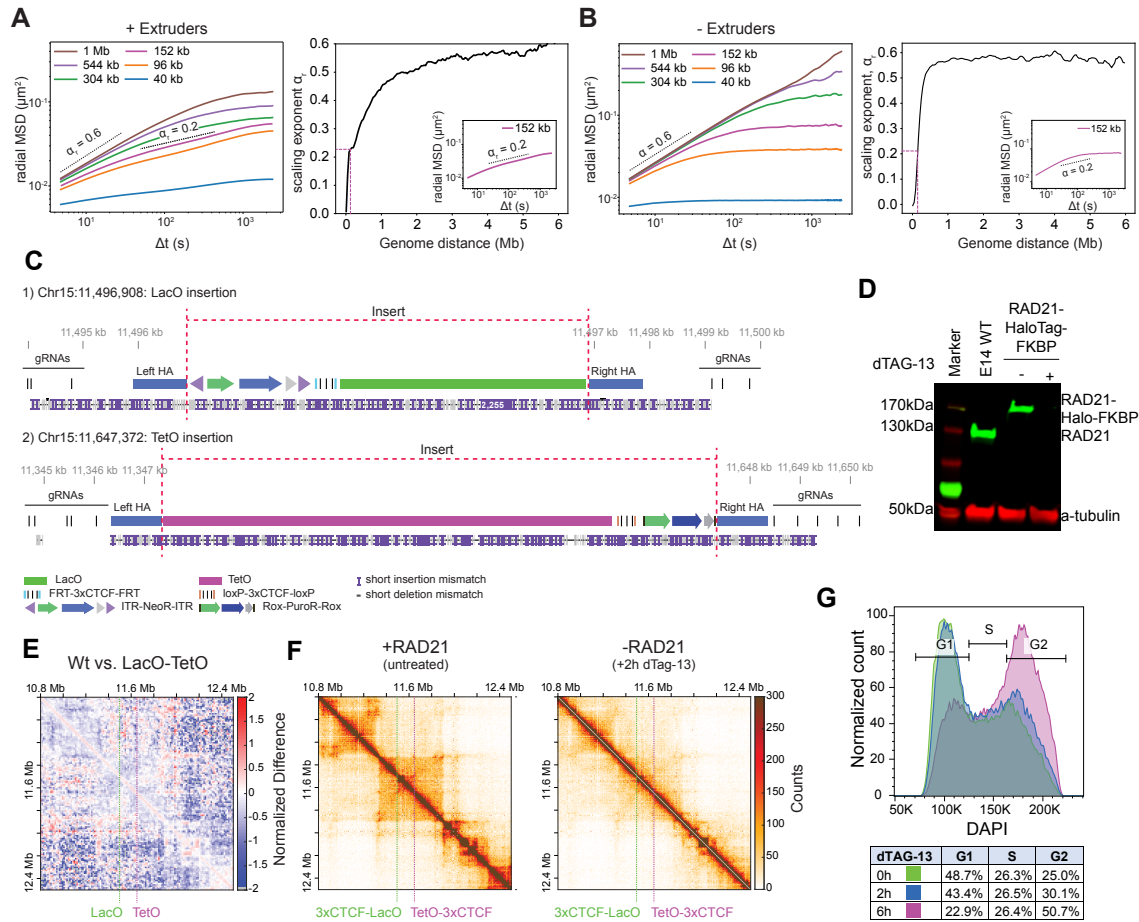


Extended Data Figure E3: Simulations of chromosome dynamics and effects of loop extrusion. A, Visual comparison of experimental Hi-C contact map with contact maps of simulations at extrusion speed 1 kb/s, extruder loading rate $0.06 (\text{Mb} \times \text{min})^{-1}$ and residence time 5.5 min. **B,** Contact maps for the polymer simulations at extrusion speed 0.1 kb/s and barriers from the range 7–16 Mb of chromosome 15. Acronyms used in this figure are indicated in the black box on the right. **C,** Pairwise comparison for conditions indicated in the title of each pair of heatmaps. Pair of heatmaps contains ratios of generalized diffusion coefficients (D) and scaling exponent (α), and represents fold change between the conditions.

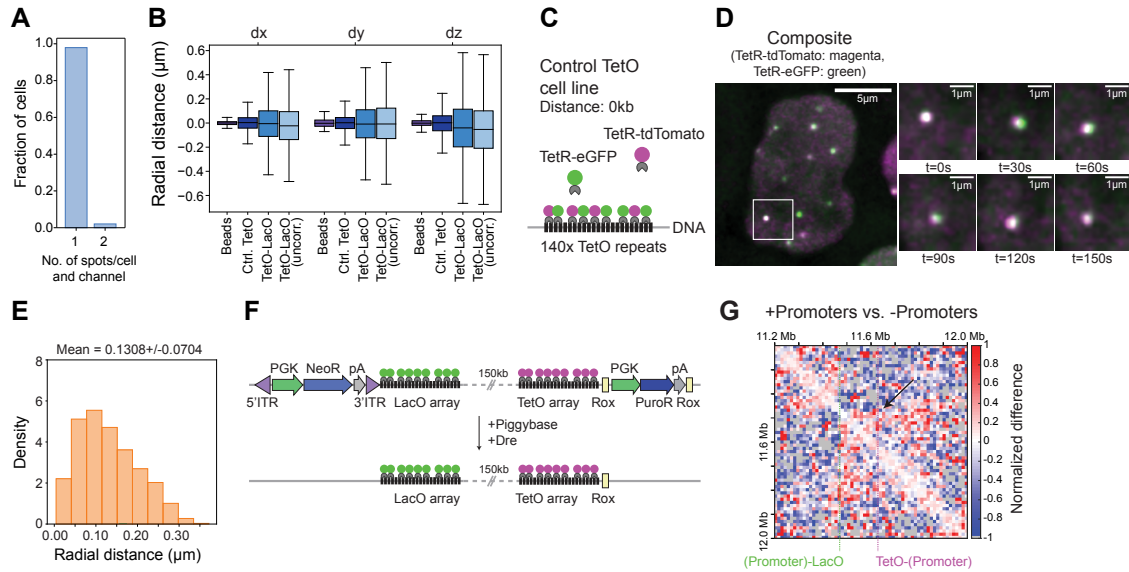


Extended Data Figure E4: MSDs of systems for two extruder speeds. **A**, MSDs for all 16 conditions for each set of loop extrusion parameters and extrusion speed of 1 kb/s. **B**, Same as A but for the extrusion speed of 0.1 kb/s.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

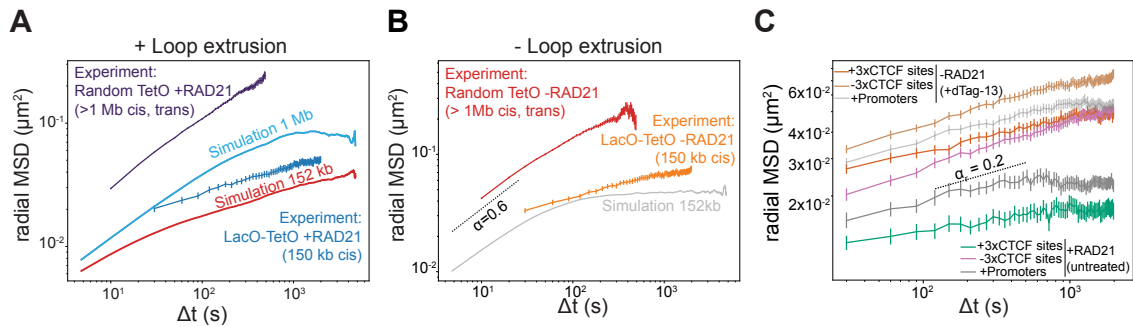


Extended Data Figure E5: Characterization of TetO and LacO array integrations. **A**, Left panel: radial MSD of distances between multiple pairs of monomers separated by distances equivalent to 40 kb - 1 Mb for a polymer with loop extrusion but no barriers. Dashed scaling exponents $\alpha = 0.2$ and $\alpha = 0.6$ serve as an eye guide. Right panel: Slopes of radial MSD curves for two loci separated by varying linear distances, estimated from linear fitting between 5 and 60 sec. Inset: detail of radial MSD and fit for monomers separated by 152 kb. **B**, Left panel: radial MSD of multiple pairs of monomers separated by various distances (40 kb-1 Mb). Simulations were performed for the polymer without extruders and barriers. Values were averaged with a sliding window without considering the first and last 200 monomers (1.6 Mb). Dashed scaling exponent $\alpha = 0.6$ serves as an eye guide. Right panel: Distance dependency of the scaling exponent (α) on the genomic distance between loci. **C**, Integrated Genomic Viewer (IGV) snapshot showing an example of a Nanopore sequencing read mapped to a modified mouse genome including the respective insertions. Reads that spanned from a guide RNA (gRNA) binding site upstream of the left homology arm (left HA) to a gRNA binding site downstream the right homology arm (right HA) confirmed single insertion of the transgene. **D**, Western Blots showing the targeted degradation of RAD21 after 2 h of treatment with 500 nM dTAG-13. Loading control: anti-tubulin, n = 2 replicates. **E**, Differential map at 6.4 kb resolution for the structural differences between a E14 WT and the E14 cell line containing LacO and TetO insertions (see Methods). Dashed lines indicate the insertion sites. No structural changes are detected upon integration of the operator arrays. **F**, Capture-C maps at 6.4 kb resolution in the region on chr15 (10.8 Mb-12.5 Mb) in the untreated cells (left) and in cells treated with 500 nM dTag-13 (left) showing that RAD21 degradation leads to loss of chromosome structure. **G**, Flow cytometry analysis of fixed cells stained with DAPI to show cell cycle stage distribution of E14 RAD21-HaloTag-FKBP cells.

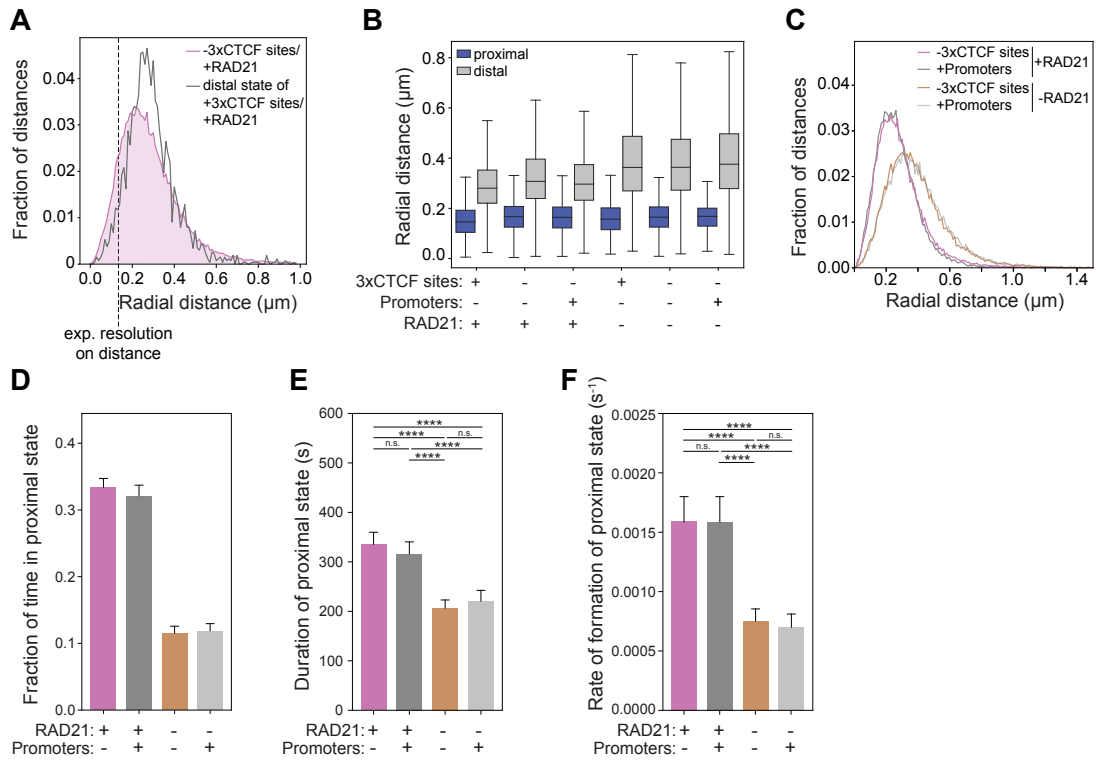


Extended Data Figure E6: Correction of chromatic aberrations and characterization of mESC lines with promoters flanking TetO and LacO arrays. **A**, Bar plot showing the number of detected spots per cell per channel for 1,400 manually annotated images subsampled from the images series. In 3% of the images 2 spots per cell are detected indicating the presence of sister-chromatids. **B**, Distribution of pairwise distances in each dimension for co-localized signals measured on beads ($n = 2,226$ timepoints) or on the control TetO cell line ($n = 69,453$ timepoints), as well as for chromatic-aberration corrected and uncorrected images from TetO-LacO cell lines (in the presence of cohesin and $3 \times$ CTCF sites, $n = 848,955$ timepoints). Boxplot: boxes denote lower and upper quartiles (Q1 and Q3, respectively); whiskers denote $1.5 \times$ the interquartile region (IQR) below Q1 and above Q3. **C**, Schematic representation of the "Control TetO" cell line that contains multiple TetO array integrations as well as stable integrations of TetR-eGFP and TetR-tdTomato. This allows labeling of each TetO array with two separate fluorophores. **D**, Representative images of the "Control TetO" cell line. The time series shows a zoomed version of the region indicated by the white square. **E**, Radial distance distribution of the "Control TetO" cell line as defined in panel C and D showing that the resolution on the 3D distance is ~ 130 nm. **F**, Schematic representation of cell line containing 3-phosphoglycerate kinase (PGK) promoters driving the expression of resistance gene directly adjacent to the operator arrays. The expression cassettes can be excised using Dre recombination or piggyBac transposition to yield the cell line with operator arrays only (PGK=PGK promoter, NeoR=Neomycin resistance gene, PuroR=Puromycin resistance gene, pA=polyadenylation signal, ITR=inverted terminal repeats for piggyBac recognition, Rox=Rox sites for Dre recombination). **G**, Differential map at 6.4 kb resolution for the structural differences between the E14 cell line containing LacO and TetO insertions with the adjacent promoters vs. the E14 cell line containing the operator arrays only (see Methods). Dashed lines indicate the insertion sites.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

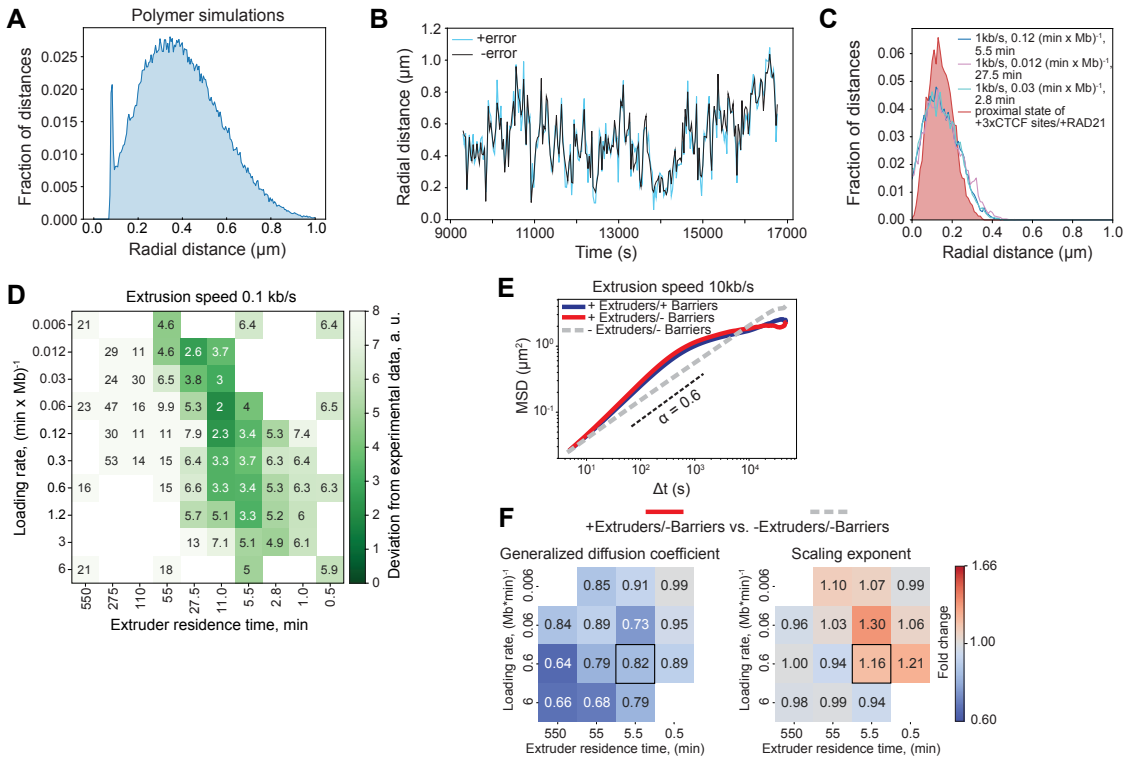


Extended Data Figure E7: Polymer simulations of two genomic locations within the same TAD. **A**, Radial MSD of TetO random integrations (mean \pm s.e.m., purple, see Fig. 2.1e, $n = 271$ cells examined over 3 pooled biological replicates) and of targeted LacO and TetO insertions on Chr15 (mean \pm s.e.m., dark blue, $n = 214$ cells examined over 4 replicates) are compared to model predictions for pairs of loci containing extrusion barriers at a distance of 1 Mb (light blue) and 152 kb (red). Note that random TetO insertions often occur on different chromosomes and thus have larger absolute radial MSD than 1 Mb simulations (but similar scaling). **B**, Radial MSD for cell lines containing multiple random integrations of TetO as shown in Extended Data Fig. E2D (mean \pm s.e.m., red, 266 cells examined over 3 pooled replicates) or the targeted integrations of LacO and TetO on chr15 (mean \pm s.e.m., orange, $n = 277$ cells examined over 6 replicates) in the absence of RAD21 compared to the predicted radial MSD of two loci at a distance of 150 kb in the absence of extruders (gray) as predicted from polymer simulations. **C**, Radial MSD of TetO-LacO distances in mESC lines with or without convergent 3xCTCF sites (or promoters, respectively), either before or after treatment with 500 nM dTag-13 for 2 hours to induce degradation of RAD21 ($dt = 30$,s). radial MSDs are plotted as mean \pm s.e.m. over conditions: +CTCF sites/+RAD21: $n = 152$ cells examined over 4 replicates, -CTCF sites/+RAD21: $n = 214$ cells examined over 4 replicates, +CTCF sites/-RAD21: $n = 248$ cells examined over 7 replicates, -CTCF sites/-RAD21: $n = 277$ cells examined over 6 replicates, +Promoters/+RAD21: $n = 155$ cells examined over 3 replicates, +Promoters/-RAD21: $n = 170$ cells examined over 3 replicates.

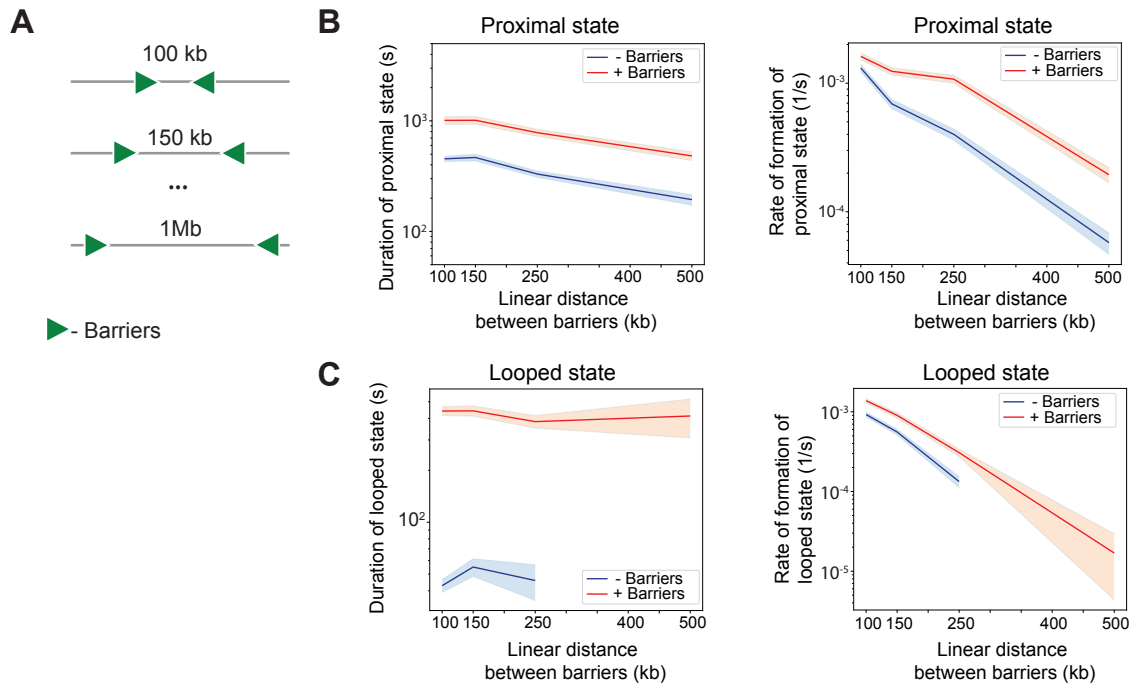


Extended Data Figure E8: Live-cell imaging of two genomic locations within the same TAD. **A**, Radial distance distribution for the condition -3xCTCF sites/+RAD21 (magenta) overlaid with the distal state called by HMM on the +3xCTCF sites/+RAD21 (gray) showing that the distal state identified by HMM largely overlaps with the distance distribution of the two loci in the absence of the CTCF sites. **B**, Boxplot for the radial distances for the proximal and distal state called by HMM on all six conditions. The horizontal line indicates the median. Box plots are as in Extended Data Fig. E1F. Boxplot: boxes denote lower and upper quartiles (Q1 and Q3, respectively); whiskers denote 1.5x the interquartile region (IQR) below Q1 and above Q3. **C**, Distribution of TetO-LacO radial distances in the four experimental conditions. -CTCF sites/+RAD21: $n = 214$ cells examined over 4 replicates, -CTCF sites/-RAD21: $n = 277$ cells examined over 6 replicates, +Promoters/+RAD21: $n = 155$ cells examined over 3 replicates, +Promoters/-RAD21: $n = 170$ cells examined over 3 replicates). **D**, Fraction of time spent in the proximal state called by HMM in the four experimental conditions comparing +Promoters vs. -Promoters +/-RAD21 (no. of cells is as indicated in panel C). Shown average across experimental conditions and error bars represent bootstrapped ($n = 10,000$) standard deviations. **E**, Average duration of proximal states (mean \pm 95% confidence interval, $n = 680$ cells (-promoter +RAD21); $n = 466$ cells (+promoter +RAD21); $n = 268$ cells (-promoter -RAD21); $n = 253$ cells (+promoter -RAD21)) for the conditions +Promoters vs. -Promoters, +/-RAD21. p -values (two-sided Kolmogorov-Smirnov): * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. p -values can be found in Supplementary Table S2. **F**, Average rates of contact formation - time elapsed between the end of a proximal state and the beginning of the next (mean \pm 95% confidence interval, $n = 726$ (-promoter +RAD21); $n = 495$ (+promoter +RAD21); $n = 323$ (-promoter -RAD21); $n = 296$ (+promoter -RAD21)) for the conditions +Promoters vs. -Promoters, +/-RAD21. P -values legend is as in panel E.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING



Extended Data Figure E9: HMM analysis of simulations compared to experimental data. **A**, Bimodal distribution of pairwise distances from simulations corresponding to the set of parameters with a loading rate of $0.06 (\text{Mb} \times \text{min})^{-1}$, extruder residence time of 5.5 min, extruder speed of 1 kb/s, and in the absence of barriers. Data were sampled every 1 s and merged from 10 simulation runs. **B**, Representative radial distance trajectory of a simulated system with and without an additional error on the distance that is in the range of the experimental error. **C**, Radial distance distribution for the proximal state of the +3xCTCF sites/+RAD21 condition overlaid with the distributions of the proximal states from the three best matching parameters sets when comparing only the average radial distances. **D**, Heatmap showing the agreement of all simulated systems (for extrusion speed 0.1 kb/s) with the experimental data. The score is as described in Fig. 2.6A (see Methods). **E**, MSDs for three conditions for extruder residence time of 5.5 min, loading rate of $0.06 (\text{Mb} \times \text{min})^{-1}$ and extrusion speed of 10 kb/s. Pairwise comparison for conditions indicated in the title of each pair of heatmaps. **F**, Heatmap showing the fold change of generalized diffusion coefficients (D) and scaling exponent (α), and represents fold change between the conditions.



Extended Data Figure E10: Polymer simulations of landscapes with two barriers at different distances.

A, Scheme of simulated polymers with varying distances between (optional) convergent loop extrusion barriers, corresponding to 100, 150, 250, 500, and 1000 kb. **B**, Duration (left) and rate of formation (right) of the HMM proximal state detected on simulated pairwise distances (after addition of experimental error) between monomers in the presence or absence of extrusion barriers, as a function of the intervening linear genomic distance. Lines are means, shaded areas are s.e.m. Note that the average duration of the HMM proximal state slightly decreases although the average duration of the underlying cohesin-mediated CTCF-CTCF interaction does not (see panel C). This is due to non-CTCF mediated interactions, which also contribute to the proximal state, and decrease with increasing genomic distance. **C**, Average duration (left) and rate of formation (right) of the looped state (that is cohesin-mediated CTCF-CTCF interaction) extracted from polymer simulations. Lines are means, shaded areas are s.e.m.

2. COHESIN AND CTCF CONTROL THE DYNAMICS OF CHROMOSOME FOLDING

Chapter 3

Cohesin dynamics on DNA: Cohesion or Extrusion?

Contribution

This project was a collaboration with Nessim Louafi^{1,2}, Simon Gaudin^{2,3}, Luca Giorgetti¹, Elphège Nora⁴, Leonid Mirny^{5,6}, Simon Grosse-Holz^{5,6} and Laura Caccianini^{7,8}. My contributions included the conception of this study as well as performance of the experiments, including the genome engineering, live-cell imaging and image analysis together with Nessim Louafi. Elphège Nora generated and kindly provided the degron system cell lines. Luca Giorgetti, Elphège Nora, Leonid Mirny, Simon Grosse-Holz and Laura Caccianini helped with the conception of the study and the discussion of the results. The results of this chapter have not yet been published.

¹ Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

² Université de Montpellier, Montpellier, France

³ École Normale Supérieure de Lyon and Université Claude Bernard Lyon I, Université de Lyon, Lyon, France

⁴ Cardiovascular Research Institute, University of California San Francisco; Department of Biochemistry and Biophysics, University of California San Francisco; San Francisco, CA, USA

⁵ Department of Physics and Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶ Institut Curie, PSL Research University, Sorbonne Université, CNRS UMR3664, Laboratoire Dynamique du Noyau, 75005 Paris, France

⁷ Laboratoire Physico Chimie Curie, Institut Curie, PSL Research University, Sorbonne Université, CNRS UMR168, 26 Rue D'Ulm, Paris, 75005, France

⁸ Massachusetts Institute of Technology, Department of Biology, 31 Ames St., Cambridge, MA 02142, USA

3.1 Abstract

Cohesin is a SMC complex involved in a variety of functions in shaping the 3D genome, of which loop extrusion and sister-chromatid cohesion are the most prominent ones. While the effects on chromosome structure have been extensively studied, little is known about the dynamics of cohesin when it is bound to DNA. Here, we combined tracking of multiple random DNA loci with single-particle tracking of cohesin in the same living cells. Applying restoration techniques based on deep learning to enhance noisy signals, we could follow dynamics of both, cohesin and DNA loci, for up to 30 min. We show that DNA motion is accelerated both in the absence of NIPBL as well as Sororin indicating that both, loop extrusion and sister-chromatid cohesion, contribute to the constraining behavior of cohesin. Furthermore, we see that cohesin motion on DNA scales with an anomalous exponent of 0.44 which is a more constrained motion than DNA, indicating that cohesin residing at loop bases or “cohesive” cohesin accounts for a high fraction of the observed dynamics.

3.2 Introduction

SMC complexes perform a variety of functions in mammalian cells. In particular, the cohesin complex has been shown to have roles in S and G2 phase where it holds together the sister chromatids after replication (Schmitz *et al.*, 2007), but also in DNA repair and recombination (Birkenbihl & Subramani, 1992; Y. Zhang *et al.*, 2022) and as a loop extruder in interphase where it establishes TAD structures by interacting with two strands of DNA in *cis* (Fudenberg *et al.*, 2016; S. Rao *et al.*, 2017). Cohesin is also the only SMC complex that has been shown to also interact with two strands of DNA in *trans* and to have an alternative, topologically entrapping binding mode to DNA (Nagasaka *et al.*, 2023).

Although the results from the previous chapter showed that cohesin constrains chromosome dynamics and that both cohesin and CTCF control the interactions of two loci in *cis*, it remains to be understood if this is because of the loop extrusion activity of cohesin or its role in sister-chromatin cohesion. As cohesion establishes bonds between sister-chromatids after replication, this could indeed also lead to a constrained chromosome motion. Therefore, we require a better understanding of the dynamics of cohesin itself, when bound to DNA. While the binding dynamics and the residence time of cohesin have been studied using FRAP, but also SMLM (Hansen *et al.*, 2017; Wutz *et al.*, 2017; Nora *et al.*, 2020), little is known about its dynamics once it has associated with DNA and performs one of either functions. For this, it would be necessary to follow single cohesin molecules over a period of multiple minutes, ideally for the length of the binding time to DNA. Furthermore, this would require being able to distinguish what function each individual cohesin molecule is engaged in whilst bound to DNA: transient interaction with DNA, loop-extruding, or bound and stabilized by either Sororin or CTCF and PDS5. While the velocity of loop extrusion by cohesin could be measured *in vitro* on otherwise naked DNA to be 1 kb/s (Davidson *et al.*, 2019), it is further in question, what is the speed of loop extrusion *in vivo*, where cohesin has to navigate a complex chromatin fiber and other bound

complexes that might interfere with its activity (Dequeker *et al.*, 2022).

While previous measurements have characterized cohesin dynamics on short time scales in the seconds range (Nora *et al.*, 2020), studying its interactions and functions on DNA that itself moves slowly, requires observation periods in the range of tens of minutes (Gabriele *et al.*, 2022; Mach *et al.*, 2022). Here, we use a SPT approach following Halo-tagged RAD21 molecules in mESCs over a time span of approximately 30 min. Optimization of numerous imaging parameters as well as novel ways of restoring noisy SPT data allow us to reconstruct trajectories of individual RAD21 molecules that can be over 600 s long. With this technique in hand, we measured cohesin and DNA dynamics, when perturbing different functionalities of cohesin through degrons systems for its regulators (NIPBL and Sororin).

3.3 Results

3.3.1 Cohesin and DNA dynamics in living cells

We set out to visualize different types of motion of cohesin on DNA, in order to potentially discriminate different sub-populations of molecules that associate with DNA in either one of four hypothesized ways: Firstly, cohesin can either temporarily associate with DNA and potentially diffuse on it without engaging in either of its functions. Secondly, cohesin can actively move DNA by performing loop extrusion. Thirdly, it can enforce the previously formed loop by stalling at CTCF sites and lastly, it performs sister-chromatid cohesion by engaging two DNA molecules in *trans*. We hypothesized that these modes should show a distinct motion with diffusion and active loop extrusion being faster processes, while loop-enforcement and cohesion should show a stalled and slow motion (Fig. 3.1a).

To discriminate these modes, we used mESCs, where the endogenous *Rad21* locus had been tagged with a C-terminal HaloTag, to visualize individual RAD21 molecules by labeling at a sub-optimal concentration (25 pM, sparse labeling) with HaloLigand-JaneliaFluor646. It had previously been shown that the cohesin core complex (including RAD21) is also stably associated when it is not bound to DNA, as single components (RAD21, SMC1 and SMC3) co-purify in immunoprecipitation assays (Hansen *et al.*, 2017). Studying RAD21 dynamics is therefore a good proxy for cohesin motion. We further engineered these cell lines to carry multiple random integrations of the TetO array that can be visualized by binding of the respective TetR fused to the fluorescent protein tdTomato as described in Mach *et al.*, 2022 (see chapter 2). These clonal cell lines allowed us to follow cohesin and DNA dynamics at the same time in the living cell (Fig. 3.1b and c).

We then studied their motion when perturbing cohesin's functions by depleting its regulators through an AID system: The endogenous loci of the *Nipbl* and *Cdca5* (Sororin) genes were tagged with the AID-tag and an eGFP (Fig. 3.1b, see genotypes in the Supplementary Information). After depletion of NIPBL (as indicated by the absence of eGFP fluorescence) by addition of 500 μ M auxin for 6 h, cohesin is not able to perform loop extrusion (Schwarzer *et al.*, 2017; Davidson *et al.*, 2019; Nagasaka

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

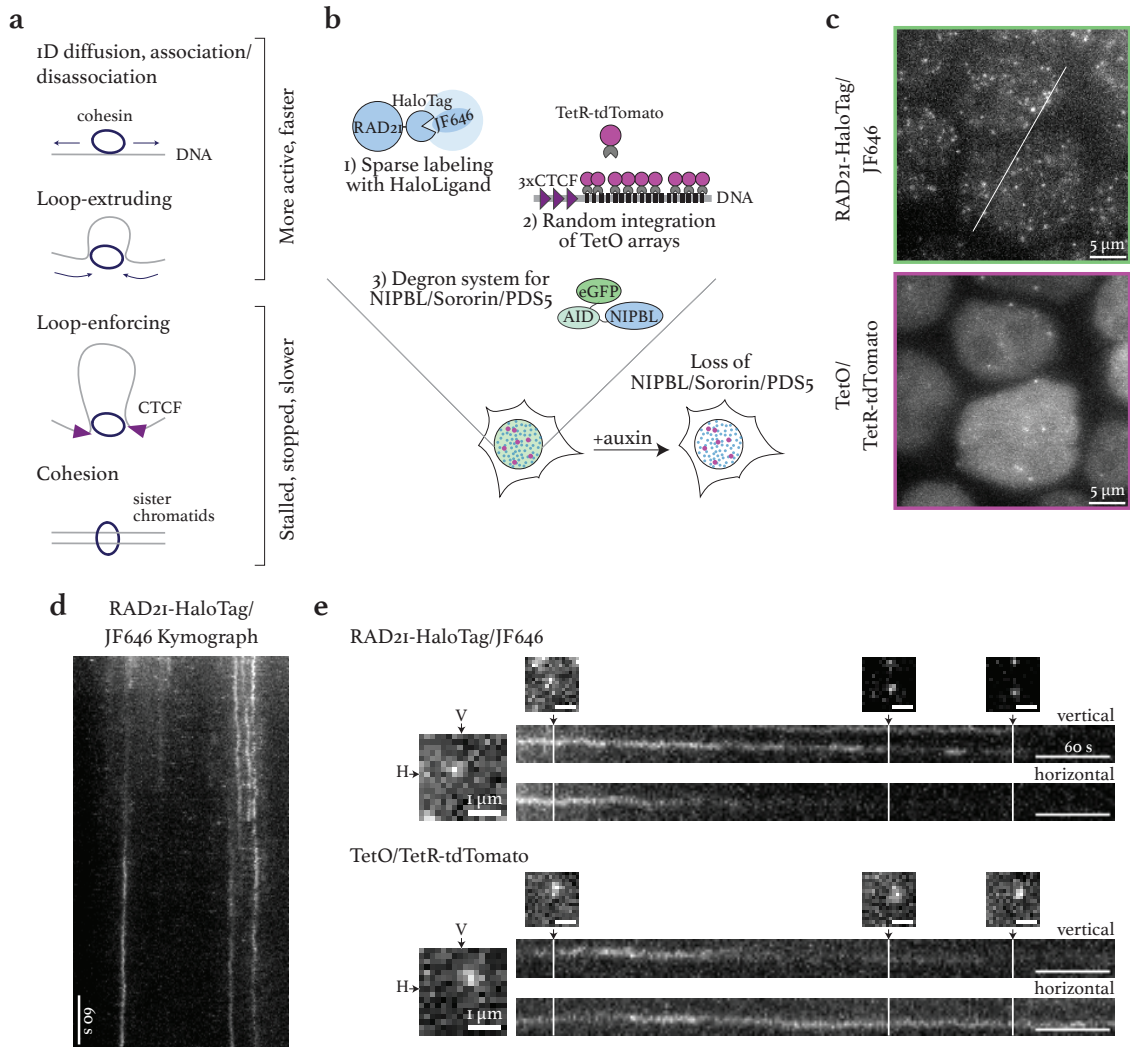


Figure 3.1: SPT of cohesin and DNA in the same cell to observe signatures of loop extrusion. **a**, Different modes of cohesin interacting with DNA that will ultimately lead to different types of motion of cohesin on DNA. **b**, Clonal cell lines carrying random integrations of the Tet operator (TetO). The cell lines are further tagged with a HaloTag at the endogenous *Rad21* locus and engineered to either be an auxin-inducible degon system for eGFP-tagged NIPBL, Sororin or PDS5, respectively. **c**, Representative images of Sororin-AID-eGFP cells containing 3 \times CTCF-TetO and RAD21-Halo (exposure time tdTomato: 20 ms, RAD21-HaloTag-JF646: 100 ms, maximum intensity projection). **d**, Kymograph of the RAD21-HaloTag-JF646 signal from the slice along the white line in the lower panel in **c**. Exposure time RAD21-HaloTag-JF646: 100 ms, max. intensity projection, dt = 2 s. **e**, Kymographs as in **d** of a single spot over time either visualizing a single cohesin molecules (top) or a DNA locus (bottom). The two kymographs per signal show a vertical and horizontal cross-section through the spot signal (cross-section demarcated by the letters H and V). Inlets show the spot at different time points along the movie (t = 30 s, 300 s and 400 s). Image depicting the initial spot from t = 0 s. Both signals (RAD21-HaloTag/JF646 and TetO/TetR-tdTomato spots) originated from the same region of interest. Exposure time RAD21-HaloTag-JF646: 100 ms, max. intensity projection, dt = 2 s.

et al., 2023), while in the absence of Sororin (3 h of depletion) sister-chromatid cohesion is impaired (Rankin *et al.*, 2005; Schmitz *et al.*, 2007; Mitter *et al.*, 2020). As a control, we observed DNA and cohesin dynamics in untreated conditions as well as in an E14 “WT” cell line that did not carry a degron system.

These clonal cell lines were then imaged before and after depletion of the regulators using highly inclined and laminated optical sheet (HILO) microscopy (Tokunaga *et al.*, 2008). We acquired a small 3D stack (z-stack of 1.2 μm , $\text{dz} = 300 \text{ nm}$) every 2 s for 10 min (or 10 s for 30 min) in order to minimize the interruption of tracks due to the spot’s movement along the z-axis (Fig. 3.1d and e). We chose to acquire the dynamics at these time scales, as previous measurements of DNA motion have shown a very slow motion with diffusion coefficients of $D \sim 10^{-2} \mu\text{m}^2/\text{s}^\alpha$ and scaling exponents of $\alpha \sim 0.5 - 0.6$ (see 2.3.1) (Gu *et al.*, 2018; Gabriele *et al.*, 2022). According to previous SPT measurements of cohesin with fast acquisition times (Nora *et al.*, 2020), 70% of all cohesin molecules are bound to DNA at a given time with slow dynamics and a distribution of D also centered around $D \sim 10^{-2} \mu\text{m}^2/\text{s}^\alpha$, while freely diffusing cohesin shows faster dynamics with $D \sim 10^1 \mu\text{m}^2/\text{s}^\alpha$. We therefore argued that we could restrict our observations to DNA-bound molecules by choosing a long exposure time of 100 ms in order to motion-blur the freely diffusing particles. This would allow a precise localization only of the bound molecules.

We further increased the length of the tracks by minimizing the effects of photo-bleaching and photo-toxicity that can be expected in SPT experiments: For this, we made use of image restoration techniques to denoise the images retroactively which allowed us to significantly decrease the laser power when acquiring image series in the first place. We verified our approach by comparing the performance of the spot detection on the raw images taken in fixed cells (see Fig. 3.2a) to the performance on denoised images or on a set of images taken of the same FOV with ideal illumination settings to achieve the best possible resolution (see Fig. 3.2c). The ground truth was a manual annotation of the images taken under the ideal settings. While spot detection performs best on the ideal images, the performance on denoised images is consistently better than on raw images. Furthermore, denoising does not decrease the localization precision by introducing artifacts as compared to the raw images, but also does not outperform images taken under ideal settings (see Fig. 3.2c). We therefore chose a laser power of 20% ($\sim 4.4 \text{ mW}$ in wide-field mode on entire FOV) and a subsequent denoising by N2V2 to increase the period over which we could image the cells and reliably detect spots, to 30 min (Höck *et al.*, 2022).

Detected spots in both channels were linked over time using a linear assignment problem tracker (LAPTrack) (Fukai, 2021; Fukai & Kawaguchi, 2022). We optimized the acquisition settings, especially the time interval as well as the labeling density with HaloLigand, to be able to track them correctly over time, as high density of molecules and long time intervals can prevent the correct linkage of molecules. Optimal tracking for a time interval of $\text{dt} = 2 \text{ s}$ over 10 min (or 10 s over 30 min) was achieved with a labeling concentration of 25 pM HaloLigand-JF646. Tracks of cohesin molecules and DNA loci were pooled and then corrected for the motion of the nucleus by subtracting the roto-translational component that all tracks within a cell have in common, from the motion of the individual track. The performance of the correction was evaluated by comparison of the motion-corrected MSD to the

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

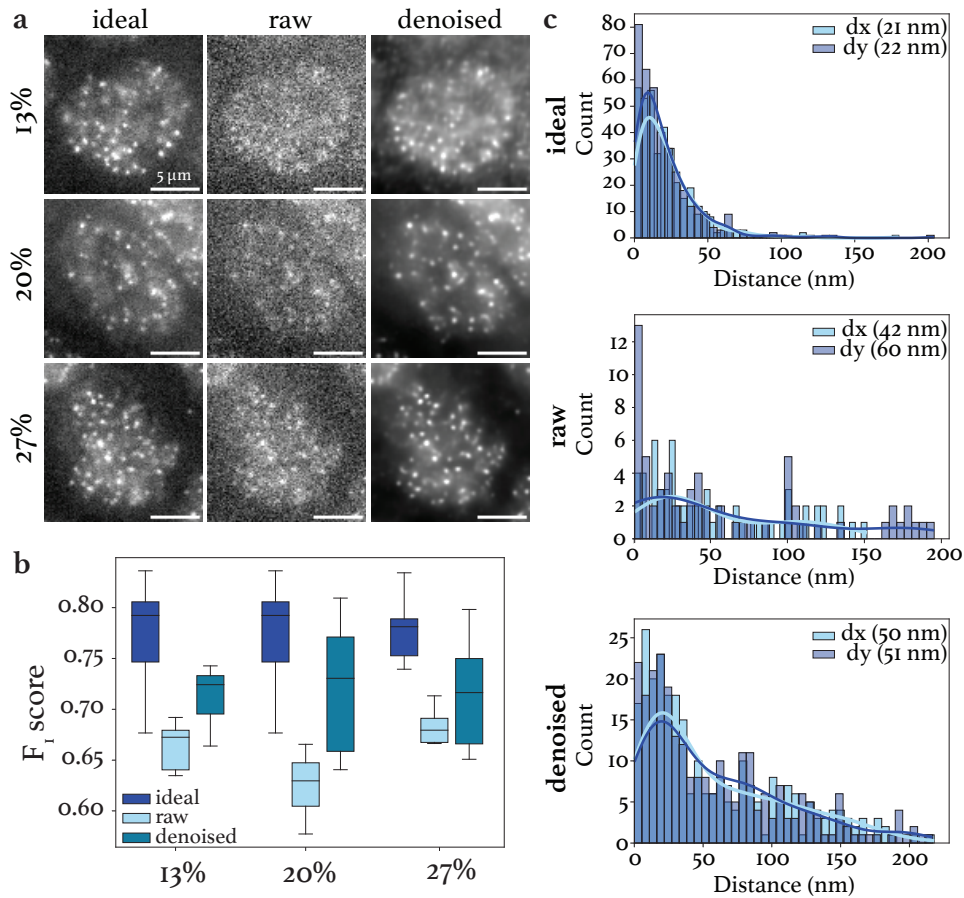


Figure 3.2: Denoising allows for extended imaging of cohesin molecules. **a**, Representative images of images taken on fixed cells with either ideal settings (exposure time: 200 ms, EMCCD gain: 50, maximal laser power), raw images (exposure time: 100 ms, EMCCD gain: 300, laser power as indicated) and denoised images (denoised with Noise2Void2 (N2V2), originating from the raw images). The same field of view (FOV) is shown for the ideal, raw and denoised image, respectively. **b**, F_1 -score (harmonic mean of precision and recall) for the performance of the spot detection on the three sets of images. F_1 -score was calculated based on comparison with a manually annotated ground truth based on the ideal images. Laser powers as indicated. **c**, Localization precision in xy-dimensions for a laser power setting of 20% for the raw and denoised images. The localization precision was calculated as the standard deviation of the Euclidean distance between the same spot acquired in two consecutive frames in a time lapse on fixed images. Values in parentheses are mean values for the localization in the indicated dimension.

MSD of pairwise distances between spots of any pair of two tracks (Fig. 3.3a). An overview of the entire image analysis pipeline is shown in the Supplementary Information.

3.3.2 Cohesive and extrusive cohesin impose constraints on DNA motion

We first set out to characterize the motion of DNA by itself under conditions where either NIPBL or Sororin have been depleted to investigate whether loop extrusion or sister-chromatid cohesion is responsible for introducing constraints to DNA motion. For this, we imaged the randomly integrated DNA loci before and after depletion of NIPBL or Sororin. As a control, we compared the motion of untreated cells from degron cell lines to an WT cell line with only TetO insertions and a *Rad21*

locus tagged with HaloTag. We found a good agreement in the scaling and the instantaneous diffusion coefficient of the MSDs across the degron cell lines (Fig. 3.3b). The observed values agree well with previously reported anomalous exponents α of $\sim 0.5 - 0.6$ and diffusion coefficients D_{inst} of $\sim 10^{-2} \mu\text{m}^2/\text{s}^\alpha$ for DNA motion with unimodal distributions for these values extracted from time-averaged MSDs of individual tracks (Fig. 3.3c and d). We note here that the control cell line had a slightly higher instantaneous diffusion coefficient. A possible explanation for this is the leakiness of the degron system that already in untreated cells might lead to a degradation of a small proportion of the tagged protein of interest.

Next, we compared the motion of the DNA loci after loss of NIPBL to their motion under untreated conditions to also potentially account for the positional bias of the integrations across different clonal cell lines. In the absence of NIPBL, chromosome dynamics showed a higher instantaneous diffusion coefficient (Fig. 3.4a and b). Since the scaling exponent remained constant (Fig. 3.4c), D_{inst} of the two conditions can be directly compared: It increased by 1.5-fold which is in excellent agreement

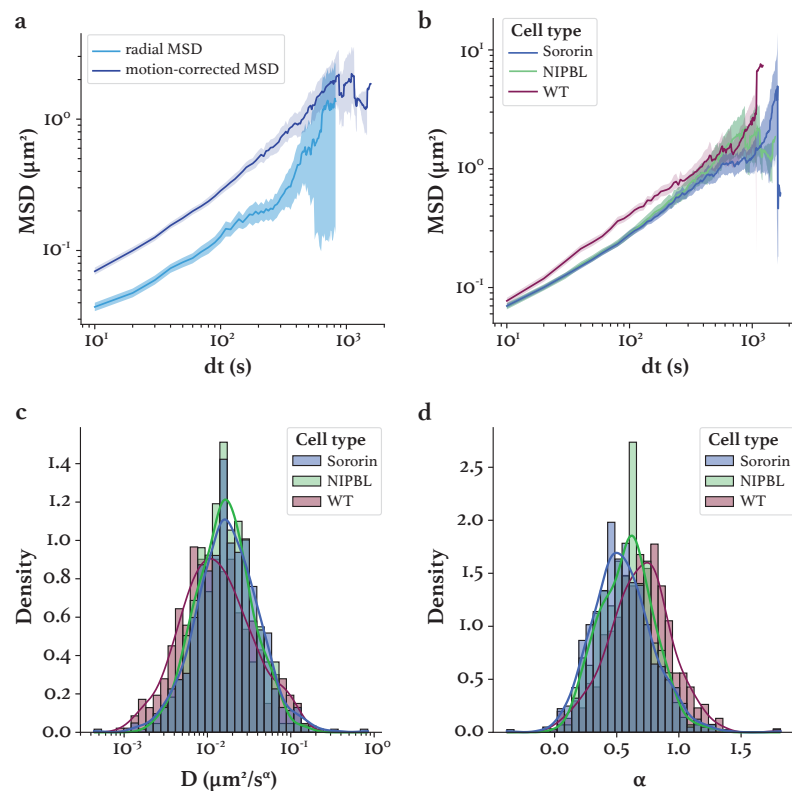


Figure 3.3: Set-up allows for accurate observation of dynamics of DNA loci. **a**, Comparison between radial MSD (light blue) calculated between any pair of tracked DNA loci over time and the motion-corrected MSD (dark blue) of the same. A factor of 2-fold difference in instantaneous diffusion coefficient is expected between radial MSD and MSD due to the fact that two loci get considered for the radial MSD. **b**, Comparison of DNA motion dynamics in three different cell lines without treatment with auxin: WT (magenta), Sororin-AID-eGFP (blue) and eGFP-AID-NIPBL (green). **c**, Distribution of instantaneous diffusion coefficient for the three cell lines in b. Average values: WT: $0.012 \mu\text{m}^2/\text{s}^\alpha$, Sororin-AID-eGFP: $0.016 \mu\text{m}^2/\text{s}^\alpha$, eGFP-AID-NIPBL: $0.017 \mu\text{m}^2/\text{s}^\alpha$. **d**, Distributions of anomalous exponents α for the three cell lines. Average values: WT: 0.69, Sororin-AID-eGFP: 0.53, eGFP-AID-NIPBL: 0.6.

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

with previous measurements in cell lines where cohesin itself was degraded (Nozaki *et al.*, 2017; Mach *et al.*, 2022). This indicates that loop extrusion contributes to imposing constraints on DNA.

We compared these results to depletion and imaging experiments in a degron cell line where Sororin, which regulates cohesin's role in sister-chromatid cohesin, was depleted. We found that also under these conditions, where cohesion could not be performed anymore, DNA motion showed higher D_{inst} while the scaling exponent remained constant and both showing a unimodal distribution for values calculated from individual tracks (time-averaged MSD). We conclude that the constraining behavior of cohesin is most likely to stem from the formation of bonds, both in *cis* from loop extrusion, but also in *trans* from cohesion.

3.3.3 Comparison of cohesin motion and DNA motion

In order to potentially discriminate cohesion's functions *in vivo*, we next described the motion of cohesin itself on DNA. Under the assumption that all observed cohesin molecules are bound to DNA, the measured motion of these spots is therefore a convolution of the motion of DNA and of the cohesin

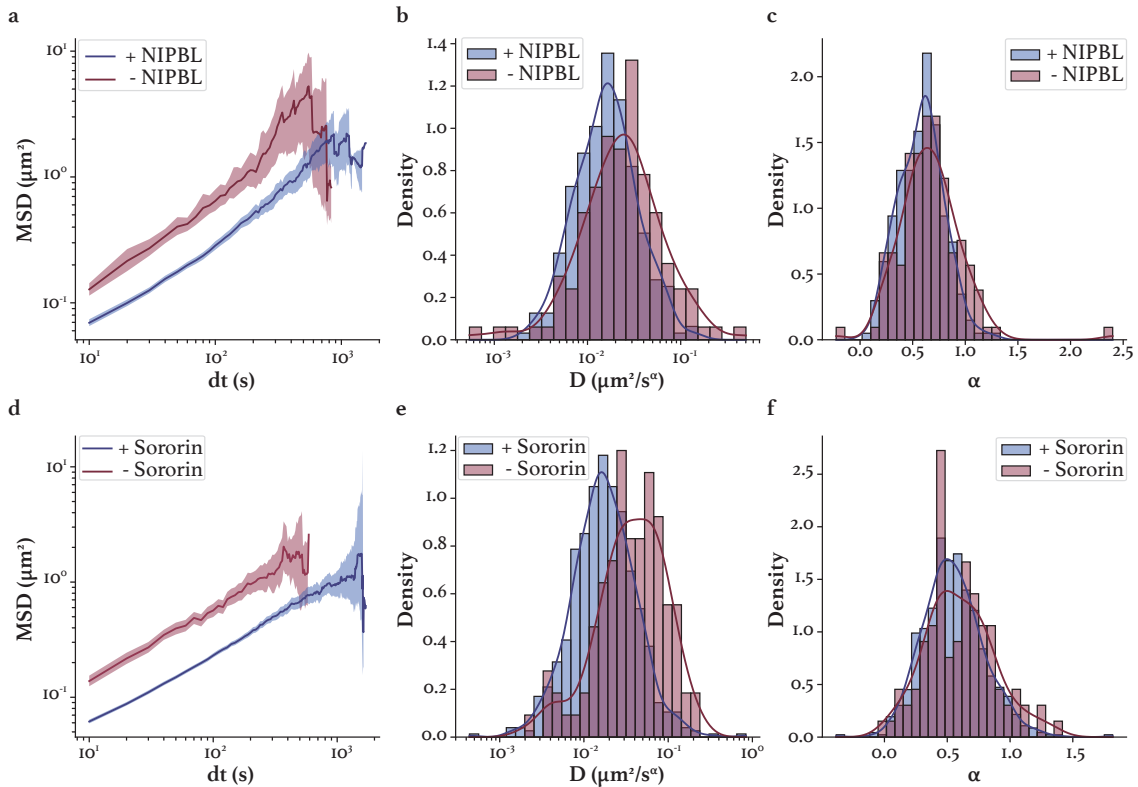


Figure 3.4: Cohesion and loop extrusion contribute to constraining chromosome motion. **a**, MSD of DNA loci motion in eGFP-AID-NIPBL cells before (blue) and after (red) treatment with auxin (loss of NIPBL). **b**, Distribution of instantaneous diffusion coefficient for the conditions in **a**. Average values: + NIPBL: $0.016 \mu\text{m}^2/\text{s}^\alpha$, - NIPBL: $0.0023 \mu\text{m}^2/\text{s}^\alpha$. **c**, Distributions of anomalous exponents α for the conditions in **a**. **d**, MSD of DNA loci motion in Sororin-AID-eGFP cells before (blue) and after (red) treatment with auxin (loss of Sororin). **e**, Distribution of instantaneous diffusion coefficient for the conditions in **d**. Average values: + Sororin: $0.017 \mu\text{m}^2/\text{s}^\alpha$, - Sororin: $0.0039 \mu\text{m}^2/\text{s}^\alpha$. **f**, Distributions of anomalous exponents α for the conditions in **d**.

molecules on DNA. Therefore, we independently measured multiple randomly integrated TetO loci as a proxy for global DNA motion so that cohesin motion could be compared to the underlying DNA motion. Fig. 3.5 shows the MSD of both tracked cohesin and DNA loci.

The initial slope of the MSD of RAD21 motion in Fig. 1.1a clearly shows that cohesin motion scales with a lower exponent (0.44 as compared to ~ 0.58 for DNA motion, extracted from linear fit on initial 5 data points) and that also the D_{inst} varies considerably. However, in the case where scaling exponents differ, D_{inst} cannot be compared directly as its values depends on α . In general, the motion of cohesin is slower and values for D_{inst} are more broadly distributed. Of note here is that the values for D_{inst} this experiment are higher (also for DNA motion: $0.055 \mu\text{m}^2/\text{s}^\alpha$ on average) than previously described. This could potentially be due to a less accurate motion correction because of the limiting number of tracks generated in these experiments. In conclusion, we could show that cohesin moves even slower than the average motion of DNA with an even more sub-diffusive behavior ($\alpha = 0.44$). This is consistent with polymer simulations that suggest that a loop base, i. e. a bond between two parts of the polymer, would show a more constrained behavior than other loci in the polymer chain or a molecule bound on the polymer and moving with it (Pavel Kos, personal communication).

We then also wanted to characterize the cohesin motion in the absence of Sororin or NIPBL to investigate whether there is a sub-population of cohesin molecules that is lost under those conditions. So far, we were not able to track a sufficient number of RAD21 molecules in Sororin- or NIPBL-depleted cells, as depletion of either factor changes the abundance of bound cohesin molecules on DNA (Nora *et al.*, 2020). The limiting labeling concentration of 25 pM that was found ideal under WT conditions, is however insufficient to observe enough tracks when the amount of cohesin bound to DNA is changed. In future experiments, we will therefore characterize the abundance of cohesin and of the observable RAD21 molecules for each experimental condition and adjust the labeling concentration.

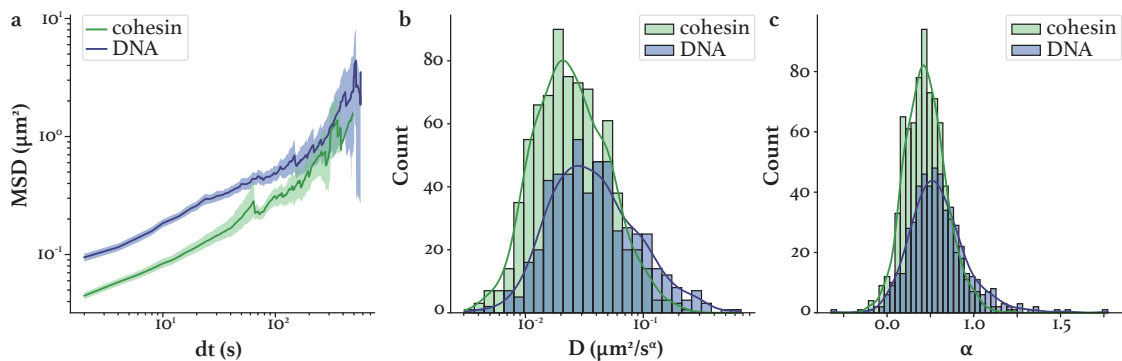


Figure 3.5: Cohesin moves slowly on DNA. **a**, MSD of DNA loci motion (blue) and RAD21 motion (green). **b**, Distribution of instantaneous diffusion coefficient for the conditions in **a**. Average values: DNA: $0.055 \mu\text{m}^2/\text{s}^\alpha$, RAD21: $0.0031 \mu\text{m}^2/\text{s}^\alpha$. **c**, Distributions of anomalous exponents α for the conditions in **a**. Average values: DNA: 0.58, RAD21: 0.44.

3.4 Discussion and outlook

In this chapter, we describe a new approach to study chromosome dynamics in combination with the dynamics of one of the motor proteins, cohesin, that shapes its folding patterns by creating bonds via loop extrusion or sister-chromatid cohesion. Using new techniques for image restoration, we are able to follow cohesin and DNA over periods of up to 30 min leading to tracks that are up to 10 min long for cohesin. This is in the range of measured residence times for cohesin (Hansen *et al.*, 2017; Wutz *et al.*, 2020), meaning that we can potentially follow an entire life cycle of a cohesin binding event on DNA. This has exciting implications for our findings: With a 150 kb-sized loop lasting on average 5-15 min, this would mean that we are able to capture the motion of cohesin while extruding a loop as well as enforcing this loop, since both events have been estimated to take place in the minutes range (Davidson *et al.*, 2019; Gabriele *et al.*, 2022). We are currently still not able to track under conditions where the abundance of cohesin on DNA is reduced, however, more tests with adjusted labeling concentrations will likely overcome this limitation.

Here, we already observe a motion of cohesin on DNA that is even slower than that of DNA itself. This could indicate that the most abundant fraction of cohesin molecules that we are observing are indeed molecules at loop bases that show a more constrained behavior, or molecules performing sister-chromatid cohesion. This is suggested from polymer simulations, where introduction of bonds would lead to more constrained behavior (Mach *et al.*, 2022). Without data from experiments in NIPBL-depleted cells, we cannot conclude whether there is also a fraction of observable cohesin molecules that are performing more active roles on DNA, such as loop extrusion.

However, due to the slow motion of DNA as well as a low estimated speed of loop extrusion *in vitro*, it is possible that signatures of loop extrusion might be masked by the motion of the underlying polymer itself. Since the motion of the RAD21 molecules observed here is a convolution of DNA motion and their motion on DNA, loop extrusion events will not be observed, if the *in vivo* velocity of it is described by a similar motion than that of DNA. In an attempt to overcome this potential limitation, we aim to perform SPT experiments also under conditions where loop extrusion might occur at increased velocities, as suggested by recent literature (van Ruiten *et al.*, 2022): This can be achieved in the context of a cell line with two orthogonal degron systems (AID and FKBP system) that allow for depletion of PDS5 with auxin and partial depletion of RAD21 by addition of 45 nM dTAG-13 (Fig. 3.6a). In this scenario, depletion of PDS5 will lead to a higher velocity and processivity of loop extrusion, that also prevents stalling by CTCF, since PDS5 no longer “brakes” cohesin. Partial depletion of RAD21 itself (only 30% of WT condition, Elphège Nora, personal communication), prevents collision and stalling of cohesin molecules, thereby potentially leading to an increased fraction of cohesin molecules actively performing extrusion as compared to a loop enforcing state (Fig. 3.6b).

Finally, we were also able to show that both functions of cohesin contribute to the decrease in constraints observed upon depletion of RAD21. We show that both, in the absence of NIPBL or Sororin, a similar increase in the instantaneous diffusion coefficient can be observed, however to

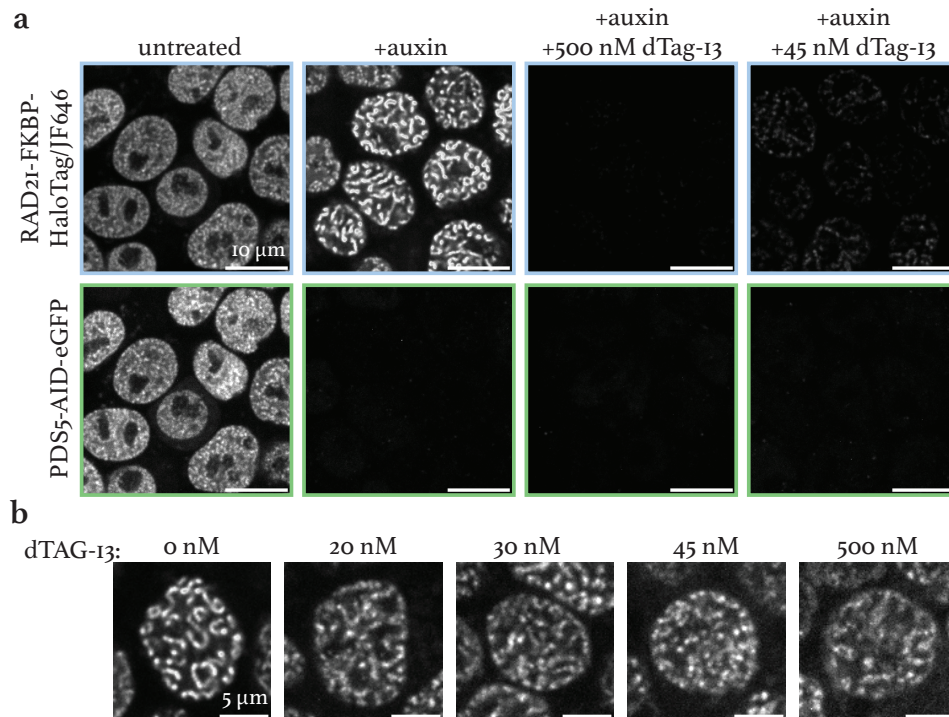


Figure 3.6: Increasing the velocity of cohesin on DNA: PDS5 and partial RAD21 depletion. **a**, Cell line carrying an AID degnon system for PDS5 which was tagged with an eGFP, as well as a FKBP system to degrade RAD21 and visualize it with a HaloTag for SPT. Depletion of PDS5 and RAD21 for 6 h with 500 μ M auxin and varying concentrations of dTAG-13 (as indicated in the panel) leads to loss of fluorescence of eGFP and HaloTag-HaloLigandJF646. **b**, "Vermicelli" chromosomes can be observed when depleting PDS5 for 6 h as indicated by cohesin positioning in the RAD21-FKBP-HaloTag-JF646 channel, however their degrees vary depending on co-depletion of RAD21. Here, contrast was adjusted manually for each image to visualize the vermicelli chromosomes, whilst accounting for changing levels in RAD21-HaloTag levels.

varying extents. It will be crucial to compare the motion of cohesin itself under these conditions in order to understand, if the observed changes stem from loss of bonds enforced by cohesin.

3.5 Methods

Culture of embryonic stem cells

All cell lines are based on E14 mouse embryonic stem cells mESCs. E14 Sororin-AID-eGFP, RAD21-Halotag, Rosa26-Tir1 (clone EN229.3.1) and RAD21-Halotag (clone EN130.1) were published in Nora *et al.*, 2020 and were kindly provided by Elphège Nora (University of California, San Francisco) together with E14 eGFP-AID-NIPBL, Tigre-Tir1 (clone EN273.7) and E14 eGFP-AID-PDS5A/B, RAD21-FKBP-Halotag, Tigre-Tir1 (clone EN387.4) (unpublished). For genome engineering, cells were cultured on gelatin-coated culture plates in Glasgow Minimum Essential Medium (GMEM) (Sigma-Aldrich, G5154) supplemented with 15% foetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μ M β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U/ml leukemia inhibitory factor (Miltenyi Biotec, premium grade) in 8% CO₂ at 37 °C. Cells were tested for mycoplasma contamination regularly and no contamination was detected. For live-cell imaging experiments, cells were cultured in Fluorobrite Dulbecco's Modified Eagle Medium (DMEM) (Gibco, A1896701) supplemented with 15% foetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μ M β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U/ml leukemia inhibitory factor (Miltenyi Biotec, premium grade) and with 2i inhibitors (1 μ M MEK inhibitor PDO35901 (Axon, 1408) and 3 μ M GSK3 inhibitor CHIR 99021 (Axon, 1386)).

Generation of mESC lines carrying random integrations of TetO arrays

Random TetO insertions in the degron cell lines (E14 eGFP-AID-NIPBL, E14 Sororin-AID-eGFP and E14 PDS5A/B-AID-eGFP) were achieved as described in the previous chapter (see 2.5). In brief, using Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's recommendations, 0.5×10^6 cells were transfected with 2 μ g PB-3xCTCF-TetO vector, 200 ng PB-TetR-tdTomato and 200 ng pBroad3_hyPBase_IRES_tagRFPt (Redolfi *et al.*, 2019). After 5 days, transfected cells were sorted FACS at 581 nm (tdTomato) to check for tdTomato expression and clonal expansion from single cells afterwards. 10 days after sorting the cells were re-plated onto glass bottom plate with high performance #1.5 cover glass (96-well, Cellvis, P96-1.5H-N). Clonal cell lines were screened by microscopy for >10 insertions of TetO/cell and a good SNR 24 h after re-seeding. Expanded clones were further genotyped by PCR for the absence of random integration of the PiggyBase itself. Primers used for genotyping are listed in Mach *et al.*, 2022.

Labeling of RAD21-Halo

For the imaging experiments, laminin-coated (1-2 μ g/ml Laminin (Sigma-Aldrich, L2020) in PBS at 37°C overnight) 35 mm glass-bottom dishes (Mattek, P35G-1.5-14-C) were prepared at least 48 h prior to imaging. 2×10^6 cells were seeded onto coated dishes in Fluorobrite medium (as described above) 24 h before imaging. Depletion of Sororin, NIPBL or PDS5A/B through the AID system was achieved by adding 500 μ M auxin (Sigma-Aldrich, I5148-2G) either 3 h (Sororin) or 6 h (NIPBL) prior to the start of the imaging to the Fluorobrite culturing medium. Partial depletion of RAD21 through the FKBP system was achieved by adding either 45 nM or 500 nM dTAG-13 (Sigma-Aldrich, SML2601-1MG) for 6 h prior to imaging.

Stochastic labeling of Halo-tagged RAD21 molecules with Halo-Ligand-JaneliaFluor® 646 (Tocris Bioscience, custom synthesis based on Cat. No. 6993) was performed as follows: Cells were incubated with Fluorobrite medium containing either 25 pM, 50 pM, 75 pM or 100 pM of Halo-Ligand-JaneliaFluor® 646 (Tocris Bioscience) for 30 min immediately prior to imaging. After incubation, cells were subsequently washed three times with PBS and fresh Fluorobrite medium was added (without HaloLigand, potentially also containing 500 μ M of auxin or dTAG-13, depending on the imaging condition).

For fixed cell measurements that help estimate the localization error and benchmark the performance of the denoising strategy and spot detection, cells were prepared as described above. After the labeling of Halo-tagged RAD21 molecules and washing steps with PBS, the cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences, 15710) in PBS for 30 min at RT. The cells were washed three times in PBS. To achieve comparable background fluorescence levels, Fluorobrite medium was added to the Mattek dish for imaging.

Live-cell imaging

Imaging was performed with a Nikon Eclipse Ti-E inverted widefield microscope equipped with a Total Internal Reflection Microscopy iLAS2 module (Roper Scientific), a Perfect Focus System (Nikon) and motorized Z-Piezo stage (ASI) using a CFI APO TIRF 100 \times , 1.49 NA oil immersion objective (Nikon). The microscope was operating in HILO mode (Tokunaga *et al.*, 2008) with a 640 nm, 150 mW Toptica iBEAM SMART laser and a 561 nm 200 mW Coherent Sapphire laser as excitation lasers. The laser power as well as the camera alignment was checked periodically to account for fluctuations in laser intensities and mis-alignments. Images were collected on two precisely aligned back-illuminated Evolve 512 Delta EMCCD cameras with a dixel size of 16 μm \times 16 μm (Photometrics). Cells were maintained at 37 $^{\circ}\text{C}$ and 8% CO_2 using an enclosed microscope environmental control setup (The BOX and The CUBE, Life Science Instruments). Movies for measurement of random TetO integrations (at 561 nm) and RAD21-Halo (at 640 nm) in the different degron cell lines were acquired with a time interval of 10 s (TetO: exposure time: 20 ms, laser power on FOV measured in wide-field mode: 1.5 mW; RAD21-Halo: exposure time: 100 ms, laser power in wide-field mode: 4.4 mW, EMCCD gain: 300 for both) in 5 z-planes (1.2 μm stack, $\text{dz} = 300$ nm) with the Visiview software (Visiview 4.4.0.12, Visitron).

To establish a test data set that showed no movement of the spots as a ground truth for spot detection and denoising, a single z-plane was acquired on fixed cells labeled with Halo-Ligand-JaneliaFluor $^{\circledR}$ 646 (Tocris Bioscience) with different laser powers to account for different levels of noise in the images. “Raw” images that model the final imaging set-up for time-lapse imaging, were acquired with either 1.4 mW, 2.8 mW, 4.4 mW or 5.8 mW and an exposure time of 100 ms (EMCCD gain: 300) on a single plane streaming 10 frames (accounting for small aberrations in between acquisition, general localization error of this set-up). “Ideal” images were subsequently acquired on the same FOV with a laser power of 22.0 mW and an exposure time of 200 ms (EMCCD gain: 50) (modeling a clean image that can act as a ground truth to validate the denoising process).

For establishing a ground truth for tracking, two sets of time lapses were acquired: Going as fast as possible, RAD21-Halo molecules were imaged on a single plane with an exposure time of 100 ms, laser power 2.8 mW or 4.4 mW, EMCCD gain: 300, streaming in time for 300 frames ($\text{dt}=100$ ms). In a “slower” imaging modality that preserves the z-stack requirements of the final imaging modalities, but chooses the fastest time interval possible, movies were acquired at a frame rate of $\text{dt}=2$ s with a z-stack of 5 planes (1.2 μm stack, $\text{dz} = 300$ nm) with exposure times and laser powers indicated in for the normal imaging settings (TetO: exposure time: 20 ms, laser power in widefield-mode: 1.5 mW;

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

RAD21-Halo: exposure time: 100 ms, laser power in wide-field mode: 2.8 mW or 4.4 mW, EMCCD gain: 300, for 300 frames).

Fluorescence microscopy of PDS5-AID-eGFP to confirm PDS5/RAD21 depletion

The cells (2×10^5) were seeded onto laminin-coated (as described above) μ -Slide 8 Well high Glass Bottom (ibidi, 80807) dishes 24 h prior to imaging. 6 h prior to imaging depletion of PDS5 and partial depletion of RAD21-Halo was achieved by adding 500 μ M auxin and 45 nM dTAG-13 (or the indicated concentration) to the culturing medium. 1 h prior to imaging the cells were stained for 30 min at 37 °C, 8% CO₂ with Halo-Ligand-JaneliaFluor® 646 (Tocris Bioscience, custom synthesis based on Cat. No. 6993) and subsequently rinsed three times with PBS. Imaging medium (containing Fluorobrite was added) and FOVs were chosen for imaging before the end of the incubation period. Imaging was started at the end of the incubation period with the predefined stage positions.

Images were acquired on a Nikon Ti2-E Eclipse inverted microscope equipped with a spinning disk confocal scanning unit (Yokogawa CSU W1 with Dual T2), a Perfect Focus System (Nikon), VisiTron VS-Homogenizer and motorized MS2000 X,Y, ZPiezo drive (ASI) using a CFI APO PLAN 100 \times , 1.45 NA oil immersion objective (Nikon). The illumination sources were a 639 nm , 200 mW Topica iBEAM SMART laser, a 561 nm , 200 mW Cobalt Jive laser, a 488 nm , 200 mW Topica iBEAM SMART laser and a 405 nm , 200 mW Topica iBEAM SMART laser. Exposure times for each channel were: 200 ms, EMCCD gain: 100, laser power: 100% (639 nm , RAD21-Halo-JF646), 200 ms, EMCCD gain: 100, laser power: 100% (560 nm , TetO/TetR-tdTomato), 300 ms, EMCCD gain: 200, laser power: 100% (488 nm , PDS5) and 25ms, EMCCD gain: 100, laser power: 50% (405 nm , Hoechst 33342). The pinhole size was 50 μ m.

Images were collected by a back-illuminated iXon-Ultra-888 EMCCD camera (Andor) with a dixel size of 13 μ m \times 13 μ m. The EMCCD gain was set as described above. The microscope was operated with the Visiview software (Visiview 6.0.0, VisiTron). Cells were maintained at 37 °C and 8% CO₂ using an enclosed microscope environmental control set-up. Z-stacks were taken with a $\delta z = 200$ nm and 51 planes.

Image restoration

In order to be able to follow RAD21 molecules over long periods of time and reduce effects of photobleaching and phototoxicity, the laser power during acquisition was reduced and images were instead restored through denoising using the self-supervised convolutional neural net (CNN) N2V2 (Krull *et al.*, 2019; Höck *et al.*, 2022). For this, the higher dimensional data (in this case five-dimensional) was considered as individual single 2D images that were denoised and then reconstituted to the original hyperstack. To create a training data set, 50,000 patches of a 128 \times 128 px size were randomly sampled from 10 movies taken on different days to account for the variability in laser intensity, labeling density or overall cell state across conditions (e. g. untreated vs. treated). From these patches, 90% were used as a training vs. 10% as a validation data set. The training was performed for 200 epochs. Denoised versions of all images from all movies were then predicted using the trained model which is available on Github (<https://github.com/fmi-basel/ggiorget-spt-analysis>). All scripts for denoising can also be

found on GitHub (<https://github.com/fmi-basel/ggiorget-spt-analysis>).

The performance of the denoising process was evaluated using the set of “raw” and “ideal” data sets acquired in fixed cells on the same FOV. The “raw” images were denoised using the pre-trained N2V2 model. “Ideal” images were manually annotated to generate a ground truth for spot detection (LoG detector with a high quality threshold with subsequent manual correction). Spots were then detected in an automated manner with h-max intensity thresholding on all three types of images: “Ideal”, “raw” and “denoised” images with parameters optimal to each data set (as determined by parameter sweeping based on the ground truth annotations of the “ideal” data set). Parameter sets and the overall performance of spot detection on “ideal”, “raw” and “denoised” images were evaluated by calculating the precision and recall (defined below), as well as the harmonic mean of both, the F_1 -score:

$$Precision = \frac{TP}{TP + FP} \quad (3.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.2)$$

$$F_1score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (3.3)$$

where TP are the true positives, FP the false positives and FN the false negatives (in this case unknown).

Estimation of localization precision

The localization error for illumination settings (i. e. different sets of laser powers) and before or after denoising was calculated as the standard deviation of the Euclidean distance of any two matching spots across the frames (max. linking distance for matching across frames = 1 px). For this, two consecutive frames from an image series (“ideal”, “raw” or “denoised”, streaming mode with $dt = 100$ ms or 200 ms for “ideal” images, 10 frames) were taken.

Table 3.1: Parameters for spot detection. Identified by parameter sweeping on “raw”, “denoised”, and “ideal” images based on ground truth annotations. LP = laser power.

| Imaging modality | Laser power | SD (h-max) | Threshold (Gaussian fit) |
|------------------|-------------|------------|--------------------------|
| “ideal” | 100% | 2 | 0.3 |
| “raw” | 13% | 1 | 0.4 |
| “raw” | 20% | 2 | 0.4 |
| “raw” | 27% | 2 | 0.5 |
| “denoised” | 13% | 2 | 0.2 |
| “denoised” | 20% | 3 | 0.3 |
| “denoised” | 27% | 3 | 0.2 |

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

Spot detection

Spots were detected by searching for local maxima in intensities with a given signal-to-background ratio (SBR) in the 3D stack (h-max intensity thresholding) using Extrema from scikit-image.morphology (Van der Walt *et al.*, 2014). The SD parameter was initially estimated from the standard deviation of intensity values in the image. The parameters used are indicated in Table 3.1. To achieve sub-pixel localization, a 2D Gaussian fit on the detected spot in the z-plane with highest signal intensity and a quality threshold of 0.5 on the goodness of the fit (square root of the diagonal of the covariance matrix of the fitted parameters) was applied to filter the spots. A parameter sweep based on a subset of movies annotated as ground truth data, was run to find robust and ideal sets of parameters that can be applied to all movies. This detection method was compared to results from the Laplacian of Gaussian (LoG) detector from scikit-image.feature, blob_log and deepBlink (Eichenberger *et al.*, 2021). The final parameters were a SD of twice the standard deviation of spots identified in 50 different frames randomly selected in the data set and a threshold on the goodness of fit of 0.5.

Spot linking and cell motion correction

The 2D spot coordinates were linked over time using LAPtrack with the parameters indicated in Table 3.2 (Fukai, 2021; Fukai & Kawaguchi, 2022). Tracks were manually checked for correct linkage.

Table 3.2: Parameters used for tracking with LAPtrack for the DNA channel and RAD21 channel and the nuclear masks.

| Channel | Max. linking distance (px) | Max. gap closing distance (px) | Max. no. of gaps |
|---------------|----------------------------|--------------------------------|------------------|
| DNA | 5 px | 10 px | 5 |
| RAD21 | 5 px | - | - |
| nuclear masks | 6 px | 32 px | 5 |

The nuclear background signal in the tdTomato channel was used to segment nuclei with StarDist and the standard model (Schmidt *et al.*, 2018; Weigert *et al.*, 2020). Nuclear masks were linked across time using LAPtrack (Fukai, 2021; Fukai & Kawaguchi, 2022) with the the parameters described in Table 3.2.

After spot linking for the tracks in the RAD21-Halo and TetO channel, these tracks were assigned to a cell mask based on a majority rule (highest occurrence of nuclear ID is assigned to the entire trajectory). Motion correction is then performed as described in Mach *et al.*, 2022, i. e. a roto-translation model that was calculated based on the common motion of all spots within a cell across two consecutive frames, is subtracted from the spot motion (matching with a minimum of 3 spots per pair of time frames). The documentation for this analysis can be found on Github (<https://github.com/fmi-basel/ggiorget-spt-analysis>).

MSD calculation

MSD calculation was also described in Mach *et al.*, 2022 and here performed in the same manner. Namely, short tracks with a length smaller than 10 data points are removed for further analysis. The time-averaged MSD for each trajectory is then calculated and subsequently an average across all time-averaged MSDs is built, also pooling the replicates from different imaging days, as chromatic aberrations are comparable. The MSDs are corrected for the estimated localization error (see 3.5) (Kepten *et al.*, 2013). To extract the scaling exponent (α) and the generalized diffusion coefficient (D), the ensemble average of the log-time averaged MSD between 10-100 s was fitted linearly. The effect of sister-chromatids in the TetO signal was neglected, since it can be regarded as a systematic error across conditions, if the cell cycle profiles of the conditions are comparable. Scripts used for the analysis are available on Github (<https://github.com/fmi-basel/ggiorget-spt-analysis>).

3. COHESIN DYNAMICS ON DNA: COHESION OR EXTRUSION?

Chapter 4

Discussion

While the architecture of chromosomes has been extensively studied using chromosome conformation capture techniques and DNA-FISH, very little is known about the dynamics of chromosomes and how they are influenced by factors that are thought to shape the 3D organization of the genome. In this thesis, I investigated chromosome dynamics, both on a global scale and local dynamics between two loci in *cis*. In chapter 2, I presented evidence that cohesin constrains chromosome motion globally and that both cohesin and CTCF modulate the interactions of two chromosomal loci separated by 150 kb of genomic sequence. In chapter 3, I then showed that cohesin's dynamics on DNA as well as DNA motion can be simultaneously followed for up to 30 min. Description of these dynamics upon depletion of NIPBL or Sororin, then led to the conclusion that cohesin's role in slowing down chromosome dynamics are due to the formation of bonds by both, sister-chromatid cohesion and loop extrusion.

In this chapter, I will place these findings in the context of current research in the field and describe their implications. I will discuss the strengths and limitations of the used approaches and will outline questions that arise from my research, as well as possible strategies to study them.

4.1 Cohesin globally slows down chromosome dynamics

Chromosome dynamics have been previously observed using live-cell imaging approaches of either genomic loci labeled through e. g. FROS or imaging of proteins bound to DNA, e. g. SPT of histones (Masui *et al.*, 2011; Nozaki *et al.*, 2017; Alexander *et al.*, 2019). However, in this study, I was interested in describing the dynamics of chromosomes at time scales where loop extrusion is thought to happen. Given that the velocity of loop extrusion was measured *in vitro* to be in the order of 1 kb/s (Davidson *et al.*, 2019), it is plausible that processes to establish an average-sized (~ 150 kb) cohesin-mediated loop would take at least ~ 2 -3 min. This is also in line with estimates of cohesin residence times in the range of 8-20 min (Gerlich *et al.*, 2006; Kueng *et al.*, 2006; Hansen *et al.*, 2017; Wutz *et al.*, 2020).

4. DISCUSSION

Hence, previous studies could only partially describe global chromosome motion at these time scales, mainly due to two technical limitations: Firstly, FROS can often observe only very few loci, since the labeling relies on integration of DNA sequences into the genome. Secondly, using a SPT approach of DNA-bound proteins leads to very short observation periods due to photo-bleaching effects (Boka *et al.*, 2021).

Therefore, we set out to overcome these limitations by measuring many chromosomal loci in the same cell with a TetO/TetR operator system. In order to achieve this, we made use of the Piggybac transposase that allows for multiple, but random integrations into the genome of even large sequences, such as operator arrays (Cadiñanos & Bradley, 2007). This led to clonal cell lines with multiple observable loci, thereby giving a good approximation of the overall chromosome dynamics. Furthermore, these loci could be followed over extended periods of time, given the good SNR and the fact that fluorophores can be replenished through maturation of newly produced fluorescent protein and binding turnover at the operator array (Boka *et al.*, 2021). However, imaging over longer periods of time also led to drawbacks, such as the fact that the motion of the cell becomes non-negligible. In order to be able to extract the absolute motion of each locus, while retaining good statistical power, we therefore developed a method to subtract the cell motion from the motion of each locus.

With this powerful tool in hand, it was then possible to observe the loci while inducing rapid degradation of factors involved in mediating or regulating folding processes, such as RAD21, CTCF and WAPL. This allowed us to study their influence on global motion. We showed that chromosome motion on the minutes scale is characterized by a single diffusive regime with an average anomalous exponent of ~ 0.6 across conditions; very comparable to previous measurements at specific loci (Gu *et al.*, 2018; Khanna *et al.*, 2019). Furthermore, we could quantify these differences that arise from the depletion of the aforementioned factors: The presence of cohesin on average led to a 30% decrease in the generalized diffusion coefficient (without changes in the anomalous exponent), with an equal, but opposite phenotype in the presence of higher levels on DNA (depletion of WAPL). This is very much in line with previous SPT studies from Nozaki *et al.*, 2017 in the seconds range. We therefore argue that cohesin serves to constrain chromosome motion and slows it down globally.

While these findings are consistent with polymer simulations of loop extrusion, I have to stress that these experiments were performed in un-synchronized mESCs where approximately 50% of cells are in either S or G2 phase (Extended Data Figs. E1c and E5g). It is therefore possible that this phenotype does not stem from cohesin's role in loop extrusion, but from its function in holding together sister chromatids after replication (cohesion) (Hirano & Mitchison, 1994). Both of these processes establish bonds between chromatin fibers; loop extrusion in *cis* and cohesion in *trans*. Therefore, loop extrusion is one of two possible explanations for our findings. Proving which function of cohesin is responsible for slowing down chromosome dynamics was, besides other questions, the goal of the experiments performed in chapter 3 and is discussed in section 4.7.

4.2 Global chromosome motion is not influenced by CTCF

We then measured the dynamics of chromosomes in the absence of CTCF (depletion via a degron system) or upon removal of CTCF binding sites adjacent to the operator arrays (via recombinases). In both cases, we did not see any changes in the global motion. This might seem surprising considering CTCF's role in loop extrusion as a barrier element. Indeed polymer simulations suggest that a locus that is in the vicinity of a loop base (as is the case at CTCF sites) should move differently than a locus far away from the loop base. However, we are here considering a locus moving in the 3D nuclear space as part of a polymer chain. This further means that a locus' motion will be influenced by and correlated with the motion of surrounding loci in the chain. In particular, what this means for a locus near a CTCF site, is that it will still experience the constraints of the loop due to correlations along the chain. According to polymer simulations, the influence of a loop base will even be felt by loci located quite far away from the CTCF site (up to 100 kb linear genomic separation) and differences in motion will be small (Pavel Kos, personal communication, unpublished data). It was also suggested by a recent study performing live-cell imaging at the *Drosophila even-skipped (eve)* locus that found that the relaxation time of DNA only shows a shallow dependence on genomic separation, thereby arguing for correlations over long distances (Brückner *et al.*, 2023). Given our experimental set-up, where our operator arrays are on average ~ 10 kb from the next endogenous CTCF site, this would mean that most of these loci experience a motion that is influenced by the next, endogenous CTCF binding sites. Therefore, it is likely that due to the high abundance of CTCF binding sites in the genome, almost all chromosome dynamics that we observed here will be a proxy for a locus close to a loop boundary. Furthermore, while the effects can be observed in polymer simulations, it is possible that the noise in experiments masks these small differences.

While this might explain the lack of any observable change when excising CTCF sites close to the operator array, it does not explain that dynamics do not change upon global depletion of CTCF. In this case, cohesin will still be loaded, however the loops will not be anchored at CTCF sites anymore. The fact that we do not see changes in the motion of our operator arrays, leads to two conclusions: First, the same amount of cohesins must be present on DNA, meaning that the residence time of cohesin is not altered by CTCF depletion. This is in line with previous measurements of cohesin dynamics on DNA upon depletion of CTCF (Nora *et al.*, 2020). Second, the overall density of the loops must also be the same, although with a different positioning which is in agreement with findings from 3C methods (Nora *et al.*, 2017).

4.3 CTCF and cohesin control looping dynamics

Since the effect of CTCF on chromosome dynamics was not observable in our experiments measuring global chromosome dynamics, we asked whether the effect can be measured when imaging two specific loci at a loop base with convergent CTCF sites. This would potentially increase the chances

4. DISCUSSION

that the effect can be observed, as indicated by polymer simulations (see 2.3.2). Also, it could answer the question of how slowing down of chromosome motion by cohesin on a global scale affects the interactions of two loci in *cis*. Understanding the dynamics of encounters of two loci in relation to each other can potentially also inform on interaction frequencies of E-P pairs.

We therefore used an ideal model set-up to investigate the effect of one specific, very strong CTCF loop: in an otherwise “neutral” TAD without any internal structure and without expressed genes or enhancers, no heterochromatic marks or association with lamina-associated domains (LADs) (Zuin *et al.*, 2022), we inserted a very strong, ectopic CTCF loop flanked by two orthogonal operator arrays. Using this bottom-up design of a regulatory landscape, we were able to image the dynamics of this specific loop without other confounding effects that could arise from adjacent, endogenous regulatory sequences. We chose 150 kb as a size for the loop, since this corresponds to the average cohesin- and CTCF loop size in mESCs as well as being approximately the average distance between E-P pairs (see Methods) (Jung *et al.*, 2019). Choosing three very strong CTCF binding motifs placed directly adjacent to each other, most likely constitutes an impermeable barrier. This further facilitated comparison of experimental data to polymer simulations with impermeable BEs.

When measuring the dynamics of the loci in *cis*, we first noticed that the MSD of their pairwise distance shows a much smaller anomalous exponent (radial MSD ~ 0.2) as observed for global chromosome dynamics (see 2.3.3). This can be explained by the linear separation of the loci, which defines their maximal distance to each other as the length of the stretched out 150 kb of chromatin. This leads to imposed constraints that become visible as a plateau in the radial MSD which is detected as an extremely low MSD power-law in the presence of experimental uncertainty on distances. Indeed, the finding that linear separation correlates with the anomalous exponent when observing dynamics at the minute scale, could explain the differences in anomalous exponents observed in other studies (Khanna *et al.*, 2019; Gabriele *et al.*, 2022).

Similar to studies using DNA-FISH and in line with the measurement of global chromosome dynamics (Szabo *et al.*, 2020), we found that the presence of cohesin leads overall to smaller distances between the two loci. Interestingly, the same holds true for CTCF sites adjacent to the operator arrays, leading to the conclusion that both cohesin and CTCF help to facilitate interactions between loci separated by a distance of 150 kb. Furthermore, we show that the variability in distances is decreased when CTCF and cohesin are present, arguing for facilitated, but also more robust interactions.

However, the advantage of live-cell imaging lies in the fact that we gain access to the temporal development of spatial distances between the two loci. We could investigate whether overall smaller distances and a reduced variability are due to a change in the duration or frequency of interactions. By inferring a proximal state with an average distance of 150 nm, we saw that cohesin and CTCF sites adjacent to the operator arrays increase the duration of this proximal state by 2-3-fold. Furthermore, the interactions are also more frequent when the operator arrays are flanked by CTCF sites and cohesin is present. With the help of polymer simulations, we could further show that this finding can be explained by the loop extrusion model, in particular with cohesins stalling at CTCF sites and

enforcing a loop. Given these measurements of proximity, the interactions between two loci in *cis* flanked by CTCF sites occur on average every 5 min and last approximately 16 min. For two loci without CTCF sites this is reduced to a duration of 6 min on average. However, considering the length of the cell cycle of mESCs of 11-14 h (Roccio *et al.*, 2013), this would mean that interactions between e. g. an E-P pair at that distance occur frequently and are quite short-lived. I will discuss the implications of this finding on transcriptional regulation in 4.6.

4.4 What is a contact?

With our experiments, we have access to the physical proximity between the loci, however the measured distances are confounded by the uncertainty limit with which we observe them. We calculated this uncertainty that mostly arises from the localization precision in diffraction-limited microscopy to be approximately 130 nm. This is slightly lower than the average distance of 150 nm in our proximal state. Of course, this raises the question of how these proximity events relate to actual molecular contacts between e. g. two CTCF sites as a loop base or a potential proximity event between an E-P pair. As summarized in 1.1.5, it is currently unclear what are the distances that are necessary for e. g. exchanging information between an enhancer and a promoter in order to achieve transcriptional activation. While a mechanism via direct protein-protein interactions would argue for very small distances in the range of up to 10 nm, models that propose communication through an enrichment of TFs and cofactors could potentially signal over larger distances up to possibly tens of nm (Hnisz *et al.*, 2017; Chong *et al.*, 2018; Field & Adelman, 2020; Karr *et al.*, 2022).

In our experiments, we do not have access to the underlying molecular state of the chromatin fiber and we must therefore conclude that our measurements likely provide an over-estimation as compared to actual functional contacts of the two loci. However, I would like to point out that our proximal state is comparable to the contact probabilities in 3C methods. Here, the cross-linking and ligation radius defines the maximal distance between two loci that is counted as a contact. This parameter was estimated to be in the range of 150-250 nm (McCord *et al.*, 2020; L. F. Chen, Lee, *et al.*, 2023) and is therefore very comparable to the average distance of 150 nm of the proximal state in our experiments. We therefore also sought to compare changes in the contact probabilities in the presence or absence of flanking CTCF sites in our Capture-C experiments to changes in the fraction of time spent in the proximal state in the live-cell imaging experiments. We found a good agreement in the fold change observed when deleting CTCF sites (2.8-fold in Capture-C, 2.3-fold in imaging experiments, see 2.3.4).

It should be noted that in theory a better resolution and thereby a smaller uncertainty on the 3D distances could be achieved with super-resolution microscopy approaches, as has been demonstrated in DNA-FISH experiments. However, so far these approaches have been limited to application to fixed cells mainly due to technical limitations, e. g. the need for stochastic activation of the fluorophores which is incompatible with tracking over time (Boka *et al.*, 2021). The rapid advance of super-resolution microscopy techniques that are compatible with live-cell approaches (Cao *et al.*, 2021;

4. DISCUSSION

Deguchi *et al.*, 2023) will likely provide measurements of E-P proximity with increased accuracy together with single-molecule resolved binding/unbinding of TFs and cofactors (Donovan *et al.*, 2019; Ferrie *et al.*, 2023), thus allowing for a more precise assessment of molecular contact dynamics.

4.5 Inference of the cohesin-mediated looped state

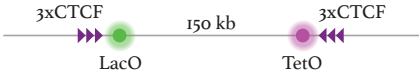
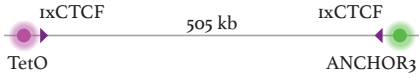
While we cannot make conclusions about molecular contacts from these experiments, polymer simulations of loop extrusion of a 150 kb-sized loop with impermeable BEs allowed us to infer the underlying cohesin-mediated looped state. We compared the experimental data to data from simulations with added noise mimicking the experimental uncertainty. Since we know the actual molecular configuration of the simulated DNA polymer, we found that for the best agreement with the experimental data, a loop lasts 5-15 min on average in the case when impermeable BEs are present, and less than 1 min without BEs. This is also in good agreement with measurements of cohesin residence times on DNA by FRAP (Hansen *et al.*, 2017; Wutz *et al.*, 2020). However, this would indicate that cohesin spends a considerable time on DNA in a loop-enforcing, stalled state as compared to an actively extruding state, given a speed of loop extrusion of 0.1-1 kb/s (Davidson *et al.*, 2019).

These results are in good agreement with a recent study by the lab of Anders S. Hansen that measured the contact dynamics near two CTCF sites that demarcate the boundaries of a 505 kb TAD at the *Fbn2* locus (Gabriele *et al.*, 2022). They found that their cohesin-mediated looped state has a comparable duration of 10-30 min, however occurs less frequently. The fraction of time the two loci spent in a looped-state was here estimated to be 3-6% as compared to 20-30% estimated for our experimental set-up based on polymer simulations.

Due to the very comparable set-up of the two experimental studies, I will now discuss the small differences in the approaches and what can potentially be learned from them (see also Table 4.1). The first difference lies in the genomic separation between the two loci, which in their case is 505 kb. This is a possible explanation for the looped state occurring only 3-6% of the time measured. This would indicate that the size of a loop mostly defines how often a loop is formed, while the duration of a looped state is unchanged by its size. This could be due to the extended time that it would take cohesin to extrude a loop of 505 kb as compared to a smaller 150 kb loop. Once stalled at CTCF sites, only the residence time of either CTCF or cohesin defines the lifetime of a loop.

The second difference is the CTCF sites used to enforce a loop between the two loci: we used three very strong CTCF motifs that had previously been described to be bound in almost 100% of alleles measured by single-molecule footprinting (R. Grand and D. Schübeler, personal communication). In the case of the study of Gabriele *et al.*, 2022, there were only one CTCF site each that were endogenously occurring at the boundary of the *Fbn2* locus. Binding strength of CTCF to specific motifs might therefore also be a determinant of looping dynamics. In this scenario, the decrease in frequency would arise from cohesins not always being stalled at the CTCF sites and continuing

Table 4.1: Comparison of studies on CTCF/cohesin-mediated looping dynamics in mESCs. The results from section 2.3.4 are compared to Gabriele *et al.*, 2022. Adapted from Giorgetti *et al.*, 2022.

| | Mach, Kos, Zhan <i>et al.</i> (this study) | Gabriele, Brandão, Grosse-Holz <i>et al.</i> |
|--|---|--|
| | in mESCs in the neutral TAD from Zuin <i>et al.</i> , 2022 | in mESCs at the <i>Fbn2</i> locus |
| |  |  |
| Physical proximity (~ 150 nm distance): | ~ 80% | n. a. |
| CTCF/cohesin-mediated looped state: | | |
| Time in looped state | ~20-30% | ~ 3-6% |
| Loop duration | 5-15 min | 10-30 min |

to extrude past them, since these sites are permeable barriers that can be overcome with a certain probability.

Besides these differences in the experimental set-up, the studies also differ in the way the cohesin-mediated looped state was inferred: We made use of a HMM to identify a proximal state from our experimental data, which we then compared to polymer simulations to get access to the underlying cohesin-mediated looped state. The study from the Hansen lab on the other hand used a newly developed Bayesian classifier (Bayesian inference of looping dynamics (BILD)). Both techniques are a first attempt at interpreting these challenging data sets generated from live-cell imaging with a high uncertainty on the actual 3D distances. Therefore, both techniques currently still suffer from drawbacks: The HMM approach assumes that there are no temporal correlations in the transitions between proximal and distal states and that transitions are instantaneous. This is likely to lead to an overestimation of the fraction of the time spent in the looped state, as e.g. relaxation times after loop release are not taken into account. Furthermore, computationally expensive polymer simulations of the exact measured scenario need to be performed in order to be able to compare the simulation to the experimental data. This makes the approach more difficult to apply. On the other hand, Bayesian inference of looping dynamics (BILD) can account for temporal correlations, because it was constrained by the actually measured MSDs in the CTCF-depleted scenario. However, the precision and recall of the classifier was tuned by training on simulated data. Here, the matching between experimental data and simulations is crucial and e.g. it is unclear how the time scales of units of simulated time were matched to the real time in experiments. For an independent review of these differences, I refer the reader to a recent review by L. F. Chen, Lee, *et al.*, 2023.

All in all, given the considerable differences in study design and analysis, the results from both studies together give a first impression of how chromosome dynamics of two loci in *cis* can be regulated by

4. DISCUSSION

cohesin-mediated loop extrusion, how this is dependent on the genomic separation and the CTCF site strength. Future studies should systematically test the relationship between chromosome structure as seen in Hi-C and dynamics measured in live-cell imaging, taking into account a large variety of regulatory landscapes.

4.6 Looping dynamics and regulation of transcription

Interestingly, the timescales of CTCF looping and physical proximity described by the two studies represent an intermediate level between those of nucleosome remodeling and TF binding/unbinding and promoter bursting: While TF binding typically occurs on the seconds timescale, promoter bursting in mammalian cells takes place over tens of minutes (Fig. 4.1a) (Coulon *et al.*, 2013). It is unclear how these processes are integrated in time to yield a robust transcriptional response with the right cell-to-cell, but also temporal variability. It should be noted that studies in budding yeast have shown more compatible time scales for TF binding and onset of transcription (Donovan *et al.*, 2019).

Given the fast binding/unbinding kinetics of many TFs and the potentially longer timescales of enhancer-promoter proximity, it is possible that time-averaged TFBS occupancies within enhancer and promoter regions are sufficient to enable the exchange of regulatory information throughout the duration of a contact. But even if this is the case, does a promoter get activated every time that an enhancer is close enough that their transiently bound TFs can interact (Fig. 4.1b, i)? Or does this happen with a delay, possibly due to the kinetics of intermediate regulatory steps (e.g. assembly or interactions with cofactors) (Fig. 4.1b, ii)? Or rather does a burst of transcription only commence after several contacts, e.g. because of cooperativity in one or more sequential events occurring at the promoter (Fig. 4.1b, iii)? Is every E-P pair actually expected to behave the same, or do alternative modes of signal integration (Lammers *et al.*, 2020) result in different behaviors at different E-P pairs? And how is this related to bursting kinetics at the promoter?

There is evidence for a 1-contact-1-burst scenario in *Drosophila* embryos, although the study only investigated ectopically induced, unusually stable E-P interactions (H. Chen *et al.*, 2018). Another study in mammalian cells found no temporal correlation between contacts and transcription bursts. However, in this study the operator array indicating the position of the enhancer was placed outside of the boundary of the TAD that harbors the E-P pair, and this might have masked potential contacts occurring within the TAD (Alexander *et al.*, 2019).

Theoretical arguments that are compatible with scenarios ii) and iii) have been recently evoked to explain striking non-linearities in how transcription levels depend on E-P contact probabilities (Xiao *et al.*, 2021; Zuin *et al.*, 2022), although none has been formally proven. First insights come from a recent study at the *Drosophila eve* locus where the enhancer was placed at varying distances from the promoter and both, E-P distances as well as nascent transcription, were observed in live-cell imaging: This indeed showed a dependence of transcriptional activation on the genomic separation which

4.7. LONG-TERM SINGLE-PARTICLE TRACKING OF COHESIN ON DNA

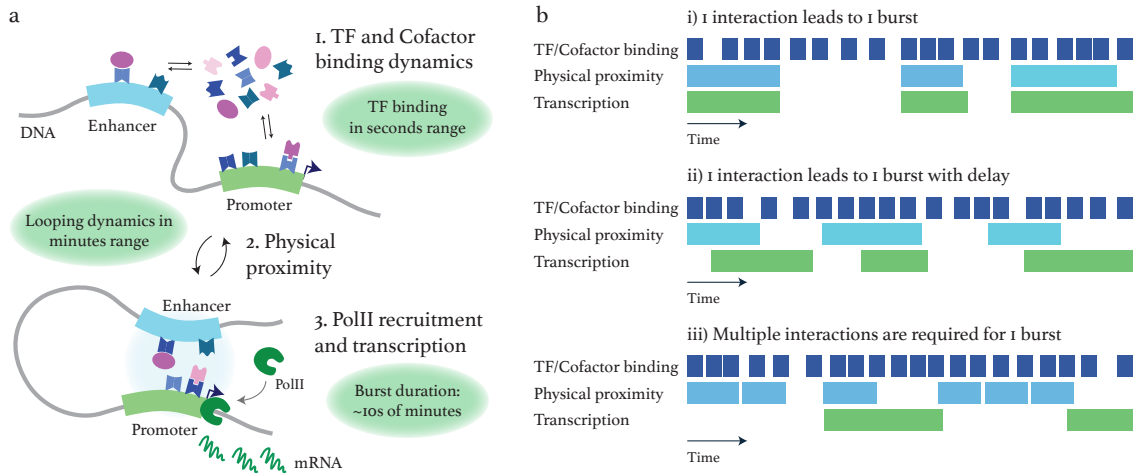


Figure 4.1: Temporal dynamics of E-P communication. **a**, Transcription initiation and the resulting transcriptional bursts are preceded by multiple processes, notably fast transcription factor and cofactor binding (typically in the second to tens of seconds range) and looping of the chromatin fiber that establishes physical proximity, estimated to occur in the few minutes range. **b**, Whilst transcription factor binding dynamics are fast and occur frequently, possibly enabling continuous occupancy of TFBS at the enhancer and the promoter, E-P proximity is presumably longer and occurs less frequent, leading to potential models of E-P communication in time: i) one interaction might lead to one event of transcription bursting, ii) possibly with a delay iii) where one bursting event can only be achieved by multiple consecutive interaction events (TF=transcription factor, PolII=RNA Polymerase II). *Fig. from Mach and Giorgetti, 2023.*

indicates that proximity might be a rate-limiting step for activation at least in *Drosophila* (Brückner *et al.*, 2023). Together with other recent measurements (Lingling *et al.*, 2023; Platania *et al.*, 2023), these studies provide the first examples of live-cell imaging experiments that are able to probe the dynamics of chromosome structure as well as transcription. Future studies at different loci, in different species or cell types, will increase our knowledge of factors that contribute to dynamics of E-P interactions.

4.7 Long-term single-particle tracking of cohesin on DNA

As described in 4.1, cohesin slowing down chromosome dynamics is compatible with the formation of bonds via loop extrusion between two strands in *cis*, but also via sister-chromatid cohesion between two strands in *trans*. It is therefore in question, whether the constrained behavior is really a consequence of loop extrusion. I therefore set out to discriminate the dynamics of cohesin's functions on DNA in living cells.

While it would generally be a possibility to answer this question by measuring chromosome dynamics during different cell cycle stages (e. g. using a fluorescent reporter for the cell cycle like the fluorescence ubiquitin cell cycle indicator (FUCCI) system (Sakaue-Sawano *et al.*, 2008)), it is technically very challenging to combine this technique with live-cell imaging. This drawback arises from the limited number of fluorophores that can simultaneously be visualized in fluorescence microscopy. It makes it difficult to image DNA loci that have very dim signals and multiple fluorescent reporters for the cell

4. DISCUSSION

cycle in the same experiment. Furthermore, this would also significantly increase the exposure to light, which should be limited to prevent photo-bleaching over time.

Therefore, I chose SPT with sparse labeling as an approach to follow RAD21 molecules and be able to characterize the motion of cohesin, potentially whilst performing either of its functions. For this, a long observation period is required, as cohesin is often associated with DNA for 8-20 min (Hansen *et al.*, 2017; Wutz *et al.*, 2020). Furthermore, the lifetime of average-sized cohesin-mediated loops was estimated to last for 5-15 min (see 2.3.4). Therefore, we followed the loci over 30 min with this new approach, possibly covering the entire lifetime of a cohesin binding event to DNA. While SPT has been previously used to study transcription factor and cofactor binding dynamics to DNA in the seconds to 1-2 min range (Donovan *et al.*, 2019; Ferrie *et al.*, 2023), its application to study cohesin dynamics on long time scales in the minutes range had been impossible until now. This is mostly due to limitations arising from photo-bleaching and photo-toxicity in SPT experiments that prevent extended imaging periods (Boka *et al.*, 2021). Here, I made use of a new image restoration technique based on deep learning (N2V2) that can significantly reduce the power needed to visualize individual molecules and thereby increases the time the molecules can be followed (Höck *et al.*, 2022). The advantage of this approach is that it leads to a quantitative description of both cohesin and DNA motion in the same cell, which can then be compared.

With this, I could show that cohesin shows a more sub-diffusive motion on DNA than the underlying DNA fiber itself. Given that the motion of RAD21 molecules that we observe is a convolution of the motion of cohesin on DNA and DNA motion, it is likely that we most often follow cohesin molecules that perform a slow, potentially bond-enforcing role on DNA. As is indicated by polymer simulations (Pavel Kos, personal communication), a loop anchor would show a more constrained behavior compared to a cohesin molecule just bound to DNA without performing any function. In the future, it remains to be seen, if this fraction of cohesin molecules is lost, when cohesin is not able to perform either loop extrusion or cohesion (as is the case when either NIPBL or Sororin is depleted). Currently, our labeling strategy prevented measurements under these conditions, as also cohesin abundance on DNA is changed, but these limitations can likely be overcome in adjusted experiments.

However, when we performed these depletion experiments and followed multiple random DNA loci on DNA, we could already conclude that both of cohesin's functions contribute to slowing down chromosome dynamics, as we saw an increase in the instantaneous diffusion coefficient in both cases, when depleting NIPBL or when depleting Sororin. This is a strong indicator that indeed bond formation is the driver of this observation, considering that this is part of the process of loop extrusion as well as of cohesion.

It is currently in question, whether it is possible to observe a signature of cohesin actively performing loop extrusion in these settings. This is due to the low velocity of cohesin performing loop extrusion *in vitro* (Davidson *et al.*, 2019) and the motion with which DNA moves itself (Nozaki *et al.*, 2017; Gu *et al.*, 2018; Gabriele *et al.*, 2022; Mach *et al.*, 2022). It is likely that both motions are very similar, making it hard to distinguish them in settings where only a convolution of cohesin and DNA motion

can be observed. In order to overcome this limitation and potentially observe actively extruding cohesin molecules, future experiments could be performed in a cell line carrying a degron system for PDS5 and RAD21, where it is potentially possible to increase the velocity and processivity of cohesin on DNA thereby making loop extrusion observable (Elphège Nora, personal communication, van Ruiten *et al.*, 2022).

In general, the approach we developed here, combining measurements of dynamics of DNA loci with measurements of the dynamics of a protein bound to DNA over long periods of time, opens up possibilities also to study other interesting questions. The dynamics of long-range transcriptional activation by enhancers and the factors involved on a mechanistic level are still very much unknown. With this technique, it is potentially possible to combine the tracking of an E-P pair with observing binding interactions of TFs or cofactors. Together with current advances in microscopy development to achieve super-resolution in live cells, such as with MINFLUX (Deguchi *et al.*, 2023), it will become increasingly more feasible to study the mechanisms of transcriptional activation by enhancers on temporal and spatial scales that are relevant to the molecular processes that are thought to mediate it (see 1.1.5 and 4.4).

4.8 Conclusion

This study provides a description of chromosome dynamics and how these are influenced by the process of loop extrusion. It shows that cohesin and CTCF regulate chromosome dynamics by constraining their motion and thereby facilitating interactions in *cis*. The duration and frequency of cohesin-mediated loops described in chapter 2 is in excellent agreement also with a recent study observing looping interactions at TAD boundaries (Gabriele *et al.*, 2022). Together, these two studies have pushed forward our understanding of the modulation of dynamics of chromosome structure by cohesin and have enabled the first estimates of duration and frequency of interactions of loci in *cis* from live-cell imaging studies.

The strength of the approaches demonstrated here, e. g. namely the bottom-up engineering of a neutral environment for live-cell imaging studies as well as simultaneous, long-term imaging of DNA loci and single proteins bound to DNA, are powerful tools to unequivocally study the effects of a specific sets of questions on chromosome dynamics. New genome engineering approaches (Zuin *et al.*, 2014; Rinzema *et al.*, 2022; Lingling *et al.*, 2023) are likely to bring an increasing complexity with which regulatory landscapes can be set-up or modified. It will be interesting to see, if technical challenges in the set-up of live-cell imaging studies can be overcome to a account for probing landscapes in a high-throughput manner.

Finally, combining live-cell imaging approaches of DNA bound proteins and DNA loci with *in vivo* proteomics approaches (Graham *et al.*, 2022; DelRosso *et al.*, 2023) will be crucial to unravel the dynamics of molecular mechanisms that modulate transcriptional activation.

4. DISCUSSION

References

- Abbé, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1), 413–468.
- Adra, C. N., Boer, P. H., & McBurney, M. W. (1987). Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. *Gene*, 60(1), 65–74. [https://doi.org/10.1016/0378-1119\(87\)90214-9](https://doi.org/10.1016/0378-1119(87)90214-9)
- Aibara, S., Schilbach, S., & Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature*, 594(7861), 124–128. <https://doi.org/10.1038/s41586-021-03554-8>
- Alerasool, N., Leng, H., Lin, Z. Y., Gingras, A. C., & Taipale, M. (2022). Identification and functional characterization of transcriptional activators in human cells. *Molecular Cell*, 82(3), 677–695.e7. <https://doi.org/10.1016/j.molcel.2021.12.008>
- Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., & Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife*, 8. <https://doi.org/10.7554/eLife.41769>
- Alipour, E., & Marko, J. F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Research*, 40(22), 11202–12. <https://doi.org/10.1093/nar/gks925>
- Aljahani, A., Hua, P., Karpinska, M. A., Quililan, K., Davies, J. O. J., & Oudelaar, A. M. (2022). Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nature Communications*, 13(1), 2139. <https://doi.org/10.1038/s41467-022-29696-5>
- Allan, D. B., Caswell, T., Keim, N. C., van der Wel, C. M., & Verweij, R. W. (2021). Soft-matter/trackpy: Trackpy v0. 5.0. *Genève: Zenodo*.
- Amano, T., Sagai, T., Tanabe, H., Mizushima, Y., Nakazawa, H., & Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: Limb bud-specific differential regulation of competence and active transcription. *Developmental Cell*, 16(1), 47–57. <https://doi.org/10.1016/j.devcel.2008.11.011>
- Anderson, D. E., Losada, A., Erickson, H. P., & Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. *Journal of Cell Biology*, 156(3), 419–24. <https://doi.org/10.1083/jcb.200111002>
- Anderson, E., Devenney, P. S., Hill, R. E., & Lettice, L. A. (2014). Mapping the *Shh* long-range regulatory domain. *Development*, 141(20), 3934–43. <https://doi.org/10.1242/dev.108480>
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raitchel, J., Lilje, B., Rapin, N., ... Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493), 455–461. <https://doi.org/10.1038/nature12787>
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryń Ł, M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339(6123), 1074–7. <https://doi.org/10.1126/>

REFERENCES

- science.1232542
- Arnould, C., Rocher, V., Finoux, A. L., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, P. E., Ricci, E. P., Mourad, R., Haber, J. E., Noordermeer, D., & Legube, G. (2021). Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature*, *590*(7847), 660–665. <https://doi.org/10.1038/s41586-021-03193-z>
- Arrastia, M. V., Jachowicz, J. W., Ollikainen, N., Curtis, M. S., Lai, C., Quinodoz, S. A., Selck, D. A., Ismagilov, R. F., & Guttman, M. (2022). Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nature Biotechnology*, *40*(1), 64–73. <https://doi.org/10.1038/s41587-021-00998-1>
- Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., & Zeitlinger, J. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nature Genetics*, *53*(3), 354–366. <https://doi.org/10.1038/s41588-021-00782-6>
- Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E., & Webb, W. W. (1976). Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical Journal*, *16*(9), 1055–69. [https://doi.org/10.1016/s0006-3495\(76\)85755-4](https://doi.org/10.1016/s0006-3495(76)85755-4)
- Ba, Z., Lou, J., Ye, A. Y., Dai, H. Q., Dring, E. W., Lin, S. G., Jain, S., Kyritsis, N., Kieffer-Kwon, K. R., Casellas, R., & Alt, F. W. (2020). CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning. *Nature*, *586*(7828), 305–310. <https://doi.org/10.1038/s41586-020-2578-0>
- Balzarotti, F., Eilers, Y., Gwosch, K. C., Gynnå, A. H., Westphal, V., Stefani, F. D., Elf, J., & Hell, S. W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science*, *355*(6325), 606–612. <https://doi.org/10.1126/science.aak9913>
- Banerji, J., Olson, L., & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, *33*(3), 729–40. [https://doi.org/10.1016/0092-8674\(83\)90015-6](https://doi.org/10.1016/0092-8674(83)90015-6)
- Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 Pt 1), 299–308. [https://doi.org/10.1016/0092-8674\(81\)90413-x](https://doi.org/10.1016/0092-8674(81)90413-x)
- Banigan, E. J., Tang, W., van den Berg, A. A., Stocsits, R. R., Wutz, G., Brandão, H. B., Busslinger, G. A., Peters, J. M., & Mirny, L. A. (2023). Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(11), e2210480120. <https://doi.org/10.1073/pnas.2210480120>
- Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F., & Mirny, L. A. (2020). Chromosome organization by one-sided and two-sided loop extrusion. *Elife*, *9*. <https://doi.org/10.7554/eLife.53558>
- Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L. M., Dostie, J., Pombo, A., & Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16173–8. <https://doi.org/10.1073/pnas.1204799109>
- Barisic, D., Stadler, M. B., Iurlaro, M., & Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature*, *569*(7754), 136–140. <https://doi.org/10.1038/s41586-019-1115-5>
- Battaglia, S., Dong, K., Wu, J., Chen, Z., Najm, F. J., Zhang, Y., Moore, M. M., Hecht, V., Shoresh, N., & Bernstein, B. E. (2022). Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nature Genetics*, *54*(10), 1504–1513. <https://doi.org/10.1038/s41588-022-01188-8>
- Batut, P. J., Bing, X. Y., Sisco, Z., Raimundo, J., Levo, M., & Levine, M. S. (2022). Genome organization controls transcriptional dynamics during development. *Science*, *375*(6580), 566–570. <https://doi.org/10.1126/science.abi7178>
- Bauer, B. W., Davidson, I. F., Canena, D., Wutz, G., Tang, W., Litos, G., Horn, S., Hinterdorfer, P., & Peters, J. M. (2021). Cohesin mediates DNA loop extrusion by a “swing and clamp” mechanism. *Cell*, *184*(21), 5448–5464.e22. <https://doi.org/10.1016/j.cell.2021.09.016>

- Beagrie, R., Scialdone, A., Schueler, M., Kraemer, D., Chotalia, M., Xie, S., Barbieri, M., de Santiago, I., Game, L., Dillon, N., Edwards, P., Nicodemi, M., & Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, *543*(7646), 519–524. <https://doi.org/10.1038/nature21411>
- Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., Chang, Y., Li, J. B., Senaratne, T. N., Williams, B. R., Rouillard, J. M., & Wu, C. T. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(52), 21301–6. <https://doi.org/10.1073/pnas.1213818110>
- Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*, *405*(6785), 482–5. <https://doi.org/10.1038/35013100>
- Bergman, D. T., Jones, T. R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H. Y., Nasser, J., Kane, M., Rios, A., Nguyen, T. H., Grossman, S. R., Fulco, C. P., Lander, E. S., & Engreitz, J. M. (2022). Compatibility rules of human enhancer and promoter sequences. *Nature*, *607*(7917), 176–184. <https://doi.org/10.1038/s41586-022-04877-w>
- Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., & Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, *313*(5793), 1642–5. <https://doi.org/10.1126/science.1127344>
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P., & Canaider, S. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology*, *40*(6), 463–71. <https://doi.org/10.3109/03014460.2013.807878>
- Bickmore, W. A. (2013). The spatial organization of the human genome. *Annual Review of Genomics and Human Genetics*, *14*, 67–84. <https://doi.org/10.1146/annurev-genom-091212-153515>
- Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N., & Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, *362*(6413). <https://doi.org/10.1126/science.aau1783>
- Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–9. [https://doi.org/10.1016/0092-8674\(85\)90312-5](https://doi.org/10.1016/0092-8674(85)90312-5)
- Birkenbihl, R. P., & Subramani, S. (1992). Cloning and characterization of *rad21* an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. *Nucleic Acids Research*, *20*(24), 6605–11. <https://doi.org/10.1093/nar/20.24.6605>
- Blayney, J., Francis, H., Camellato, B., Mitchell, L., Stolper, R., Boeke, J., Higgs, D., & Kassouf, M. (2022). Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression. *bioRxiv*, 2022.06.20.496856. <https://doi.org/10.1101/2022.06.20.496856>
- Bobroff, N. (1986). Position measurement with a resolution and noise-limited instrument. *Review of Scientific Instruments*, *57*(6), 1152–1157.
- Boka, A. P., Mukherjee, A., & Mir, M. (2021). Single-molecule tracking technologies for quantifying the dynamics of gene regulation in cells, tissue and embryos. *Development*, *148*(18). <https://doi.org/10.1242/dev.199744>
- Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M. R., & Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLOS Biology*, *3*(5), e157. <https://doi.org/10.1371/journal.pbio.0030157>
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., & Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, *171*(3), 557–572.e24. <https://doi.org/10.1016/j.cell.2017.09.043>
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008).

REFERENCES

- High-resolution mapping and characterization of open chromatin across the genome. *Cell*, *132*(2), 311–22. <https://doi.org/10.1016/j.cell.2007.12.014>
- Branco, M. R., & Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLOS Biology*, *4*(5), e138. <https://doi.org/10.1371/journal.pbio.0040138>
- Brandão, H. B., Gabriele, M., & Hansen, A. S. (2021). Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Current Opinion in Cell Biology*, *70*, 18–26. <https://doi.org/10.1016/j.ceb.2020.11.002>
- Brandão, H. B., Ren, Z., Karaboja, X., Mirny, L. A., & Wang, X. (2021). DNA-loop-extruding SMC complexes can traverse one another *in vivo*. *Nature Structural & Molecular Biology*, *28*(8), 642–651. <https://doi.org/10.1038/s41594-021-00626-1>
- Breschi, A., Gingeras, T. R., & Guigó, R. (2017). Comparative transcriptomics in human and mouse. *Nature Reviews Genetics*, *18*(7), 425–440. <https://doi.org/10.1038/nrg.2017.19>
- Bronstein, I., Israel, Y., Kepten, E., Mai, S., Shav-Tal, Y., Barkai, E., & Garini, Y. (2009). Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Physical Review Letters*, *103*(1), 018102. <https://doi.org/10.1103/PhysRevLett.103.018102>
- Brosh, R., Coelho, C., Ribeiro-Dos-Santos, A. M., Ellis, G., Hogan, M. S., Ashe, H. J., Somogyi, N., Ordoñez, R., Luther, R. D., Huang, E., Boeke, J. D., & Maurano, M. T. (2023). Synthetic regulatory genomics uncovers enhancer context dependence at the *Sox2* locus. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2023.02.027>
- Brückner, D. B., Chen, H., Barinov, L., Zoller, B., & Gregor, T. (2023). Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, *380*(6652), 1357–1362. <https://doi.org/10.1126/science.adf5568>
- Bulger, M., & Groudine, M. (1999). Looping versus linking: Toward a model for long-distance gene activation. *Genes & Development*, *13*(19), 2465–77. <https://doi.org/10.1101/gad.13.19.2465>
- Bürmann, F., Lee, B. G., Than, T., Sinn, L., O'Reilly, F. J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K., & Löwe, J. (2019). A folded conformation of mukBEF and cohesin. *Nature Structural & Molecular Biology*, *26*(3), 227–236. <https://doi.org/10.1038/s41594-019-0196-z>
- Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J. M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, *544*(7651), 503–507. <https://doi.org/10.1038/nature22063>
- Bystricky, K. (2015). Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Letters*, *589*(20 Pt A), 3014–22. <https://doi.org/10.1016/j.febslet.2015.07.012>
- C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, *282*(5396), 2012–8. <https://doi.org/10.1126/science.282.5396.2012>
- Cadiñanos, J., & Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Research*, *35*(12), e87. <https://doi.org/10.1093/nar/gkm446>
- Calderon, L., Weiss, F. D., Beagan, J. A., Oliveira, M. S., Georgieva, R., Wang, Y. F., Carroll, T. S., Dharmalingam, G., Gong, W., Tossell, K., de Paola, V., Whilding, C., Ungless, M. A., Fisher, A. G., Phillips-Cremins, J. E., & Merkenschlager, M. (2022). Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *Elife*, *11*. <https://doi.org/10.7554/eLife.76539>
- Çamdere, G., Guacci, V., Stricklin, J., & Koshland, D. (2015). The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. *Elife*, *4*. <https://doi.org/10.7554/eLife.11315>
- Cameron, G., Gruszka, D., Xie, S., Nasmyth, K. A., Srinivasan, M., & Yardimci, H. (2022). Sister chromatid cohesion establishment during DNA replication termination. *bioRxiv*, 2022.09.15.508094. <https://doi.org/10.1101/2022.09.15.508094>

- Cao, B., Coelho, S., Li, J., Wang, G., & Pertsinidis, A. (2021). Volumetric interferometric lattice light-sheet imaging. *Nature Biotechnology*, *39*(11), 1385–1393. <https://doi.org/10.1038/s41587-021-01042-y>
- Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J. B., Espinola, S. M., Gurgo, J., Messina, O., Houbbron, C., Ogiyama, Y., Papadopoulos, G. L., Cavalli, G., Lagha, M., & Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell*, *74*(1), 212–222.e5. <https://doi.org/10.1016/j.molcel.2019.01.011>
- Cattoglio, C., Pustova, I., Walther, N., Ho, J. J., Hantsche-Grininger, M., Inouye, C. J., Hossain, M. J., Dailey, G. M., Ellenberg, J., Darzacq, X., Tjian, R., & Hansen, A. S. (2019). Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *Elife*, *8*. <https://doi.org/10.7554/eLife.40164>
- Cattoni, D. I., Cardozo Gizzi, A. M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbbron, C., Déjardin, S., Fiche, J. B., González, I., Chang, J. M., Sexton, T., Marti-Renom, M. A., Bantignies, F., Cavalli, G., & Nollmann, M. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nature Communications*, *8*(1), 1753. <https://doi.org/10.1038/s41467-017-01962-x>
- Chakraborty, S., Kopitchinski, N., Zuo, Z., Eraso, A., Awasthi, P., Chari, R., Mitra, A., Tobias, I. C., Moorthy, S. D., Dale, R. K., Mitchell, J. A., Petros, T. J., & Rocha, P. P. (2023). Enhancer-promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nature Genetics*, *55*(2), 280–290. <https://doi.org/10.1038/s41588-022-01295-6>
- Chambeyron, S., & Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the *HoxB* locus upon induction of transcription. *Genes & Development*, *18*(10), 1119–30. <https://doi.org/10.1101/gad.292104>
- Chan, K. L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K., & Beckouët, F. (2013). Pds5 promotes and protects cohesin acetylation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 13020–5. <https://doi.org/10.1073/pnas.1306900110>
- Chang, L. H., Ghosh, S., & Noordermeer, D. (2020). TADs and Their Borders: Free Movement or Building a Wall? *Journal of Molecular Biology*, *432*(3), 643–652. <https://doi.org/10.1016/j.jmb.2019.11.025>
- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., & Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479–91. <https://doi.org/10.1016/j.cell.2013.12.001>
- Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. *Nature Genetics*, *50*(9), 1296–1303. <https://doi.org/10.1038/s41588-018-0175-z>
- Chen, L. F., Lee, J., & Boettiger, A. (2023). Recent progress and challenges in single-cell imaging of enhancer-promoter interaction. *Current Opinion in Genetics & Development*, *79*, 102023. <https://doi.org/10.1016/j.gde.2023.102023>
- Chen, L. F., Long, H. K., Park, M., Swigut, T., Boettiger, A. N., & Wysocka, J. (2023). Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2023.03.009>
- Cheng, A. W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A., & Wang, H. (2016). Casilio: A versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Research*, *26*(2), 254–7. <https://doi.org/10.1038/cr.2016.3>
- Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., Chin, S. F., Caldas, C., Liu, S. J., Horlbeck, M. A., Lim, D. A., Weissman, J. S., Curtis, C., & Chang, H. Y. (2018). Promoter of lncRNA gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell*, *173*(6), 1398–1412.e22. <https://doi.org/10.1016/j.cell.2018.03.068>

REFERENCES

- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, *361*(6400). <https://doi.org/10.1126/science.aar2555>
- Chuang, C. H., Carpenter, A. E., Fuchsova, B., Johnson, T., de Lanerolle, P., & Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Current Biology*, *16*(8), 825–31. <https://doi.org/10.1016/j.cub.2006.03.059>
- Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., & Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Molecular Cell*, *5*(2), 243–54. [https://doi.org/10.1016/s1097-2765\(00\)80420-7](https://doi.org/10.1016/s1097-2765(00)80420-7)
- Cisse, I., Izeddin, I., Causse, S. Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., & Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. *Science*, *341*(6146), 664–7. <https://doi.org/10.1126/science.1239053>
- Clarkson, C. T., Deeks, E. A., Samarista, R., Mamayusupova, H., Zhurkin, V. B., & Teif, V. B. (2019). CTCF-dependent chromatin boundaries formed by asymmetric nucleosome arrays with decreased linker length. *Nucleic Acids Research*, *47*(21), 11181–11196. <https://doi.org/10.1093/nar/gkz908>
- Clow, P. A., Du, M., Jillette, N., Taghbalout, A., Zhu, J. J., & Cheng, A. W. (2022). CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nature Communications*, *13*(1), 1871. <https://doi.org/10.1038/s41467-022-29343-z>
- Corsi, F., Rusch, E., & Goloborodko, A. (2023). Loop extrusion rules: The next generation. *Current Opinion in Genetics & Development*, *81*, 102061. <https://doi.org/10.1016/j.gde.2023.102061>
- Coulon, A., Chow, C. C., Singer, R. H., & Larson, D. R. (2013). Eukaryotic transcriptional dynamics: From single molecules to cell populations. *Nature Review Genetics*, *14*(8), 572–84. <https://doi.org/10.1038/nrg3484>
- Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, *2*(4), 292–301. <https://doi.org/10.1038/35066075>
- Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L., & Kirsch-Volders, M. (1982). Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Human Genetics*, *62*(3), 201–9. <https://doi.org/10.1007/bf00333519>
- Cuartero, S., Weiss, F. D., Dharmalingam, G., Guo, Y., Ing-Simmons, E., Masella, S., Robles-Rebollo, I., Xiao, X., Wang, Y. F., Barozzi, I., Djeghloul, D., Amano, M. T., Niskanen, H., Petretto, E., Dowell, R. D., Tachibana, K., Kaikkonen, M. U., Nasmyth, K. A., Lenhard, B., ... Merckenschlager, M. (2018). Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nature Immunology*, *19*(9), 932–941. <https://doi.org/10.1038/s41590-018-0184-1>
- Cugliandolo, L. F., Gonnella, G., & Suma, A. (2015). Rotational and translational diffusion in an interacting active dumbbell system. *Physical Review E*, *91*(6), 062124. <https://doi.org/10.1103/PhysRevE.91.062124>
- da Costa-Nunes, J. A., & Noordermeer, D. (2023). TADs: Dynamic structures to create stable regulatory functions. *Current Opinion in Structural Biology*, *81*, 102622. <https://doi.org/10.1016/j.sbi.2023.102622>
- Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D., & Singer, R. H. (2007). *In vivo* dynamics of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, *14*(9), 796–806. <https://doi.org/10.1038/nsmb1280>
- Dauban, L., Montagne, R., Thierry, A., Lazar-Stefanita, L., Bastié, N., Gadal, O., Cournac, A., Koszul, R., & Beckouët, F. (2020). Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other partners. *Molecular Cell*, *77*(6), 1279–1293.e4. <https://doi.org/10.1016/j.molcel.2020.01.019>
- Davidson, I. F., Barth, R., Zaczek, M., van der Torre, J., Tang, W., Nagasaka, K., Janissen, R., Kerssemakers, J., Wutz, G., Dekker, C., & Peters, J. M. (2023). CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature*, *616*(7958), 822–827. <https://doi.org/10.1038/s41586-023-05961-5>
- Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., & Peters, J. M. (2019). DNA loop extrusion by human

- cohesin. *Science*, 366(6471), 1338–1345. <https://doi.org/10.1126/science.aaz3418>
- Davidson, I. F., & Peters, J. M. (2021). Genome folding through loop extrusion by SMC complexes. *Nature Reviews Molecular Cell Biology*, 22(7), 445–464. <https://doi.org/10.1038/s41580-021-00349-7>
- de Almeida, B. P., Reiter, F., Pagani, M., & Stark, A. (2022). DeepSTARR predicts enhancer activity from DNA sequence and enables the *de novo* design of synthetic enhancers. *Nature Genetics*, 54(5), 613–624. <https://doi.org/10.1038/s41588-022-01048-5>
- Deguchi, T., Iwanski, M. K., Schentarra, E. M., Heidebrecht, C., Schmidt, L., Heck, J., Weihs, T., Schnorrenberg, S., Hoess, P., Liu, S., Chevyreva, V., Noh, K. M., Kapitein, L. C., & Ries, J. (2023). Direct observation of motor protein stepping in living cells using MINFLUX. *Science*, 379(6636), 1010–1015. <https://doi.org/10.1126/science.ade2676>
- Dekker, J., & Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. *Cell*, 164(6), 1110–1121. <https://doi.org/10.1016/j.cell.2016.02.007>
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306–11. <https://doi.org/10.1126/science.1067799>
- DelRosso, N., Tycko, J., Suzuki, P., Andrews, C., Aradhana, Mukund, A., Liongson, I., Ludwig, C., Spees, K., Fordyce, P., Bassik, M. C., & Bintu, L. (2023). Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature*, 616(7956), 365–372. <https://doi.org/10.1038/s41586-023-05906-y>
- Dequeker, B. J. H., Scherr, M. J., Brandão, H. B., Gassler, J., Powell, S., Gaspar, I., Flyamer, I. M., Lalic, A., Tang, W., Stocsits, R., Davidson, I. F., Peters, J. M., Duderstadt, K. E., Mirny, L. A., & Tachibana, K. (2022). MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature*, 606(7912), 197–203. <https://doi.org/10.1038/s41586-022-04730-0>
- Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B. J., Flegel, V., Bucher, P., Jongeneel, C. V., & Antonarakis, S. E. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature*, 420(6915), 578–82. <https://doi.org/10.1038/nature01251>
- de Wit, E., & Nora, E. P. (2023). New insights into genome folding by loop extrusion from inducible degron technologies. *Nature Review Genetics*, 24(2), 73–85. <https://doi.org/10.1038/s41576-022-00530-4>
- Diebold-Durand, M. L., Lee, H., Ruiz Avila, L. B., Noh, H., Shin, H. C., Im, H., Bock, F. P., Bürmann, F., Durand, A., Basfeld, A., Ham, S., Basquin, J., Oh, B. H., & Gruber, S. (2017). Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Molecular Cell*, 67(2), 334–347.e5. <https://doi.org/10.1016/j.molcel.2017.06.010>
- Dixon, J. R., Gorkin, D. U., & Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell*, 62(5), 668–80. <https://doi.org/10.1016/j.molcel.2016.05.018>
- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A., & Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539), 331–6. <https://doi.org/10.1038/nature14222>
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376–80. <https://doi.org/10.1038/nature11082>
- Donovan, B. T., Huynh, A., Ball, D. A., Patel, H. P., Poirier, M. G., Larson, D. R., Ferguson, M. L., & Lenstra, T. L. (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *EMBO Journal*, 38(12). <https://doi.org/10.15252/embj.2018100809>
- Dorigi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., Still, 2., C. D., Garcia, B. A., Adelman, K., & Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription

REFERENCES

- from Promoters Independently of H3K4 Monomethylation. *Molecular Cell*, 66(4), 568–576.e4. <https://doi.org/10.1016/j.molcel.2017.04.018>
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D., & Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 16(10), 1299–309. <https://doi.org/10.1101/gr.5571506>
- Driscoll, M. C., Dobkin, C. S., & Alter, B. P. (1989). Gamma delta beta-thalassemia due to a *de novo* mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proceedings of the National Academy of Sciences of the United States of America*, 86(19), 7470–4. <https://doi.org/10.1073/pnas.86.19.7470>
- Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C., & Taddei, A. (2011). Tight protein-DNA interactions favor gene silencing. *Genes & Development*, 25(13), 1365–70. <https://doi.org/10.1101/gad.611011>
- Dünweg, B., & Kremer, K. (1993). Molecular dynamics simulation of a polymer chain in solution. *The Journal of Chemical Physics*, 99(9), 6983–6997. <https://doi.org/10.1063/1.465445>
- Dynan, W. S., & Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, 316(6031), 774–8. <https://doi.org/10.1038/316774a0>
- Eichenberger, B. T., Zhan, Y., Rempfler, M., Giorgetti, L., & Chao, J. A. (2021). DeepBlink: Threshold-independent detection and localization of diffraction-limited spots. *Nucleic Acids Research*, 49(13), 7292–7297. <https://doi.org/10.1093/nar/gkab546>
- Emerson, D. J., Zhao, P. A., Cook, A. L., Barnett, R. J., Klein, K. N., Saulebekova, D., Ge, C., Zhou, L., Simandi, Z., Minsk, M. K., Titus, K. R., Wang, W., Gong, W., Zhang, D., Yang, L., Venev, S. V., Gibcus, J. H., Yang, H., Sasaki, T., ... Phillips-Cremins, J. E. (2022). Cohesin-mediated loop anchors confine the locations of human replication origins. *Nature*, 606(7915), 812–819. <https://doi.org/10.1038/s41586-022-04803-0>
- ENCODE Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., & Mirny, L. A. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, 570(7761), 395–399. <https://doi.org/10.1038/s41586-019-1275-3>
- Ferrie, J. J., Karr, J. P., Graham, T. G., Dailey, G. M., Zhang, G., Tjian, R., & Darzacq, X. (2023). P300 Is an Obligate Integrator of Combinatorial Transcription Factor Inputs. *bioRxiv*, 2023.05.18.541220. <https://doi.org/10.1101/2023.05.18.541220>
- Field, A., & Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annual Review of Biochemistry*, 89, 213–234. <https://doi.org/10.1146/annurev-biochem-011420-095916>
- Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C., & Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. *Nature*, 513(7516), 120–3. <https://doi.org/10.1038/nature13695>
- Finn, E. H., & Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science*, 365(6457). <https://doi.org/10.1126/science.aaw9498>
- Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., Mirny, L., & Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, 176(6), 1502–1515.e10. <https://doi.org/10.1016/j.cell.2019.01.020>
- Flavahan, W. A., Drier, Y., Liao, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L., & Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature*, 529(7584), 110–4. <https://doi.org/10.1038/nature16490>
- Flemming, W. (1882). *Zellsubstanz, Kern und Zelltheilung*. Vogel.
- Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A.,

- & Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, *544*(7648), 110–114. <https://doi.org/10.1038/nature21711>
- Flyamer, I. M., Illingworth, R. S., & Bickmore, W. A. (2020). Coolpup.py: Versatile pile-up analysis of Hi-C data. *Bioinformatics*, *36*(10), 2980–2985. <https://doi.org/10.1093/bioinformatics/btaa073>
- Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., & Mirny, L. A. (2017). Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harbor Symposia on Quantitative Biology*, *82*, 45–55. <https://doi.org/10.1101/sqb.2017.82.034710>
- Fudenberg, G., & Imakaev, M. (2017). FISH-ing for captured contacts: Towards reconciling FISH and 3C. *Nature Methods*, *14*(7), 673–678. <https://doi.org/10.1038/nmeth.4329>
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–49. <https://doi.org/10.1016/j.celrep.2016.04.085>
- Fukai, Y. T. (2021). Laptrack. <https://doi.org/10.5281/zenodo.5519537>
- Fukai, Y. T., & Kawaguchi, K. (2022). LapTrack: Linear Assignment Particle Tracking with Tunable Metrics. *Bioinformatics*, btac799. <https://doi.org/10.1093/bioinformatics/btac799>
- Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, *51*(12), 1664–1669. <https://doi.org/10.1038/s41588-019-0538-0>
- Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., Hsieh, T. S., Mirny, L., Zechner, C., & Hansen, A. S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, *376*(6592), 496–501. <https://doi.org/10.1126/science.abn6583>
- Galupa, R., Alvarez-Canales, G., Borst, N. O., Fuqua, T., Gandara, L., Misunou, N., Richter, K., Alves, M. R. P., Karumbi, E., Perkins, M. L., Kocijan, T., Rushlow, C. A., & Crocker, J. (2023). Enhancer architecture and chromatin accessibility constrain phenotypic space during *Drosophila* development. *Developmental Cell*, *58*(1), 51–62.e4. <https://doi.org/10.1016/j.devcel.2022.12.003>
- Galupa, R., & Heard, E. (2015). X-chromosome inactivation: New insights into *cis* and *trans* regulation. *Current Opinion in Genetics & Development*, *31*, 57–66. <https://doi.org/10.1016/j.gde.2015.04.002>
- Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science*, *360*(6384), 102–105. <https://doi.org/10.1126/science.aar7831>
- Gasparini, M., Tome, J. M., & Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics*, *21*(5), 292–310. <https://doi.org/10.1038/s41576-019-0209-0>
- Gautier, A., Juillerat, A., Heinis, C., Corrêa, J., I. R., Kindermann, M., Beaufils, F., & Johnsson, K. (2008). An engineered protein tag for multiprotein labeling in living cells. *ACS Chemical Biology*, *15*(2), 128–36. <https://doi.org/10.1016/j.chembiol.2008.01.007>
- Gerlich, D., Hirota, T., Koch, B., Peters, J. M., & Ellenberg, J. (2006). Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Current Biology*, *16*(4), 333–44. <https://doi.org/10.1016/j.cub.2005.12.040>
- Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H. A., Sellou, H., Politi, A. Z., Ellenberg, J., Gallardo, F., & Bystricky, K. (2017). Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophysical Journal*, *113*(7), 1383–1394. <https://doi.org/10.1016/j.bpj.2017.08.014>
- Gibcus, J. H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M. T., Xie, L., Paulson, J. R., Earnshaw, W. C., Mirny, L. A., & Dekker, J. (2018). A pathway for mitotic chromosome

REFERENCES

- formation. *Science*, 359(6376). <https://doi.org/10.1126/science.aao6135>
- Gillies, S. D., Morrison, S. L., Oi, V. T., & Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, 33(3), 717–28. [https://doi.org/10.1016/0092-8674\(83\)90014-4](https://doi.org/10.1016/0092-8674(83)90014-4)
- Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F. J., & Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology*, 38(4), 433–438. <https://doi.org/10.1038/s41587-020-0407-5>
- Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G., & Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4), 950–63. <https://doi.org/10.1016/j.cell.2014.03.025>
- Giorgetti, L., & Heard, E. (2016). Closing the loop: 3C versus DNA FISH. *Genome Biology*, 17(1), 215. <https://doi.org/10.1186/s13059-016-1081-2>
- Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., Chen, C. J., Kaplan, N., Chang, H. Y., Heard, E., & Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature*, 535(7613), 575–9. <https://doi.org/10.1038/nature18589>
- Giorgetti, L., Hansen, A., Mirny, L., Grosse-Holz, S., Mach, P., Kos, P., & Zhan, Y. (2022). *Comparison of looping dynamics*. Retrieved October 3, 2023, from <https://twitter.com/LucaGiorgetti/status/1599854837466038272>
- Goel, V. Y., Huseyin, M. K., & Hansen, A. S. (2023). Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nature Genetics*, 55(6), 1048–1056. <https://doi.org/10.1038/s41588-023-01391-1>
- Golfier, S., Quail, T., Kimura, H., & Brugués, J. (2020). Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *Elife*, 9. <https://doi.org/10.7554/eLife.53885>
- Goloborodko, A., Marko, J. F., & Mirny, L. A. (2016). Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal*, 110(10), 2162–8. <https://doi.org/10.1016/j.bpj.2016.02.041>
- Göös, H., Kinnunen, M., Salokas, K., Tan, Z., Liu, X., Yadav, L., Zhang, Q., Wei, G. H., & Varjosalo, M. (2022). Human transcription factor protein interaction networks. *Nature Communications*, 13(1), 766. <https://doi.org/10.1038/s41467-022-28341-5>
- Graham, T. G. W., Ferrie, J. J., Dailey, G. M., Tjian, R., & Darzacq, X. (2022). Detecting molecular interactions in live-cell single-molecule imaging with proximity-assisted photoactivation (PAPA). *Elife*, 11. <https://doi.org/10.7554/eLife.76870>
- Grimm, J. B., Muthusamy, A. K., Liang, Y., Brown, T. A., Lemon, W. C., Patel, R., Lu, R., Macklin, J. J., Keller, P. J., Ji, N., & Lavis, L. D. (2017). A general method to fine-tune fluorophores for live-cell and *in vivo* imaging. *Nature Methods*, 14(10), 987–994. <https://doi.org/10.1038/nmeth.4403>
- Grosberg, A. Y., Nechaev, S. K., & Shakhnovich, E. I. (1988). The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de physique*, 49(12), 2095–2100.
- Grosberg, A. Y., Khokhlov, A. R., Stanley, H. E., Mallinckrodt, A. J., & McKay, S. (1995). Statistical physics of macromolecules. *Computers in Physics*, 9(2), 171–172.
- Gu, B., Swigut, T., Spencley, A., Bauer, M. R., Chung, M., Meyer, T., & Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian *cis*-regulatory elements. *Science*, 359(6379), 1050–1055. <https://doi.org/10.1126/science.aao3136>
- Guacci, V., Koshland, D., & Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell*, 91(1), 47–57. [https://doi.org/10.1016/s0092-8674\(01\)80008-8](https://doi.org/10.1016/s0092-8674(01)80008-8)
- Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., Bates, J. G., Richards, N., Myers, D., Patel, H., Gallagher, M., Schlissel, M. S., Murre, C., Busslinger, M.,

- Giallourakis, C. C., & Alt, F. W. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature*, *477*(7365), 424–30. <https://doi.org/10.1038/nature10495>
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T., & Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–10. <https://doi.org/10.1016/j.cell.2015.07.038>
- Gurskaya, N. G., Verkhusha, V. V., Shcheglov, A. S., Staroverov, D. B., Chepurnykh, T. V., Fradkov, A. F., Lukyanov, S., & Lukyanov, K. A. (2006). Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature Biotechnology*, *24*(4), 461–5. <https://doi.org/10.1038/nbt1191>
- Gustafsson, M. G., Shao, L., Carlton, P. M., Wang, C. J., Golubovskaya, I. N., Cande, W. Z., Agard, D. A., & Sedat, J. W. (2008). Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical Journal*, *94*(12), 4957–70. <https://doi.org/10.1529/biophysj.107.120345>
- Gustavsson, A. K., Ghosh, R. P., Petrov, P. N., Liphardt, J. T., & Moerner, W. E. (2022). Fast and parallel nanoscale three-dimensional tracking of heterogeneous mammalian chromatin dynamics. *Molecular Biology of the Cell*, *33*(6), ar47. <https://doi.org/10.1091/mbc.E21-10-0514>
- Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Flach, K. D., Teunissen, H., Willems, L., Brummelkamp, T. R., Rowland, B. D., & de Wit, E. (2022). A Mediator-cohesin axis controls heterochromatin domain formation. *Nature Communications*, *13*(1), 754. <https://doi.org/10.1038/s41467-022-28377-7>
- Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P. H. L., Teunissen, H., Medema, R. H., van Steensel, B., Brummelkamp, T. R., de Wit, E., & Rowland, B. D. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*, *169*(4), 693–707.e14. <https://doi.org/10.1016/j.cell.2017.04.013>
- Haberle, V., Arnold, C. D., Pagani, M., Rath, M., Schernhuber, K., & Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. *Nature*, *570*(7759), 122–126. <https://doi.org/10.1038/s41586-019-1210-7>
- Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J., & Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature*, *454*(7202), 297–301. <https://doi.org/10.1038/nature07098>
- Hanahan, D. (1985). Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. *Nature*, *315*(6015), 115–22. <https://doi.org/10.1038/315115a0>
- Hansen, A. S. (2020). CTCF as a boundary factor for cohesin-mediated loop extrusion: Evidence for a multi-step mechanism. *Nucleus*, *11*(1), 132–148. <https://doi.org/10.1080/19491034.2020.1782024>
- Hansen, A. S., Cattoglio, C., Darzacq, X., & Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*, *9*(1), 20–32. <https://doi.org/10.1080/19491034.2017.1389365>
- Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*, *6*. <https://doi.org/10.7554/eLife.25776>
- Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D. R., & Yu, H. (2014). Structure of cohesin subcomplex pinpoints direct Shugoshin-Wapl antagonism in centromeric cohesion. *Nature Structural & Molecular Biology*, *21*(10), 864–70. <https://doi.org/10.1038/nsmb.2880>
- Hardison, R. C., Oeltjen, J., & Miller, W. (1997). Long human-mouse sequence alignments reveal novel regulatory elements: A reason to sequence the mouse genome. *Genome Research*, *7*(10), 959–66. <https://doi.org/10.1101/gr.7.10.959>
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–8. <https://doi.org/10.1038/ng1966>

REFERENCES

- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrbücher für Wissenschaftliche Botanik*, 69, 762–818.
- Hendy, O., Serebreni, L., Bergauer, K., Muerdter, F., Huber, L., Nemčko, F., & Stark, A. (2022). Developmental and housekeeping transcriptional programs in *Drosophila* require distinct chromatin remodelers. *Molecular Cell*, 82(19), 3598–3612.e7. <https://doi.org/10.1016/j.molcel.2022.08.019>
- Higashi, T. L., Pobegalov, G., Tang, M., Molodtsov, M. I., & Uhlmann, F. (2021). A Brownian ratchet model for DNA loop extrusion by the cohesin complex. *Elife*, 10. <https://doi.org/10.7554/eLife.67530>
- Hill, L., Ebert, A., Jaritz, M., Wutz, G., Nagasaka, K., Tagoh, H., Kostanova-Poliakova, D., Schindler, K., Sun, Q., Bönelt, P., Fischer, M., Peters, J. M., & Busslinger, M. (2020). *Wapl* repression by Pax5 promotes V gene recombination by *Igh* loop extrusion. *Nature*, 584(7819), 142–147. <https://doi.org/10.1038/s41586-020-2454-y>
- Hirano, T., Kobayashi, R., & Hirano, M. (1997). Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a *Xenopus* homolog of the *Drosophila* Barren protein. *Cell*, 89(4), 511–21. [https://doi.org/10.1016/s0092-8674\(00\)80233-0](https://doi.org/10.1016/s0092-8674(00)80233-0)
- Hirano, T., & Mitchison, T. J. (1994). A heterodimeric coiled-coil protein required for mitotic chromosome condensation *in vitro*. *Cell*, 79(3), 449–58. [https://doi.org/10.1016/0092-8674\(94\)90254-2](https://doi.org/10.1016/0092-8674(94)90254-2)
- Hirota, T., Gerlich, D., Koch, B., Ellenberg, J., & Peters, J. M. (2004). Distinct functions of condensin I and II in mitotic chromosome assembly. *Journal of Cell Science*, 117(Pt 26), 6435–45. <https://doi.org/10.1242/jcs.01604>
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, 169(1), 13–23. <https://doi.org/10.1016/j.cell.2017.02.007>
- Höck, E., Buchholz, T.-O., Brachmann, A., Jug, F., & Freytag, A. (2022). N2V2 – Fixing Noise2Void Checkerboard Artifacts with Modified Sampling Strategies and a Tweaked Network Architecture. *arXiv*. <https://doi.org/10.48550/arXiv.2211.08512>
- Hoencamp, C., & Rowland, B. D. (2023). Genome control by SMC complexes. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/s41580-023-00609-8>
- Holzmann, J., Politi, A. Z., Nagasaka, K., Hantsche-Grininger, M., Walther, N., Koch, B., Fuchs, J., Dürnberger, G., Tang, W., Ladurner, R., Stocsits, R. R., Busslinger, G. A., Novák, B., Mechtler, K., Davidson, I. F., Ellenberg, J., & Peters, J. M. (2019). Absolute quantification of cohesin, CTCF and their regulators in human cells. *Elife*, 8. <https://doi.org/10.7554/eLife.46269>
- Hou, C., Li, L., Qin, Z. S., & Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Molecular Cell*, 48(3), 471–84. <https://doi.org/10.1016/j.molcel.2012.08.031>
- Houlard, M., Cutts, E. E., Shamim, M. S., Godwin, J., Weisz, D., Presser Aiden, A., Lieberman Aiden, E., Schermelleh, L., Vannini, A., & Nasmyth, K. (2021). MCPH1 inhibits Condensin II during interphase by regulating its SMC2-Kleisin interface. *Elife*, 10. <https://doi.org/10.7554/eLife.73348>
- Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Darzacq, X., & Tjian, R. (2022). Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nature Genetics*, 54(12), 1919–1932. <https://doi.org/10.1038/s41588-022-01223-8>
- Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell*, 78(3), 539–553.e8. <https://doi.org/10.1016/j.molcel.2020.03.002>
- Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K. L., Upcher, W., Godlee, C., Roig, M. B., Shirahige, K., & Nasmyth, K. (2011). ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Current Biology*, 21(1), 12–24. <https://doi.org/10.1016/j.cub.2010.12.004>
- Hu, M., Deng, K., Qin, Z., Dixon, J., Selvaraj, S., Fang, J., Ren, B., & Liu, J. S. (2013). Bayesian inference of spatial organizations of chromosomes. *PLOS Computational Biology*, 9(1), e1002893. <https://doi.org/10.1371/journal.pcbi.1002893>

1/journal.pcbi.1002893

- Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A. M., Yu, M., Hu, R., Tastemel, M., Juric, I., Hu, M., Nicodemi, M., Zhuang, X., & Ren, B. (2021). CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics*, *53*(7), 1064–1074. <https://doi.org/10.1038/s41588-021-00863-6>
- Huis in 't Veld, P. J., Herzog, F., Ladurner, R., Davidson, I. F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R., & Peters, J. M. (2014). Characterization of a DNA exit gate in the human cohesin ring. *Science*, *346*(6212), 968–72. <https://doi.org/10.1126/science.1256904>
- Hunt, G., Boija, A., & Mannervik, M. (2022). P300/CBP sustains Polycomb silencing by non-enzymatic functions. *Molecular Cell*, *82*(19), 3580–3597.e9. <https://doi.org/10.1016/j.molcel.2022.09.005>
- Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J., & Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, *9*(10), 999–1003. <https://doi.org/10.1038/nmeth.2148>
- Imakaev, M. V., Fudenberg, G., & Mirny, L. A. (2015). Modeling chromosomes: Beyond pretty pictures. *FEBS Letters*, *589*(20 Pt A), 3031–6. <https://doi.org/10.1016/j.febslet.2015.09.004>
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–45. <https://doi.org/10.1038/nature03001>
- Isbel, L., Grand, R. S., & Schübeler, D. (2022). Generating specificity in genome regulation through transcription factor sensitivity to chromatin. *Nature Reviews Genetics*, *23*(12), 728–740. <https://doi.org/10.1038/s41576-022-00512-6>
- Iurlaro, M., Stadler, M. B., Masoni, F., Jagani, Z., Galli, G. G., & Schübeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nature Genetics*, *53*(3), 279–287. <https://doi.org/10.1038/s41588-020-00768-w>
- Jaeger, M. G., Schwalb, B., Mackowiak, S. D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., Ferguson, F. M., Müller, A. C., Bergthaler, A., Gray, N. S., Bradner, J. E., Bock, C., Hnisz, D., Cramer, P., & Winter, G. E. (2020). Selective Mediator dependence of cell-type-specifying transcription. *Nature Genetics*, *52*(7), 719–727. <https://doi.org/10.1038/s41588-020-0635-0>
- Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., ... Fraser, P. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, *167*(5), 1369–1384.e19. <https://doi.org/10.1016/j.cell.2016.09.037>
- Jerković, I., & Cavalli, G. (2021). Understanding 3D genome organization by multidisciplinary methods. *Nature Reviews Molecular Cell Biology*, *22*(8), 511–528. <https://doi.org/10.1038/s41580-021-00362-w>
- Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., & Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, *527*(7578), 384–8. <https://doi.org/10.1038/nature15518>
- Jost, D., Carrivain, P., Cavalli, G., & Vaillant, C. (2014). Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, *42*(15), 9553–61. <https://doi.org/10.1093/nar/gku698>
- Jung, I., Schmitt, A., Diao, Y., Lee, A. J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliah, E., Barr, C. L., Li, B., Kuan, S., Kim, D., & Ren, B. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*, *51*(10), 1442–1449. <https://doi.org/10.1038/s41588-019-0494-8>
- Kaaij, L. J. T., Mohn, F., van der Weide, R. H., de Wit, E., & Bühler, M. (2019). The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. *Cell*, *178*(6),

REFERENCES

- 1437–1451.e14. <https://doi.org/10.1016/j.cell.2019.08.007>
- Kane, L., Williamson, I., Flyamer, I. M., Kumar, Y., Hill, R. E., Lettice, L. A., & Bickmore, W. A. (2022). Cohesin is required for long-range enhancer action at the *Shh* locus. *Nature Structural & Molecular Biology*, 29(9), 891–897. <https://doi.org/10.1038/s41594-022-00821-8>
- Karr, J. P., Ferrie, J. J., Tjian, R., & Darzacq, X. (2022). The transcription factor activity gradient (TAG) model: Contemplating a contact-independent mechanism for enhancer-promoter communication. *Genes & Development*, 36(1-2), 7–16. <https://doi.org/10.1101/gad.349160.121>
- Keizer, V. I. P., Grosse-Holz, S., Woringer, M., Zambon, L., Aizel, K., Bongaerts, M., Delille, F., Kolar-Znika, L., Scolari, V. F., Hoffmann, S., Banigan, E. J., Mirny, L. A., Dahan, M., Fachinetti, D., & Coulon, A. (2022). Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science*, 377(6605), 489–495. <https://doi.org/10.1126/science.abi9810>
- Kepten, E., Bronshtein, I., & Garini, Y. (2013). Improved estimation of anomalous diffusion exponents in single-particle tracking experiments. *Physical Review E*, 87(5), 052713. <https://doi.org/10.1103/PhysRevE.87.052713>
- Khanna, N., Zhang, Y., Lucas, J. S., Dudko, O. K., & Murre, C. (2019). Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nature Communications*, 10(1), 2771. <https://doi.org/10.1038/s41467-019-10628-9>
- Khokhlov, A. R., Grosberg, A. Y., & Pande, V. S. (1994). *Statistical physics of macromolecules* (Vol. 1). Springer.
- Kiefer, L., Chiosso, A., Langen, J., Buckley, A., Gaudin, S., Rajkumar, S. M., Servito, G. I. F., Cha, E. S., Vijay, A., Yeung, A., Horta, A., Mui, M. H., & Canzio, D. (2023). WAPL functions as a rheostat of Protocadherin isoform diversity that controls neural wiring. *Science*, 380(6651), eadf8440. <https://doi.org/10.1126/science.adf8440>
- Kim, D. S., Risca, V. I., Reynolds, D. L., Chappell, J., Rubin, A. J., Jung, N., Donohue, L. K. H., Lopez-Pajares, V., Kathiria, A., Shi, M., Zhao, Z., Deep, H., Sharmin, M., Rao, D., Lin, S., Chang, H. Y., Snyder, M. P., Greenleaf, W. J., Kundaje, A., & Khavari, P. A. (2021). The dynamic, combinatorial *cis*-regulatory lexicon of epidermal differentiation. *Nature Genetics*, 53(11), 1564–1576. <https://doi.org/10.1038/s41588-021-00947-3>
- Kim, E., Kerssemakers, J., Shaltiel, I. A., Haering, C. H., & Dekker, C. (2020). DNA-loop extruding condensin complexes can traverse one another. *Nature*, 579(7799), 438–442. <https://doi.org/10.1038/s41586-020-2067-5>
- Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative *cis*-regulatory code. *Molecular Cell*, 83(3), 373–392. <https://doi.org/10.1016/j.molcel.2022.12.032>
- Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanekov, V. V., & Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128(6), 1231–45. <https://doi.org/10.1016/j.cell.2006.12.048>
- Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., & Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. *Science*, 366(6471), 1345–1349. <https://doi.org/10.1126/science.aaz4475>
- Kioussis, D., Vanin, E., deLange, T., Flavell, R. A., & Grosfeld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, 306(5944), 662–6. <https://doi.org/10.1038/306662a0>
- Kitajima, T. S., Sakuno, T., Ishiguro, K., Iemura, S., Natsume, T., Kawashima, S. A., & Watanabe, Y. (2006). Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature*, 441(7089), 46–52. <https://doi.org/10.1038/nature04663>
- Klar, T. A., Jakobs, S., Dyba, M., Egnér, A., & Hell, S. W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15), 8206–10. <https://doi.org/10.1073/pnas.97.15.8206>

- Kong, M., Cutts, E. E., Pan, D., Beuron, F., Kaliyappan, T., Xue, C., Morris, E. P., Musacchio, A., Vannini, A., & Greene, E. C. (2020). Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Molecular Cell*, *79*(1), 99–114.e9. <https://doi.org/10.1016/j.molcel.2020.04.026>
- Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., Nowaczyk, M. J., Toriello, H., Bamshad, M. J., Carey, J. C., Rappaport, E., Kawachi, S., Lander, A. D., Calof, A. L., Li, H. H., ... Jackson, L. G. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genetics*, *36*(6), 631–5. <https://doi.org/10.1038/ng1364>
- Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., & Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell*, *67*(3), 411–422.e4. <https://doi.org/10.1016/j.molcel.2017.06.027>
- Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S., & Krebs, A. R. (2023). Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. *Molecular Cell*, *83*(5), 787–802.e9. <https://doi.org/10.1016/j.molcel.2023.01.017>
- Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, *78*(3), 554–565.e7. <https://doi.org/10.1016/j.molcel.2020.03.003>
- Krull, A., Buchholz, T.-O., & Jug, F. (2019). Noise2Void – Learning Denoising from Single Noisy Images. *arXiv*. <https://doi.org/10.48550/arXiv.1811.10980>
- Kschonsak, M., Merkel, F., Bisht, S., Metz, J., Rybin, V., Hassler, M., & Haering, C. H. (2017). Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*, *171*(3), 588–600.e24. <https://doi.org/10.1016/j.cell.2017.09.008>
- Kueng, S., Hegemann, B., Peters, B. H., Lipp, J. J., Schleiffer, A., Mechtler, K., & Peters, J. M. (2006). Wapl controls the dynamic association of cohesin with chromatin. *Cell*, *127*(5), 955–67. <https://doi.org/10.1016/j.cell.2006.09.040>
- Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., & Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(28), 10684–9. <https://doi.org/10.1073/pnas.0600326103>
- Ladurner, R., Kreidl, E., Ivanov, M. P., Ekker, H., Idarraga-Amado, M. H., Busslinger, G. A., Wutz, G., Cisneros, D. A., & Peters, J. M. (2016). Sororin actively maintains sister chromatid cohesion. *EMBO Journal*, *35*(6), 635–53. <https://doi.org/10.15252/embj.201592532>
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>
- Lammers, N. C., Kim, Y. J., Zhao, J., & Garcia, H. G. (2020). A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Current Opinion in Cell Biology*, *67*, 147–157. <https://doi.org/10.1016/j.ceb.2020.08.001>
- Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature*, *547*(7662), 236–240. <https://doi.org/10.1038/nature22822>
- Lau, I. F., Filipe, S. R., Søballe, B., Økstad, O. A., Barre, F. X., & Sherratt, D. J. (2003). Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Molecular Microbiology*, *49*(3), 731–43. <https://doi.org/10.1046/j.1365-2958.2003.03640.x>

REFERENCES

- Lawrimore, J., Aicher, J. K., Hahn, P., Fulp, A., Kompa, B., Vicci, L., Falvo, M., Taylor, 2., R. M., & Bloom, K. (2016). ChromoShake: A chromosome dynamics simulator reveals that chromatin loops stiffen centromeric chromatin. *Molecular Biology of the Cell*, *27*(1), 153–66. <https://doi.org/10.1091/mbc.E15-08-0575>
- Le, T. B., Imakaev, M. V., Mirny, L. A., & Laub, M. T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, *342*(6159), 731–4. <https://doi.org/10.1126/science.1242059>
- Lee, B. G., Merkel, F., Allegretti, M., Hassler, M., Cawood, C., Lecomte, L., O'Reilly, F. J., Sinn, L. R., Gutierrez-Escribano, P., Kschonsak, M., Bravo, S., Nakane, T., Rappsilber, J., Aragon, L., Beck, M., Löwe, J., & Haering, C. H. (2020). Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. *Nature Structural & Molecular Biology*, *27*(8), 743–751. <https://doi.org/10.1038/s41594-020-0457-x>
- Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K., & Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, *430*(6999), 573–8. <https://doi.org/10.1038/nature02742>
- Levi, V., Ruan, Q., Plutz, M., Belmont, A. S., & Gratton, E. (2005). Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophysical Journal*, *89*(6), 4275–85. <https://doi.org/10.1529/biophysj.105.066670>
- Levings, P. P., & Bungert, J. (2002). The human beta-globin locus control region. *European Journal of Biochemistry*, *269*(6), 1589–99. <https://doi.org/10.1046/j.1432-1327.2002.02797.x>
- Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, *128*(4), 707–19. <https://doi.org/10.1016/j.cell.2007.01.015>
- Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B. K., & Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. *Nature*, *524*(7564), 186–91. <https://doi.org/10.1038/nature14685>
- Li, Q., Zhou, B., Powers, P., Enver, T., & Stamatoyannopoulos, G. (1991). Primary structure of the goat beta-globin locus control region. *Genomics*, *9*(3), 488–99. [https://doi.org/10.1016/0888-7543\(91\)90415-b](https://doi.org/10.1016/0888-7543(91)90415-b)
- Li, Y., Haarhuis, J. H. I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M. S., Willems, L., Teunissen, H., Muir, K. W., de Wit, E., Rowland, B. D., & Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature*, *578*(7795), 472–476. <https://doi.org/10.1038/s41586-019-1910-z>
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–93. <https://doi.org/10.1126/science.1181369>
- Lingling, C., Chayan, D., Jieru, L., & Alexandros, P. (2023). Mechanisms of transcription control by distal enhancers from high-resolution single-gene imaging. *bioRxiv*, 2023.03.19.533190. <https://doi.org/10.1101/2023.03.19.533190>
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–22. <https://doi.org/10.1038/nature08514>
- Liu, H., Dong, P., Ioannou, M. S., Li, L., Shea, J., Pasolli, H. A., Grimm, J. B., Rivlin, P. K., Lavis, L. D., Koyama, M., & Liu, Z. (2018). Visualizing long-term single-molecule dynamics *in vivo* by stochastic protein labeling. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), 343–348. <https://doi.org/10.1073/pnas.1713895115>
- Liu, M., Lu, Y., Yang, B., Chen, Y., Radda, J. S. D., Hu, M., Katz, S. G., & Wang, S. (2020). Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nature Communications*, *11*(1), 2907. <https://doi.org/10.1038/s41467-020-16732-5>
- Liu, N. Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M., Teunissen, H., Bruneau, B. G., Nora, E. P., &

- de Wit, E. (2021). WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics*, 53(1), 100–109. <https://doi.org/10.1038/s41588-020-00744-4>
- Liu, Y., & Dekker, J. (2022). CTCF-CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nature Cell Biology*, 24(10), 1516–1527. <https://doi.org/10.1038/s41556-022-00992-y>
- Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken *c-myc* gene. *Oncogene*, 5(12), 1743–53.
- Long, H. K., Osterwalder, M., Welsh, I. C., Hansen, K., Davies, J. O. J., Liu, Y. E., Koska, M., Adams, A. T., Aho, R., Arora, N., Ikeda, K., Williams, R. M., Sauka-Spengler, T., Porteus, M. H., Mohun, T., Dickel, D. E., Swigut, T., Hughes, J. R., Higgs, D. R., ... Wysocka, J. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell*, 27(5), 765–783.e14. <https://doi.org/10.1016/j.stem.2020.09.001>
- Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., & Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, 288(5463), 136–40. <https://doi.org/10.1126/science.288.5463.136>
- Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., & Wood, K. V. (2008). HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology*, 3(6), 373–82. <https://doi.org/10.1021/cb800025k>
- Losada, A., Hirano, M., & Hirano, T. (1998). Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes & Development*, 12(13), 1986–97. <https://doi.org/10.1101/gad.12.13.1986>
- Lu, T., Ang, C. E., & Zhuang, X. (2022). Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell*, 185(23), 4448–4464.e17. <https://doi.org/10.1016/j.cell.2022.09.035>
- Lucas, J. S., Zhang, Y., Dudko, O. K., & Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. *Cell*, 158(2), 339–352. <https://doi.org/10.1016/j.cell.2014.05.036>
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–60. <https://doi.org/10.1038/38444>
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., ... Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), 1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004>
- Ma, Y., Wang, M., Li, W., Zhang, Z., Zhang, X., Tan, T., Zhang, X. E., & Cui, Z. (2017). Live cell imaging of single genomic loci with quantum dot-labeled TALEs. *Nature Communications*, 8, 15318. <https://doi.org/10.1038/ncomms15318>
- Mach, P., Kos, P. I., Zhan, Y., Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., & Giorgetti, L. (2022). Cohesin and CTCF control the dynamics of chromosome folding. *Nature Genetics*, 54(12), 1907–1918. <https://doi.org/10.1038/s41588-022-01232-7>
- Mach, P., & Giorgetti, L. (2023). Integrative approaches to study enhancer–promoter communication. *Current Opinion in Genetics & Development*, 80, 102052. <https://doi.org/https://doi.org/10.1016/j.gde.2023.102052>

REFERENCES

- 02052
- MacPherson, M. J., & Sadowski, P. D. (2010). The CTCF insulator protein forms an unusual DNA structure. *BMC Molecular Biology*, *11*, 101. <https://doi.org/10.1186/1471-2199-11-101>
- Magde, D., Elson, E., & Webb, W. W. (1972). Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Physical Review Letters*, *29*(11), 705.
- Marchal, C., Sima, J., & Gilbert, D. M. (2019). Control of DNA replication timing in the 3D genome. *Nature Reviews Molecular Cell Biology*, *20*(12), 721–737. <https://doi.org/10.1038/s41580-019-0162-y>
- Margot, J. B., Demers, G. W., & Hardison, R. C. (1989). Complete nucleotide sequence of the rabbit beta-like globin gene cluster. Analysis of intergenic sequences and comparison with the human beta-like globin gene cluster. *Journal of Molecular Biology*, *205*(1), 15–40. [https://doi.org/10.1016/0022-2836\(89\)90362-8](https://doi.org/10.1016/0022-2836(89)90362-8)
- Marko, J. F., De Los Rios, P., Barducci, A., & Gruber, S. (2019). DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes. *Nucleic Acids Research*, *47*(13), 6956–6972. <https://doi.org/10.1093/nar/gkz497>
- Marshall, W. F., Straight, A., Marko, J. F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A. W., Agard, D. A., & Sedat, J. W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology*, *7*(12), 930–9. [https://doi.org/10.1016/s0960-9822\(06\)00412-x](https://doi.org/10.1016/s0960-9822(06)00412-x)
- Martin, B. J., Ablondi, E. F., Goglia, C., & Adelman, K. (2023). Global identification of direct SWI/SNF targets reveals compensation by EP400. *bioRxiv*, 2023.03.07.531379. <https://doi.org/10.1101/2023.03.07.531379>
- Martinez, S. R., & Miranda, J. L. (2010). CTCF terminal segments are unstructured. *Protein Science*, *19*(5), 1110–6. <https://doi.org/10.1002/pro.367>
- Martinez-Ara, M., Comoglio, F., van Arensbergen, J., & van Steensel, B. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Molecular Cell*, *82*(13), 2519–2531.e6. <https://doi.org/10.1016/j.molcel.2022.04.009>
- Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A. S., & Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell*, *145*(3), 447–58. <https://doi.org/10.1016/j.cell.2011.03.032>
- Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., & Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, *568*(7750), 49–54. <https://doi.org/10.1038/s41586-019-1035-4>
- Mazza, D., Abernathy, A., Golob, N., Morisaki, T., & McNally, J. G. (2012). A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Research*, *40*(15), e119. <https://doi.org/10.1093/nar/gks701>
- McArthur, E., & Capra, J. A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *The American Journal of Human Genetics*, *108*(2), 269–283. <https://doi.org/10.1016/j.ajhg.2021.01.001>
- McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell*, *77*(4), 688–708. <https://doi.org/10.1016/j.molcel.2019.12.021>
- McKnight, S. L., & Miller, J., O. L. (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of *Drosophila melanogaster*. *Cell*, *8*(2), 305–19. [https://doi.org/10.1016/0092-8674\(76\)90014-3](https://doi.org/10.1016/0092-8674(76)90014-3)
- Mercola, M., Wang, X. F., Olsen, J., & Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science*, *221*(4611), 663–5. <https://doi.org/10.1126/science.6306772>
- Michaelis, C., Ciosk, R., & Nasmyth, K. (1997). Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, *91*(1), 35–45. [https://doi.org/10.1016/s0092-8674\(01\)80007-6](https://doi.org/10.1016/s0092-8674(01)80007-6)

- Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*(6), 598–606. <https://doi.org/10.1038/ng.3286>
- Minnoye, L., Taskiran, I., Mauduit, D., Fazio, M., Van Aerschot, L., Hulselmans, G., Christiaens, V., Makhzami, S., Seltenhammer, M., Karras, P., Primot, A., Cadieu, E., van Rooijen, E., Marine, J. C., Egidy, G., Ghanem, G. E., Zon, L., Wouters, J., & Aerts, S. (2020). Cross-species analysis of enhancer logic using deep learning. *Genome Research*, *30*(12), 1815–1834. <https://doi.org/10.1101/gr.260844.120>
- Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M. B., & Rijli, F. M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science*, *355*(6332). <https://doi.org/10.1126/science.aal2913>
- Mirny, L. A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(52), 22534–9. <https://doi.org/10.1073/pnas.0913805107>
- Mitter, M., Gasser, C., Takacs, Z., Langer, C. C. H., Tang, W., Jessberger, G., Beales, C. T., Neuner, E., Ameres, S. L., Peters, J. M., Goloborodko, A., Micura, R., & Gerlich, D. W. (2020). Conformation of sister chromatids in the replicated human genome. *Nature*, *586*(7827), 139–144. <https://doi.org/10.1038/s41586-020-2744-4>
- Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J. M., Taneja, N., Folco, H. D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., & Grewal, S. I. S. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature*, *516*(7531), 432–435. <https://doi.org/10.1038/nature13833>
- Monahan, K., Horta, A., & Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature*, *565*(7740), 448–453. <https://doi.org/10.1038/s41586-018-0845-0>
- Mönnich, M., Banks, S., Eccles, M., Dickinson, E., & Horsfield, J. (2009). Expression of cohesin and condensin genes during zebrafish development supports a non-proliferative role for cohesin. *Gene Expression Patterns*, *9*(8), 586–94. <https://doi.org/10.1016/j.gep.2009.08.004>
- Moon, A. M., & Ley, T. J. (1990). Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(19), 7693–7. <https://doi.org/10.1073/pnas.87.19.7693>
- Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., & Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Research*, *9*(22), 6047–68. <https://doi.org/10.1093/nar/9.22.6047>
- Mukund, A. X., Tycko, J., Allen, S. J., Robinson, S. A., Andrews, C., Sinha, J., Ludwig, C. H., Spees, K., Bassik, M. C., & Bintu, L. (2023). High-throughput functional characterization of combinations of transcriptional activators and repressors. *Cell Systems*, *14*(9), 746–763.e5. <https://doi.org/10.1016/j.cels.2023.07.001>
- Müller, T. G., Zila, V., Peters, K., Schifferdecker, S., Stanic, M., Lucic, B., Laketa, V., Lusic, M., Müller, B., & Kräusslich, H. G. (2021). HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *Elife*, *10*. <https://doi.org/10.7554/eLife.64776>
- Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., Caron, A. W., Pilote, A., & Massie, B. (2006). The cumate gene-switch: A system for regulated expression in mammalian cells. *BMC Biotechnology*, *6*, 43. <https://doi.org/10.1186/1472-6750-6-43>
- Murayama, Y., & Uhlmann, F. (2015). DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*, *163*(7), 1628–40. <https://doi.org/10.1016/j.cell.2015.11.030>
- Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A. L., Erb, M. A., Lawlor, M. A., Souza, A., Scott, T. G., Vittori, S., Perry, J. A., Qi, J., Winter, G. E., Wong, K. K., Gray, N. S., & Bradner, J. A. (2020). A genome-wide screen identifies a set of genes that are essential for the establishment and maintenance of Polycomb target gene silencing. *Cell*, *181*(2), 407–421.e17. <https://doi.org/10.1016/j.cell.2020.02.020>

REFERENCES

- J. E. (2018). The dTAG system for immediate and target-specific protein degradation. *Nature Chemical Biology*, *14*(5), 431–441. <https://doi.org/10.1038/s41589-018-0021-8>
- Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. <https://doi.org/10.1038/nature12593>
- Nagasaka, K., Davidson, I. F., Stocsits, R. R., Tang, W., Wutz, G., Batty, P., Panarotto, M., Litos, G., Schleiffer, A., Gerlich, D. W., & Peters, J. M. (2023). Cohesin mediates DNA loop extrusion and sister chromatid cohesion by distinct mechanisms. *Molecular Cell*, *83*(17), 3049–3063.e6. <https://doi.org/10.1016/j.molcel.2023.07.024>
- Naqvi, S., Kim, S., Hoskens, H., Matthews, H. S., Spritz, R. A., Klein, O. D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J. K., & Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nature Genetics*. <https://doi.org/10.1038/s41588-023-01366-2>
- Nasmyth, K. (2001). Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annual Reviews of Genetics*, *35*, 673–745. <https://doi.org/10.1146/annurev.genet.35.102401.091334>
- Neumayr, C., Haberle, V., Serebreni, L., Karner, K., Hendy, O., Boija, A., Henninger, J. E., Li, C. H., Stejskal, K., Lin, G., Bergauer, K., Pagani, M., Rath, M., Mechtler, K., Arnold, C. D., & Stark, A. (2022). Differential cofactor dependencies define distinct types of human enhancers. *Nature*, *606*(7913), 406–413. <https://doi.org/10.1038/s41586-022-04779-x>
- Nguyen, H. Q., Chatteraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., Hershberg, E. A., Martins, N. M. C., Reginato, P. L., Hannan, M., Beliveau, B. J., Church, G. M., Daugharthy, E. R., Marti-Renom, M. A., & Wu, C. T. (2020). 3D mapping and accelerated super-resolution imaging of the human genome using *in situ* sequencing. *Nature Methods*, *17*(8), 822–832. <https://doi.org/10.1038/s41592-020-0890-0>
- Nichols, M. H., & Corces, V. G. (2018). A tethered-inchworm model of SMC DNA translocation. *Nature Structural & Molecular Biology*, *25*(10), 906–910. <https://doi.org/10.1038/s41594-018-0135-4>
- Niki, H., Jaffé, A., Imamura, R., Ogura, T., & Hiraga, S. (1991). The new gene mukB codes for a 177 kD protein with coiled-coil domains involved in chromosome partitioning of *E. coli*. *EMBO Journal*, *10*(1), 183–93. <https://doi.org/10.1002/j.1460-2075.1991.tb07935.x>
- Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., Lee, S. D., Nguyen, S. C., McCole, R. B., Chatteraj, S., Erceg, J., AlHajj Abed, J., Martins, N. M. C., Nguyen, H. Q., Hannan, M. A., Russell, S., Durand, N. C., Rao, S. S. P., Kishi, J. Y., ... Wu, C. T. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLOS Genetics*, *14*(12), e1007872. <https://doi.org/10.1371/journal.pgen.1007872>
- Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A. A., Mechtler, K., & Peters, J. M. (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell*, *143*(5), 737–49. <https://doi.org/10.1016/j.cell.2010.10.031>
- Nora, E. P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A. L., Coulon, A., Mirny, L. A., Pollard, K. S., Dahan, M., & Bruneau, B. G. (2020). Molecular basis of CTCF binding polarity in genome folding. *Nature Communications*, *11*(1), 5612. <https://doi.org/10.1038/s41467-020-19283-x>
- Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, *169*(5), 930–944.e22. <https://doi.org/10.1016/j.cell.2017.05.004>
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial

- partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), 381–5. <https://doi.org/10.1038/nature11049>
- Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W., & Mostafavi, S. (2023). Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Review Genetics*, 24(2), 125–137. <https://doi.org/10.1038/s41576-022-00532-2>
- Noviello, G., Gjaltema, R. A. F., & Schulz, E. G. (2023). CasTuner is a degron and CRISPR/Cas-based toolkit for analog tuning of endogenous gene expression. *Nature Communications*, 14(1), 3225. <https://doi.org/10.1038/s41467-023-38909-4>
- Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., Joti, Y., Tomita, M., Hibino, K., Kanemaki, M. T., Wendt, K. S., Okada, Y., Nagai, T., & Maeshima, K. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Molecular Cell*, 67(2), 282–293.e7. <https://doi.org/10.1016/j.molcel.2017.06.018>
- Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., & Mirny, L. A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), E6697–e6706. <https://doi.org/10.1073/pnas.1717730115>
- Oldenkamp, R., & Rowland, B. D. (2022). A walk through the SMC cycle: From catching DNAs to shaping the genome. *Molecular Cell*, 82(9), 1616–1630. <https://doi.org/10.1016/j.molcel.2022.04.006>
- Olins, A. L., & Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, 183(4122), 330–2. <https://doi.org/10.1126/science.183.4122.330>
- Oliveira, R. A., Hamilton, R. S., Pauli, A., Davis, I., & Nasmyth, K. (2010). Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. *Nature Cell Biology*, 12(2), 185–92. <https://doi.org/10.1038/ncb2018>
- Ong, C. T., & Corces, V. G. (2014). CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*, 15(4), 234–46. <https://doi.org/10.1038/nrg3663>
- Ono, T., Fang, Y., Spector, D. L., & Hirano, T. (2004). Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Molecular Biology of the Cell*, 15(7), 3296–308. <https://doi.org/10.1091/mbc.e04-03-0242>
- Ono, T., Losada, A., Hirano, M., Myers, M. P., Neuwald, A. F., & Hirano, T. (2003). Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell*, 115(1), 109–21. [https://doi.org/10.1016/s0092-8674\(03\)00724-4](https://doi.org/10.1016/s0092-8674(03)00724-4)
- Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., & Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics*, 36(10), 1065–71. <https://doi.org/10.1038/ng1423>
- Ostersehl, L. M., Jans, D. C., Wittek, A., Keller-Findeisen, J., Inamdar, K., Sahl, S. J., Hell, S. W., & Jakobs, S. (2022). DNA-PAINT MINFLUX nanoscopy. *Nature Methods*, 19(9), 1072–1075. <https://doi.org/10.1038/s41592-022-01577-1>
- Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., Kato, M., Garvin, T. H., Pham, Q. T., Harrington, A. N., Akiyama, J. A., Afzal, V., Lopez-Rios, J., Dickel, D. E., Visel, A., & Pennacchio, L. A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554(7691), 239–243. <https://doi.org/10.1038/nature25461>
- Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., & O’Shea, C. C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349). <https://doi.org/10.1126/science.aag0025>
- Paulson, J. R., & Laemmli, U. K. (1977). The structure of histone-depleted metaphase chromosomes. *Cell*, 12(3),

REFERENCES

- 817–28. [https://doi.org/10.1016/0092-8674\(77\)90280-x](https://doi.org/10.1016/0092-8674(77)90280-x)
- Peng, Q., Huang, Z., Sun, K., Liu, Y., Yoon, C. W., Harrison, R. E. S., Schmitt, D. L., Zhu, L., Wu, Y., Tasan, I., Zhao, H., Zhang, J., Zhong, S., Chien, S., & Wang, Y. (2022). Engineering inducible biomolecular assemblies for genome imaging and manipulation in living cells. *Nature Communications*, *13*(1), 7933. <https://doi.org/10.1038/s41467-022-35504-x>
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & Rubin, E. M. (2006). *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499–502. <https://doi.org/10.1038/nature05295>
- Petela, N. J., Gligoris, T. G., Metson, J., Lee, B. G., Voulgaris, M., Hu, B., Kikuchi, S., Chapard, C., Chen, W., Rajendra, E., Srinivisan, M., Yu, H., Löwe, J., & Nasmyth, K. A. (2018). Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Molecular Cell*, *70*(6), 1134–1148.e7. <https://doi.org/10.1016/j.molcel.2018.05.022>
- Philipsen, S., Talbot, D., Fraser, P., & Grosveld, F. (1990). The beta-globin dominant control region: Hypersensitive site 2. *EMBO Journal*, *9*(7), 2159–67. <https://doi.org/10.1002/j.1460-2075.1990.tb07385.x>
- Phillips, J. E., & Corces, V. G. (2009). CTCF: Master weaver of the genome. *Cell*, *137*(7), 1194–211. <https://doi.org/10.1016/j.cell.2009.06.001>
- Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., Ong, C. T., Hookway, T. A., Guo, C., Sun, Y., Bland, M. J., Wagstaff, W., Dalton, S., McDevitt, T. C., Sen, R., Dekker, J., Taylor, J., & Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, *153*(6), 1281–95. <https://doi.org/10.1016/j.cell.2013.04.053>
- Plachta, N., Bollenbach, T., Pease, S., Fraser, S. E., & Pantazis, P. (2011). Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nature Cell Biology*, *13*(2), 117–23. <https://doi.org/10.1038/ncb2154>
- Platania, A., Erb, C., Barbieri, M., Molcette, B., Grandgirard, E., Kort, M. A. d., Meaburn, K., Taylor, T., Shchuka, V. M., Kocanova, S., Oliveira, G. M., Mitchell, J. A., Soutoglou, E., Lenstra, T. L., Molina, N., Papantonis, A., Bystricky, K., & Sexton, T. (2023). Competition between transcription and loop extrusion modulates promoter and enhancer dynamics. *bioRxiv*, 2023.04.25.538222. <https://doi.org/10.1101/2023.04.25.538222>
- Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V., & Hackett, J. A. (2022). Systematic Epigenome Editing Captures the Context-dependent Instructive Function of Chromatin Modifications. *bioRxiv*, 2022.09.04.506519. <https://doi.org/10.1101/2022.09.04.506519>
- Pollex, T., & Heard, E. (2019). Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation. *Nature Genetics*, *51*(2), 285–295. <https://doi.org/10.1038/s41588-018-0305-7>
- Popay, T. M., & Dixon, J. R. (2022). Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. *Journal of Biological Chemistry*, *298*(8), 102117. <https://doi.org/10.1016/j.jbc.2022.102117>
- Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., ... Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature*, *515*(7527), 402–5. <https://doi.org/10.1038/nature13986>
- Popp, A. P., Hettich, J., & Gebhardt, J. C. M. (2021). Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Research*, *49*(11), 6249–6266. <https://doi.org/10.1093/nar/gkab443>
- Pradhan, B., Barth, R., Kim, E., Davidson, I. F., Bauer, B., van Laar, T., Yang, W., Ryu, J. K., van der Torre, J.,

- Peters, J. M., & Dekker, C. (2022). SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports*, *41*(3), 111491. <https://doi.org/10.1016/j.celrep.2022.111491>
- Pugacheva, E. M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A. L., Strunnikov, A. V., Zentner, G. E., Ren, B., & Lobanenko, V. V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 2020–2031. <https://doi.org/10.1073/pnas.1911708117>
- Quililan, K., Oberbeckmann, E., Cramer, P., & Oudelaar, A. M. (2023). *In vitro* reconstitution of chromatin domains. *bioRxiv*, 2023.02.27.530214. <https://doi.org/10.1101/2023.02.27.530214>
- Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M., Shishkin, A., Bhat, P., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, *174*(3), 744–757.e24. <https://doi.org/10.1016/j.cell.2018.05.024>
- Ramasamy, S., Aljahani, A., Karpinska, M. A., Cao, T. B. N., Velychko, T., Cruz, J. N., Lidschreiber, M., & Oudelaar, A. M. (2023). The Mediator complex regulates enhancer-promoter interactions. *Nature Structural & Molecular Biology*, *30*(7), 991–1000. <https://doi.org/10.1038/s41594-023-01027-2>
- Rankin, S., Ayad, N. G., & Kirschner, M. W. (2005). Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. *Molecular Cell*, *18*(2), 185–200. <https://doi.org/10.1016/j.molcel.2005.03.017>
- Rao, S., Ahmad, K., & Ramachandran, S. (2021). Cooperative binding between distant transcription factors is a hallmark of active enhancers. *Molecular Cell*, *81*(8), 1651–1665.e4. <https://doi.org/10.1016/j.molcel.2021.02.014>
- Rao, S., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., ... Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305–320.e24. <https://doi.org/10.1016/j.cell.2017.09.026>
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–80. <https://doi.org/10.1016/j.cell.2014.11.021>
- Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., Pollex, T., Grand, R. S., Mulugeta, E., Kind, J., Tiana, G., Smallwood, S. A., de Laat, W., & Giorgetti, L. (2019). DamC reveals principles of chromatin folding *in vivo* without crosslinking and ligation. *Nature Structural & Molecular Biology*, *26*(6), 471–480. <https://doi.org/10.1038/s41594-019-0231-0>
- Reitman, M., & Felsenfeld, G. (1990). Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. *Molecular and Cellular Biology*, *10*(6), 2774–86. <https://doi.org/10.1128/mcb.10.6.2774-2786.1990>
- Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., & Cramer, P. (2021). Structure of the human Mediator-RNA polymerase II pre-initiation complex. *Nature*, *594*(7861), 129–133. <https://doi.org/10.1038/s41586-021-03555-7>
- Rhodes, J., Mazza, D., Nasmyth, K., & Uphoff, S. (2017). Scc2/Nipbl hops between chromosomal cohesin rings after loading. *Elife*, *6*. <https://doi.org/10.7554/eLife.30000>
- Richter, W. F., Nayak, S., Iwasa, J., & Taatjes, D. J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *23*(11), 732–749. <https://doi.org/10.1038/s41580-022-00498-3>
- Rickels, R., Herz, H. M., Sze, C. C., Cao, K., Morgan, M. A., Collings, C. K., Gause, M., Takahashi, Y. H., Wang, L., Rendleman, E. J., Marshall, S. A., Krueger, A., Bartom, E. T., Piunti, A., Smith, E. R., Abshiru, N. A.,

REFERENCES

- Kelleher, N. L., Dorsett, D., & Shilatifard, A. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nature Genetics*, *49*(11), 1647–1653. <https://doi.org/10.1038/ng.3965>
- Riedel, C. G., Katis, V. L., Katou, Y., Mori, S., Itoh, T., Helmhart, W., Gálová, M., Petronczki, M., Gregan, J., Cetin, B., Mudrak, I., Ogris, E., Mechtler, K., Pelletier, L., Buchholz, F., Shirahige, K., & Nasmyth, K. (2006). Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature*, *441*(7089), 53–61. <https://doi.org/10.1038/nature04664>
- Riggs, A. D. (1990). DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. *Philosophical Transactions of the Royal Society B*, *326*(1235), 285–97. <https://doi.org/10.1098/rstb.1990.0012>
- Rinaldi, L., Fettweis, G., Kim, S., Garcia, D. A., Fujiwara, S., Johnson, T. A., Tettey, T. T., Ozbun, L., Pegoraro, G., Puglia, M., Blagoev, B., Upadhyaya, A., Stavreva, D. A., & Hager, G. L. (2022). The glucocorticoid receptor associates with the cohesin loader NIPBL to promote long-range gene regulation. *Science Advances*, *8*(13), eabj8360. <https://doi.org/10.1126/sciadv.abj8360>
- Rinzema, N. J., Sofiadis, K., Tjalsma, S. J. D., Versteegen, M., Oz, Y., Valdes-Quezada, C., Felder, A. K., Filipovska, T., van der Elst, S., de Andrade Dos Ramos, Z., Han, R., Krijger, P. H. L., & de Laat, W. (2022). Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nature Structural & Molecular Biology*, *29*(6), 563–574. <https://doi.org/10.1038/s41594-022-00787-7>
- Roayaei Ardakany, A., Gezer, H. T., Lonardi, S., & Ay, F. (2020). Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biology*, *21*(1), 256. <https://doi.org/10.1186/s13059-020-02167-0>
- Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., & Belmont, A. S. (1996). *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *Journal of Cell Biology*, *135*(6 Pt 2), 1685–700. <https://doi.org/10.1083/jcb.135.6.1685>
- Roccio, M., Schmitter, D., Knobloch, M., Okawa, Y., Sage, D., & Lutolf, M. P. (2013). Predicting stem cell fate changes by differential cell cycle progression patterns. *Development*, *140*(2), 459–70. <https://doi.org/10.1242/dev.086215>
- Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., & Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science*, *321*(5888), 563–6. <https://doi.org/10.1126/science.1157774>
- Rosa, A., Becker, N. B., & Everaers, R. (2010). Looping probabilities in model interphase chromosomes. *Biophysical Journal*, *98*(11), 2410–9. <https://doi.org/10.1016/j.bpj.2010.01.054>
- Rosa, A., & Everaers, R. (2008). Structure and dynamics of interphase chromosomes. *PLOS Computational Biology*, *4*(8), e1000153. <https://doi.org/10.1371/journal.pcbi.1000153>
- Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A., & Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(24), 8309–14. <https://doi.org/10.1073/pnas.0801273105>
- Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics*, *43*(4), 379–86. <https://doi.org/10.1038/ng.790>
- Rust, M. J., Bates, M., & Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, *3*(10), 793–5. <https://doi.org/10.1038/nmeth929>
- Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., & Bystricky, K. (2014). DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLOS Genetics*,

- 10(3), e1004187. <https://doi.org/10.1371/journal.pgen.1004187>
- Sabari, B. R., Dall'Agnesse, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I., ... Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, *361*(6400). <https://doi.org/10.1126/science.aar3958>
- Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C. O., Cramer, P., Kivioja, T., & Taipale, J. (2022). Sequence determinants of human gene regulatory elements. *Nature Genetics*, *54*(3), 283–294. <https://doi.org/10.1038/s41588-021-01009-4>
- Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., Imamura, T., Ogawa, M., Masai, H., & Miyawaki, A. (2008). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell*, *132*(3), 487–98. <https://doi.org/10.1016/j.cell.2007.12.033>
- Salari, H., Di Stefano, M., & Jost, D. (2022). Spatial organization of chromosomes leads to heterogeneous chromatin motion and drives the liquid- or gel-like dynamical behavior of chromatin. *Genome Research*, *32*(1), 28–43. <https://doi.org/10.1101/gr.275827.121>
- Salic, A., Waters, J. C., & Mitchison, T. J. (2004). Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. *Cell*, *118*(5), 567–78. <https://doi.org/10.1016/j.cell.2004.08.016>
- Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., & Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6456–65. <https://doi.org/10.1073/pnas.1518552112>
- Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell Detection with Star-Convex Polygons. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*, 265–273. https://doi.org/10.1007/978-3-030-00934-2_30
- Schmitz, J., Watrin, E., Lénárt, P., Mechtler, K., & Peters, J. M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Current Biology*, *17*(7), 630–6. <https://doi.org/10.1016/j.cub.2007.02.029>
- Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews Genetics*, *20*(8), 437–455. <https://doi.org/10.1038/s41576-019-0128-0>
- Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S. W., Dimitrova, E., Dimond, A., Edelman, L. B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., ... Fraser, P. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research*, *25*(4), 582–97. <https://doi.org/10.1101/gr.185272.114>
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C. H., Mirny, L., & Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, *551*(7678), 51–56. <https://doi.org/10.1038/nature24281>
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*, 259. <https://doi.org/10.1186/s13059-015-0831-x>
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, *148*(3), 458–72. <https://doi.org/10.1016/j.cell.2012.01.010>

REFERENCES

- Shaban, H. A., Barth, R., & Bystricky, K. (2020). Navigating the crowd: Visualizing coordination between genome dynamics, structure, and transcription. *Genome Biology*, *21*(1), 278. <https://doi.org/10.1186/s13059-020-02185-y>
- Shaltiel, I. A., Datta, S., Lecomte, L., Hassler, M., Kschonsak, M., Bravo, S., Stober, C., Ormanns, J., Eustermann, S., & Haering, C. H. (2022). A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. *Science*, *376*(6597), 1087–1094. <https://doi.org/10.1126/science.abm4012>
- Shechner, D. M., Hacisuleyman, E., Younger, S. T., & Rinn, J. L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nature Methods*, *12*(7), 664–70. <https://doi.org/10.1038/nmeth.3433>
- Shi, Z., Gao, H., Bai, X. C., & Yu, H. (2020). Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science*, *368*(6498), 1454–1459. <https://doi.org/10.1126/science.abb0981>
- Shintomi, K., & Hirano, T. (2011). The relative ratio of condensin I to II determines chromosome shapes. *Genes & Development*, *25*(14), 1464–9. <https://doi.org/10.1101/gad.2060311>
- Shrinivas, K., Sabari, B. R., Coffey, E. L., Klein, I. A., Boija, A., Zamudio, A. V., Schuijers, J., Hannett, N. M., Sharp, P. A., Young, R. A., & Chakraborty, A. K. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. *Molecular Cell*, *75*(3), 549–561.e7. <https://doi.org/10.1016/j.molcel.2019.07.009>
- Simon, J. A., & Kingston, R. E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Molecular Cell*, *49*(5), 808–24. <https://doi.org/10.1016/j.molcel.2013.02.013>
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., & de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, *38*(11), 1348–54. <https://doi.org/10.1038/ng1896>
- Soh, Y. M., Bürmann, F., Shin, H. C., Oda, T., Jin, K. S., Toseland, C. P., Kim, C., Lee, H., Kim, S. J., Kong, M. S., Durand-Diebold, M. L., Kim, Y. G., Kim, H. M., Lee, N. K., Sato, M., Oh, B. H., & Gruber, S. (2015). Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Molecular Cell*, *57*(2), 290–303. <https://doi.org/10.1016/j.molcel.2014.11.023>
- Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., & Krebs, A. R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity *In Vivo*. *Molecular Cell*, *81*(2), 255–267.e6. <https://doi.org/10.1016/j.molcel.2020.11.015>
- Spielmann, M., Lupiáñez, D. G., & Mundlos, S. (2018). Structural variation in the 3D genome. *Nature Reviews Genetics*, *19*(7), 453–467. <https://doi.org/10.1038/s41576-018-0007-0>
- Splinter, E., de Wit, E., van de Werken, H. J., Klous, P., & de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods*, *58*(3), 221–30. <https://doi.org/10.1016/j.ymeth.2012.04.009>
- Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N., & de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development*, *20*(17), 2349–54. <https://doi.org/10.1101/gad.399506>
- Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S., & Stamatoyannopoulos, J. A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science*, *368*(6498), 1449–1454. <https://doi.org/10.1126/science.aaz1646>
- Strick, T. R., Kawaguchi, T., & Hirano, T. (2004). Real-time detection of single-molecule DNA compaction by condensin I. *Current Biology*, *14*(10), 874–80. <https://doi.org/10.1016/j.cub.2004.04.038>
- Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, *18*(1), 100–106. <https://doi.org/10.1038/s41592-020-01018-x>
- Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, *547*(7662), 241–245. <https://doi.org/10.1038/na>

ture22989

- Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B. H., & Peters, J. M. (2000). Characterization of vertebrate cohesin complexes and their regulation in prophase. *Journal of Cell Biology*, *151*(4), 749–62. <https://doi.org/10.1083/jcb.151.4.749>
- Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., & Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, *24*(3), 390–400. <https://doi.org/10.1101/gr.163519.113>
- Szabo, Q., Bantignies, F., & Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Science Advances*, *5*(4), eaaw1668. <https://doi.org/10.1126/sciadv.aaw1668>
- Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., Nora, E. P., Bruneau, B. G., Bantignies, F., & Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*, *52*(11), 1151–1157. <https://doi.org/10.1038/s41588-020-00716-8>
- Talbot, D., Philipson, S., Fraser, P., & Grosveld, F. (1990). Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO Journal*, *9*(7), 2169–77. <https://doi.org/10.1002/j.1460-2075.1990.tb07386.x>
- Tamm, M. V., & Polovnikov, K. (2017). Dynamics of Polymers: Classic Results and Recent Developments. In *Order, disorder and criticality* (pp. 113–172). WORLD SCIENTIFIC. [https://doi.org/doi:10.1142/9789813232105_0003](https://doi.org/doi:10.1142/9789813232105_000310.1142/9789813232105_0003)
- Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., & Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, *159*(3), 635–46. <https://doi.org/10.1016/j.cell.2014.09.039>
- Taskiran, I. I., Spanier, K. I., Christiaens, V., Mauduit, D., & Aerts, S. (2022). Cell type directed design of synthetic enhancers. *bioRxiv*, 2022.07.26.501466. <https://doi.org/10.1101/2022.07.26.501466>
- Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H. N., Huynh, T. N., Zhen, C. Y., Ma, B., Wang, H., & Ren, X. (2019). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *Journal of Biological Chemistry*, *294*(5), 1451–1463. <https://doi.org/10.1074/jbc.RA118.006620>
- Tebo, A. G., & Gautier, A. (2019). A split fluorescent reporter with rapid and reversible complementation. *Nature Communications*, *10*(1), 2822. <https://doi.org/10.1038/s41467-019-10855-0>
- Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I. F., Tang, W., Cisneros, D. A., Bhaskara, V., Nishiyama, T., Vaziri, A., Wutz, A., Ellenberg, J., & Peters, J. M. (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*, *501*(7468), 564–8. <https://doi.org/10.1038/nature12471>
- Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., in 't Veld, P. J., Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R., Stevens, M. J., Tranchida, J., Trott, C., & Plimpton, S. J. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, *271*, 108171. <https://doi.org/https://doi.org/10.1016/j.cpc.2021.108171>
- Thompson, R. E., Larson, D. R., & Webb, W. W. (2002). Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, *82*(5), 2775–83. [https://doi.org/10.1016/s0006-3495\(02\)75618-x](https://doi.org/10.1016/s0006-3495(02)75618-x)
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., ... Stamatooyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. <https://doi.org/10.1038/nature11232>
- Tkačik, G., & Gregor, T. (2021). The many bits of positional information. *Development*, *148*(2). <https://doi.org/>

REFERENCES

- 10.1242/dev.176065
- Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods*, *5*(2), 159–61. <https://doi.org/10.1038/nmeth1171>
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*, *10*(6), 1453–65. [https://doi.org/10.1016/s1097-2765\(02\)00781-5](https://doi.org/10.1016/s1097-2765(02)00781-5)
- Tuan, D., Solomon, W., Li, Q., & London, I. M. (1985). The “beta-like-globin” gene domain in human erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(19), 6384–8. <https://doi.org/10.1073/pnas.82.19.6384>
- Tycko, J., DelRosso, N., Hess, G. T., Aradhana, Banerjee, A., Mukund, A., Van, M. V., Ego, B. K., Yao, D., Spees, K., Suzuki, P., Marinov, G. K., Kundaje, A., Bassik, M. C., & Bintu, L. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell*, *183*(7), 2020–2035.e16. <https://doi.org/10.1016/j.cell.2020.11.024>
- Uhlmann, F., Lottspeich, F., & Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature*, *400*(6739), 37–42. <https://doi.org/10.1038/21831>
- Unal, E., Heidinger-Pauli, J. M., Kim, W., Guacci, V., Onn, I., Gygi, S. P., & Koshland, D. E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. *Science*, *321*(5888), 566–9. <https://doi.org/10.1126/science.1157880>
- Valton, A. L., Venev, S. V., Mair, B., Khokhar, E. S., Tong, A. H. Y., Usaj, M., Chan, K., Pai, A. A., Moffat, J., & Dekker, J. (2022). A cohesin traffic pattern genetically linked to gene regulation. *Nature Structural & Molecular Biology*, *29*(12), 1239–1251. <https://doi.org/10.1038/s41594-022-00890-9>
- van den Engh, G., Sachs, R., & Trask, B. J. (1992). Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science*, *257*(5075), 1410–2. <https://doi.org/10.1126/science.1388286>
- Van der Ploeg, L. H., Konings, A., Oort, M., Roos, D., Bernini, L., & Flavell, R. A. (1980). Gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature*, *283*(5748), 637–42. <https://doi.org/10.1038/283637a0>
- Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in Python. *PeerJ*, *2*, e453. <https://doi.org/10.7717/peerj.453>
- van Bommel, J. G., Galupa, R., Gard, C., Servant, N., Picard, C., Davies, J., Szempruch, A. J., Zhan, Y., Żylicz, J. J., Nora, E. P., Lameiras, S., de Wit, E., Gentien, D., Baulande, S., Giorgetti, L., Guttman, M., Hughes, J. R., Higgs, D. R., Gribnau, J., & Heard, E. (2019). The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of *Tsix* and *Xist*. *Nature Genetics*, *51*(6), 1024–1034. <https://doi.org/10.1038/s41588-019-0412-0>
- van Ruiten, M. S., van Gent, D., Sedeño Cacciatore, Á., Fauster, A., Willems, L., Hekkelman, M. L., Hoekman, L., Altelaar, M., Haarhuis, J. H. I., Brummelkamp, T. R., de Wit, E., & Rowland, B. D. (2022). The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. *Nature Structural & Molecular Biology*, *29*(6), 586–591. <https://doi.org/10.1038/s41594-022-00773-z>
- van Staalduinen, J., van Staveren, T., Grosveld, F., & Wendt, K. S. (2023). Live-cell imaging of chromatin contacts opens a new window into chromatin dynamics. *Epigenetics & Chromatin*, *16*(1), 27. <https://doi.org/10.1186/s13072-023-00503-9>
- Vazquez Nunez, R., Polyhach, Y., Soh, Y. M., Jeschke, G., & Gruber, S. (2021). Gradual opening of SMC arms in prokaryotic condensin. *Cell Reports*, *35*(4), 109051. <https://doi.org/10.1016/j.celrep.2021.109051>
- Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., Jabs, E. W., Inui, K., & Joenje, H. (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nature Genetics*, *37*(5), 468–70. <https://doi.org/10.1038/ng1548>

- Vettorel, T., Grosberg, A. Y., & Kremer, K. (2009). Statistics of polymer rings in the melt: A numerical simulation study. *Physical Biology*, 6(2), 025013. <https://doi.org/10.1088/1478-3975/6/2/025013>
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), 854–8. <https://doi.org/10.1038/nature07730>
- Walther, N., Hossain, M. J., Politi, A. Z., Koch, B., Kueblbeck, M., Ødegård-Fougner, Ø., Lampe, M., & Ellenberg, J. (2018). A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. *Journal of Cell Biology*, 217(7), 2309–2328. <https://doi.org/10.1083/jcb.201801048>
- Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., Rando, O. J., Birney, E., Myers, R. M., Noble, W. S., Snyder, M., & Weng, Z. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research*, 22(9), 1798–812. <https://doi.org/10.1101/gr.139105.112>
- Wang, X., Hughes, A. C., Brandão, H. B., Walker, B., Lierz, C., Cochran, J. C., Oakley, M. G., Kruse, A. C., & Rudner, D. Z. (2018). *In Vivo* Evidence for ATPase-Dependent DNA Translocation by the *Bacillus subtilis* SMC Condensin Complex. *Molecular Cell*, 71(5), 841–847.e5. <https://doi.org/10.1016/j.molcel.2018.07.006>
- Weigert, M., Schmidt, U., Haase, R., Sugawara, K., & Myers, G. (2020). Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. *The IEEE Winter Conference on Applications of Computer Vision (WACV)*. <https://doi.org/10.1109/WACV45572.2020.9093435>
- Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, 171(7), 1573–1588.e28. <https://doi.org/10.1016/j.cell.2017.11.008>
- Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., & Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451(7180), 796–801. <https://doi.org/10.1038/nature06634>
- Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K. D., & Nienhaus, G. U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), 15905–10. <https://doi.org/10.1073/pnas.0403668101>
- Wilhelm, L., Bürmann, F., Minnen, A., Shin, H. C., Toseland, C. P., Oh, B. H., & Gruber, S. (2015). SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in *Bacillus subtilis*. *Elife*, 4. <https://doi.org/10.7554/eLife.06659>
- Wolff, J. O., Scheiderer, L., Engelhardt, T., Engelhardt, J., Matthias, J., & Hell, S. W. (2023). MINFLUX dissects the unimpeded walking of kinesin-1. *Science*, 379(6636), 1004–1010. <https://doi.org/10.1126/science.ade2650>
- Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., Fabre, E., & Zimmer, C. (2012). A predictive computational model of the dynamic 3D interphase yeast nucleus. *Current Biology*, 22(20), 1881–90. <https://doi.org/10.1016/j.cub.2012.07.069>
- Wutz, G., Ladurner, R., St Hilaire, B. G., Stocsits, R. R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Várnai, C., Ivanov, M. P., Schoenfelder, S., van der Lelij, P., Huang, X., Dürnberger, G., Roitinger, E., Mechtler, K., Davidson, I. F., Fraser, P., Lieberman-Aiden, E., & Peters, J. M. (2020). ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. *Elife*, 9. <https://doi.org/10.7554/eLife.52091>
- Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M. J., Walther, N., Koch, B., Kueblbeck, M., Ellenberg, J., Zuber, J., Fraser, P.,

REFERENCES

- & Peters, J. M. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO Journal*, *36*(24), 3573–3599. <https://doi.org/10.15252/embj.201798004>
- Xiao, J. Y., Hafner, A., & Boettiger, A. N. (2021). How subtle changes in 3D structure can create large changes in transcription. *Elife*, *10*. <https://doi.org/10.7554/eLife.64320>
- Xu, H., Boone, C., & Brown, G. W. (2007). Genetic dissection of parallel sister-chromatid cohesion pathways. *Genetics*, *176*(3), 1417–29. <https://doi.org/10.1534/genetics.107.072876>
- Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A. P., Lobanenko, V., Klenova, E., & Ohlsson, R. (2004). Poly(ADP-ribosylation) regulates CTCF-dependent chromatin insulation. *Nature Genetics*, *36*(10), 1105–10. <https://doi.org/10.1038/ng1426>
- Zaret, K. S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annual Review of Genomics and Human Genetics*, *54*, 367–385. <https://doi.org/10.1146/annurev-genet-030220-015007>
- Zhan, Y., Mariani, L., Barozzi, I., Schulz, E. G., Blüthgen, N., Stadler, M., Tiana, G., & Giorgetti, L. (2017). 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*(3), 479–490. <https://doi.org/10.1101/gr.212803.116>
- Zhang, H., Shi, Z., Banigan, E. J., Kim, Y., Yu, H., Bai, X. C., & Finkelstein, I. J. (2023). CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Molecular Cell*, *83*(16), 2856–2871.e8. <https://doi.org/10.1016/j.molcel.2023.07.006>
- Zhang, S., Übelmesser, N., Josipovic, N., Forte, G., Slotman, J. A., Chiang, M., Gothe, H. J., Gusmao, E. G., Becker, C., Altmüller, J., Houtsmuller, A. B., Roukos, V., Wendt, K. S., Marenduzzo, D., & Papantonis, A. (2021). RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. *Science Advances*, *7*(43), eabg8205. <https://doi.org/10.1126/sciadv.abg8205>
- Zhang, Y., Zhang, X., Dai, H. Q., Hu, H., & Alt, F. W. (2022). The role of chromatin loop extrusion in antibody diversification. *Nature Reviews Immunology*, *22*(9), 550–566. <https://doi.org/10.1038/s41577-022-00679-3>
- Zhen, C. Y., Tatavosian, R., Huynh, T. N., Duc, H. N., Das, R., Kokotovic, M., Grimm, J. B., Lavis, L. D., Lee, J., Mejia, F. J., Li, Y., Yao, T., & Ren, X. (2016). Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *Elife*, *5*. <https://doi.org/10.7554/eLife.17667>
- Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S. O., Nitta, K. R., Morgunova, E., Taipale, M., Cramer, P., & Taipale, J. (2018). The interaction landscape between transcription factors and the nucleosome. *Nature*, *562*(7725), 76–81. <https://doi.org/10.1038/s41586-018-0549-5>
- Zidovska, A., Weitz, D. A., & Mitchison, T. J. (2013). Micron-scale coherence in interphase chromatin dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15555–60. <https://doi.org/10.1073/pnas.1220313110>
- Zuin, J., Dixon, J. R., van der Reijden, M. I., Ye, Z., Kolovos, P., Brouwer, R. W., van de Corput, M. P., van de Werken, H. J., Knoch, T. A., van, I. W. F., Grosveld, F. G., Ren, B., & Wendt, K. S. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), 996–1001. <https://doi.org/10.1073/pnas.1317788111>
- Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through enhancer-promoter interactions. *Nature*, *604*(7906), 571–577. <https://doi.org/10.1038/s41586-022-04570-y>

Supplementary Information

Chapter 2 – Cohesin and CTCF control the dynamics of chromosome folding

Methods

Generation of targeting vectors for random integration of TetO array and TetR-tdTomato

To generate an 8 kb TetO array within piggyBac inverted terminal repeats (ITRs), the TetO array was obtained from the pSO2.Pac.TetO, a gift from Edith's Heard lab (Masui *et al.*, 2011), by growing bacteria at 37 °C to reduce the size of the original 30 kb operator array by recombining. The array was then excised from the vector by restriction digest with BamHI (NEB, R0136S) and cloned into the PB-empty (Redolfi *et al.*, 2019) (final vector PB-empty-DSE-TetO-8kb). A cassette carrying three strong CTCF sites was excised with XhoI (NEB, R0146S) from PB-empty_DSE_TetO_2.7kb_3xCTCF and ligated into the PB-empty-DSE-TetO-8kb vector using T4 DNA Ligase (NEB, M0202L). Clones were screened by Sanger sequencing (Microsynth) for CTCF sites inserted facing toward the TetO array. The final vector (PB-3xCTCF-TetO) was validated by restriction digest with EcoRI (NEB, R3101L) and NotI (NEB, R3189L) for the correct size of the operator array and Sanger sequencing (Microsynth) for the correct insertion of the CTCF cassette.

To express the TetR and Lac repressor (LacI) fused to a fluorescent protein flanked by ITRs for piggyBac transposition, the PB-empty vector was first linearized by digestion with XhoI (NEB, R0146L) and the TetR-eGFP was amplified with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) from pBroad3-TetR-ICP22-eGFP kindly provided by Tim Pollex (Pollex & Heard, 2019) with Gibson overhangs. PB-Ubc-TetR-eGFP was assembled using Gibson cloning (NEB, E2611L). The tdTomato was amplified with Gibson overhangs and assembled with the digested PB-TetR-eGFP (BamHI, NEB, R3136L and EcoRI, NEB, R3101L) to yield the PB-Ubc-TetR-tdTomato vector. To increase expression levels of the fusion proteins, the Ubc promoter was exchanged for the stronger CAGGS promoter that was amplified with Gibson overhangs from a pCAGGS plasmid. The final PB-TetR-tdTomato was made using Gibson assembly of the amplified CAGGS promoter (from Addgene plasmid #20733) with the PB-Ubc-TetR-tdTomato digested with BglII (NEB, R0144L) and AgeI (NEB, R3552S). To generate PB-CAGGS-TetR-eGFP, PB-Ubc-TetR-eGFP and PB-TetR-tdTomato were

digested with XhoI (NEB, R0146L) and ligated using T4 DNA Ligase (NEB, M0202L). PB-LacI-eGFP was generated by amplification of the LacI with overhangs for subsequent Gibson assembly with the digested PB-TetR-eGFP (AgeI, NEBR3552S). Primers used for cloning can be found in Supplementary Table S3.

Generation of targeting vectors for TetO and LacO

Vector for targeting the TetO array to the genomic locus on chr15:11,647,372: The vector pMK-chr15-Rox-PuroR-Rox containing the homology arms for chromosome 15 as well as the Puromycin resistance gene flanked by Rox sites was custom synthesized by GeneArt Synthesis (Thermo Fisher Scientific) and linearized with SbfI (NEB, R0642S) and SpeI (NEB, R3133S). A short linker sequence including a XhoI restriction site was introduced into the vector by PCR amplification from pMK-chr15-Rox-PuroR-Rox with Gibson overhangs. The XhoI restriction site was then used to ligate the 3xCTCF-TetO (cut from the PB-3xCTCF-TetO vector) into the vector leading to pMK-3xCTCF-TetO-Rox-PuroR-Rox. Vector for targeting the LacO array to the genomic locus on chr15:11,496,908: The vector pUC19-empty was linearized with EcoRI (NEB, R3101L) and BamHI (NEB, R3136L). The left homology arm was amplified from E14 wild-type genomic DNA with overhangs for Gibson assembly. The 5'-ITR was amplified from PB-empty with Gibson overhangs and both PCR products were assembled into the pUC19 vector (Addgene, #50005) to yield pUC19-IHA-5'ITR. pUC19 was digested with KpnI (NEB, R3142L) and BamHI (NEB, R2126L) to be assembled into pUC19-3'ITR-3xCTCF-rHA with the following PCR products with respective Gibson overhangs: The 3'ITR was amplified from PB-empty, the CTCF cassette was amplified from pMK-chr15-Rox-PuroR-Rox and the right homology arm was amplified from E14 wild-type genomic DNA. To make the targeting vector, pUC19-IHA-5'ITR was linearized with EcoRI and BamHI, pUC19-3'ITR-3xCTCF-rHA was linearized with KpnI and HindIII (NEB, R3104L) and the Neomycin resistance gene was amplified from pEN113 (Addgene, #86233). All three parts were assembled using Gibson assembly to pUC19-ITR-NeoR-ITR-3xCTCF. The LacO array was excised from the pLAU43_LacO_plus vector (Lau *et al.*, 2003) with XhoI (NEB, R0146L) and ligated into the linearized targeting vector (cut with XhoI) using T4 DNA Ligase resulting in the final targeting vector pUC19-ITR-NeoR-ITR-3xCTCF-LacO. Primers used for cloning can be found in Supplementary Table S3.

Removal of CTCF sites by Cre recombination

To selectively remove the three CTCF binding sites flanking the operator arrays, 0.5×10^6 cells of each cell line were transfected 1 μ g of pIC-Cre (a gift from the Schübeler lab) using Lipofectamine3000 according to the manufacturer's instructions (Thermo Fisher Scientific, L3000008). 7 days after transfection, the cells were sorted and genotyped as described previously. Primers used in genotyping are listed in Supplementary Table S3.

Generation of control cell lines expressing OsTir1

E14 wild-type cells were transfected with the targeting vector pEN396-pCAGGS-Tir1-V5-2A-PuroR TIGRE donor and the gRNA vector pX330-EN1201 (Addgene #92144 and #92142) using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza,

V4XP-3024 KT). 2×10^6 cells were harvested using accutase (Sigma Aldrich, A6964) and resuspended in 100 μ l transfection solution (82 μ l primary solution, 18 μ l supplement, 15 μ g Tir1 targeting vector and 5 μ g of pX330-EN1201) and transferred to a single Nucleocuvette (Lonza). Nucleofection was performed using the protocol CG110. Transfected cells were directly seeded in pre-warmed E14 standard medium. 48 hours after transfection, 1 μ g/ml of puromycin (InvivoGen, ant-pr-1) was added to the medium for 3 days to select cells for insertion of the Tir1 integration. Cells were sorted and genotyped as described previously. Primers used in genotyping are listed in Supplementary Table S3.

Tagging of endogenous *Rad21* locus and removal of resistance genes *NeoR* and *HygroR*

The gRNA vector and the targeting vector from N. Q. Liu *et al.*, 2021 were purchased from Addgene (gRNA pX330-EN1082: #156450, targeting vector pEN313: #156431). The gRNA sequence can be found in Supplementary Table S3. The targeting vector was modified using a restriction digest with NdeI and EcoRI (NEB, R0111S and NEB, R3101L) and subsequent Gibson assembly (NEB, E2611L) to insert an FKBP-tag at the end of the coding sequence as well as a Rox-HygroR-Rox cassette for selection (final vector: RAD21-Halo-FKBP-Rox-HygroR-Rox). The clonal lines carrying the TetO and LacO cassettes as well as TeR-tdTomato and LacI-eGFP (clones 1B4 and 2C10) were transfected with the targeting vector (RAD21-Halo-FKBP-Rox-HygroR-Rox) and the gRNA vector PX330-EN1082 using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024 KT) as described above. 48 hours after transfection, 160 μ g/ml of Hygromycin B (Thermo Fisher Scientific, 10687010) was added to the medium for 3 days to select cells for insertion of the HaloTag-FKBP-tag cassette. 7 days after selection, 0.5×10^6 cells were transfected with 2 μ g of pCAGGS-Dre-IRES-bsd (purchased from Gene Bridges) using Lipofectamine3000 according to the manufacturer's instructions (Thermo Fisher Scientific, L3000008) to remove the Neomycin and Hygromycin resistances used as selection markers for previous integrations. Prior to single-cell sorting, the cells were incubated with 100 nM JF646 HaloTag Ligand (Grimm *et al.*, 2017) in full culturing medium for 30 min at 37 °C, 8%CO₂ and washed three times with PBS. The cells were sorted for fluorescent emission at 664 nm. Sorted cells were cultured and genotyped as described for the Tir1 integration. Primers used for genotyping are listed in Supplementary Table S3. Cell lines showing the corrected genotype were selected, expanded and validated for homozygous insertion of the HaloTag-FKBP-tag and correct functioning of the degron system by Western Blot (as described below). Clones 1B1 (-PuroR,-NeoR,-HygroR) and 2C11 (+PuroR,+NeoR,-HygroR) were used for further engineering.

Removal of CTCF sites in dual array TetO-LacO cell lines

To selectively remove the three CTCF binding sites flanking the operator arrays, 0.5×10^6 cells of the TetO-LacO dual array cell line + RAD21-HaloTag-FKBP (clone 1B1) were transfected 1 μ g of pIC-Cre using Lipofectamine3000 according to the manufacturer's instructions (Thermo Fisher Scientific, L3000008). 5 days after transfection, 0.5×10^6 cells of the pool were further transfected with 1 μ g pCAG-FlpO-P2A-HygroR (Zuin *et al.*, 2022). Cells were then sorted and genotyped as described previously. Primers used for genotyping are listed in Supplementary Table S3. Since the recombination with Flippase did not work sufficiently to attain correct clones, the CTCF sites

flanking the LacO array were then removed using CRISPR-Cas9 deletion. The gRNA sequences for the CRISPR/Cas9 knock-out of the CTCF sites flanking the LacO array were designed using the online tool https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE and purchased from Microsynth. The gRNA sequence can be found in Supplementary Table S3. The gRNA sequence was cloned into the PX330 plasmid (Addgene, #58778) using the BsaI restriction site as described previously. The pool of cells transfected with pIC-Cre was further transfected with 0.5 μ g of each gRNA/Cas9 vector and subsequently sorted and genotyped. Correct clones were expanded and validated by genotyping PCR that was analyzed on a 1% agarose gel imaged with a Typhoon FLA 9500 scanner (GE Healthcare). Subsequent Sanger sequencing of the PCR product (Microsynth) confirmed the removal and clones 1B1 (+CTCF sites, -promoters, +RAD21-HaloTag-FKBP), 1A2 (-CTCF sites, -promoters, +RAD21-HaloTag-FKBP) and 1F4 (-CTCF sites, +promoters from resistance gene cassette, +RAD21-HaloTag-FKBP) were used in live-cell imaging and Capture-Hi-C experiments.

Generation of control cell lines to measure localization error and experimental uncertainty

To generate a control cell line for the dual array imaging, one clone of RAD21-AID-eGFP + 3xCTCF-TetO + TetR-tdTomato (clone 2B10) was transfected with 200 ng pBroad3_hyPBase_IRES_tagRFPT and 200 ng PB-TetR-eGFP using Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. Cells were cultured in standard E14 medium for 7 days and sorted (as described previously) for fluorescent emission at 507 nm (eGFP) and 581 nm (tdTomato). Clonal lines were screened for a good SNR by microscopy on Corning High-Content Imaging Glass Bottom Microplates (96-well, Corning, 4580) and were used for estimation of localization error and the experimental uncertainty on the distance by live-cell imaging using the same analysis pipeline as for the TetO-LacO dual-array cell line (see description below).

piggyBac insertion site mapping

The integration sites of the random integrations by PiggyBase were mapped as described in Redolfi *et al.*, 2019. In short, genomic DNA (2 μ g) was fragmented to an average of 500 bp by sonication (Covaris) and ligation of full-length barcoded Illumina adapters was performed using the TruSeq DNA PCR-free kit (Illumina) according to the manufacturer's guidelines, with the exception that large DNA fragments were not removed. Libraries were pooled together and capture of desired fragments was performed using biotinylated probes against the piggyBac ITRs sequences using xGEN Hybridisation reagents (IDT). Following capture, libraries were amplified for 14 cycles (KAPA HiFi Hotstart). Sequencing was performed on the NextSeq500 platform (Illumina) as paired-end 300 cycles.

Capture-Hi-C sample preparation

Capture-Hi-C sample preparation was performed as described previously (Zuin *et al.*, 2022). In short, for RAD21 depletion, 2×10^7 cells were treated with 500 nM dTag-13 (Sigma-Aldrich, SML2601-1MG) for 2 h at 37 °C. All cells were then crosslinked with 1% formaldehyde (EMS, 15710) for 10 min at RT. The reaction was quenched with glycine (final concentration 0.125 M). Lysis was performed in 1 M Tris-HCl pH 8.0, 5 M NaCl and 10% NP40 (Sigma-Aldrich, I8896-50ML) and Complete protease inhibitor (Sigma-Aldrich, 11836170001). Cells were digested using 100 U of MboI (NEB, R0147) and

ligated at 16 °C with 10,000 U of T4 DNA ligase (NEB, M0202) in ligase buffer supplemented with 0.8% Triton X-100 (Sigma-Aldrich, T8787) and 240 µg of BSA (NEB, B9000). De-crosslinking was achieved with 400 µg Proteinase K (Macherey Nagel, 740506) at 65 °C. The 3C sample was purified using a phenol/chloroform extraction. 3C library preparation and target enrichment using a custom-designed collection of 6979 biotinylated RNA “baits” targeting single MboI restriction fragments chr15:10,283,500-13,195,800 (mm9) (Supplementary Table S3; Agilent Technologies; as in Zuin *et al.*, 2022) were performed following the SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol. However, 9 µg of 3C input material instead of 3 µg was used for the capture and DNA was sheared using Covaris sonication with the following settings: Duty Factor: 10%; Peak Incident Power (PIP): 175; Cycles per Burst: 200; Treatment Time: 480 s; Bath Temperature: 4 °C to 8 °C).

Targeted nanopore sequencing with Cas9-guided adapter ligation (nCATS)

nCATS was performed as described previously in Zuin *et al.*, 2022. In short, 3-5 gRNAs sequences each (targeting the upstream and downstream regions 2-3 kb external of the respective integration cassette, either LacO or TetO integration) were designed using the IDT online tool https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE (Supplementary Table S3). Custom designed Alt-R CRISPR-Cas9 crRNAs (3-5 crRNAs targeting the region 5' and 3-5 crRNAs targeting the region 3' of the integrated transgene), Alt-R CRISPR-Cas9 tracrRNA (IDT, 1072532) and Alt-R S.p. Cas9 enzyme (IDT, 1081060) were purchased from IDT. Genomic DNA from clones 2G5 and 1F11 was extracted with Genra Puregene Cell Kit (Qiagen, 158745) following the manufacturer's instructions. Quality of the High Molecular Weight (HMW) DNA was checked with the TapeStation (Agilent) and 5 µg of HMW DNA were de-phosphorylated using Shrimp Alkaline Phosphatase (rSAP; NEB, M0371) for 3 min at 37 °C followed by 5 min at 65 °C. To assemble the Alt-R guide RNA duplex (crRNA:tracrRNA), the six Alt-R CRISPR-Cas9 crRNAs were pooled to a final concentration of 100 µM and subsequently incubated in a ratio of 1:1 with 100 µM of Alt-R CRISPR-Cas9 tracrRNA at 95 °C for 5 min. 4 pmol of Alt-R S.p. Cas9 enzyme were incubated with 8 pmol Alt-R guide RNA (crRNA:tracrRNA) at RT for 20 min to assemble the RNP complex. *In vitro* digestion and A-tailing of the DNA were performed by adding 10 µl of the RNP complex, 10 mM of dATP (NEB, N0440) and 5 U of Taq Polymerase (NEB, M0267) and incubating the samples at 30 min, 37 °C followed by 5 min, 72 °C. Nanopore sequencing adaptors were ligated using the Ligation Sequencing Kit (Oxford Nanopore Technologies, SQK-CAS109) according to the manufacturer's instructions. After purification with AMPure PB beads (Beckman Coulter, A63881), samples were loaded into MinION selecting SQL-CAS109 protocol (Oxford Nanopore Technologies).

Hi-C sample preparation

Hi-C sample preparation was performed as described previously in Redolfi *et al.*, 2019. Briefly, 6×10^6 cells were treated with 500 µM auxin (Sigma-Aldrich, I5148-2G) for 90 min and cells were crosslinked with 1% formaldehyde (EMS, 15710) and quenched with 0.125 M glycine for 5 min at RT. Cells were lysed in 10 nM Tris-HCl pH 8.0, 10 nM NaCl, 0.2% NP-40 (Sigma-Aldrich, I8896-50ML), complete protease inhibitor (Sigma-Aldrich, 11836170001) and nuclei were digested with 400 U of MboI (NEB,

R0147) at 37 °C overnight. End-repair was performed using 40 μM Biotin-11-dATP (Life Technologies, 19524-016) and 50 U DNA Polymerase I Large Klenow fragment (NEB, M0210M) incubating at 37 °C for 45 min. The end-repaired samples were ligated using 10,000 U T4 DNA ligase (NEB, M0202M) in ligase buffer supplemented with 0.8% Triton X-100 (Sigma-Aldrich, T8787) and 120 μg BSA (NEB, B9000) at 16 °C overnight. De-crosslinking was performed by adding 20 μl Proteinase K (20 mg/ml, Macherey-Nagel, 740506) to the ligation mix (1.2 ml) and incubating at 65 °C overnight. DNA was purified using phenol/chloroform and 2 μg of purified 3C sample was sonicated using the Bioruptor Pico (Diagenode). Biotinylated DNA was captured using MyOne Streptavidin T1 magnetic beads (Life Technologies, No. 65601) followed by A-tailing. Library preparation was performed according to NEBNext Ultra DNA Library prep kit instruction (NEB, E7370L) and samples were purified with magnetic AMPure bead (Beckman Coulter, A63881). Hi-C libraries were sequenced on an Illumina Nextseq500 platform (2x42 bp paired-end).

4C sample preparation

Sample preparation for 4C was performed as previously described (Splinter *et al.*, 2012). In short, 10⁷ cells were treated with 500 μM auxin (Sigma-Aldrich, I5148-2G) for 90 min and cross-linked in 2% paraformaldehyde (EMS, 15710) for 10 min and quenched with 0.125 M glycine (final concentration). Lysis was performed in 150 mM NaCl, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.5% NP-40 (Sigma-Aldrich, I8896-50ML), 1% Triton X-100 (Sigma-Aldrich, T8787). The samples were digested with 200 U DpnII (NEB, R0543M) and subsequently ligated at 16 °C with 50 U T4 DNA ligase (Roche, #10799009001) in a final reaction volume of 7 ml. De-crosslinking was performed with Proteinase K (0.05 μg/μl) at 65 °C and samples were then purified using phenol/chloroform extraction. The second digest was performed with 50 U Csp6I (Thermo Fisher Scientific, ER0211). Samples were ligated with 100 U T4 DNA ligase in a final volume of 14 ml and purified by precipitation in 100% ethanol. The resulting products were used directly as a PCR template for the TetO 4C viewpoint. Primers for PCR were designed according to the set-up used in Redolfi *et al.*, 2019 and can be found in Supplementary Table S3. Library preparation was performed with the NEBNext Ultra DNA Library prep kit according to the manufacturer's instructions (NEB, E7370L). 4C libraries were sequenced on an Illumina Hiseq2500 platform (50 cycles, single-end reads).

Cell cycle analysis of RAD21 depleted cells by flow cytometry

To validate that in auxin or dTAG-13 treated cells the distribution of cells in different stages of the cell cycle is not skewed towards cells in S and G2 phase (as cells are arrested in mitosis upon depletion of RAD21), cell cycle stage analysis was performed by flow cytometry. For this, cells were cultured until confluency on a 6-well and treated with the corresponding compound (500 μM auxin (Sigma-Aldrich, I5148-2G) or 500 nM dTAG-13 (Sigma-Aldrich, SML2601-1MG) resuspended in culturing medium) for the time indicated (0 h, 1.5 h, 2 h, 6 h). The cells were then harvested and 3x10⁶ cells were fixed in 4% paraformaldehyde (EMS, 15710) at RT for 15 min. The cells were washed in PBS and stained with 5 μg/ml 4',6-Diamidino-2-phenylindole (DAPI) (D9564-10MG, Sigma-Aldrich) in 1xPBS+0.1% Triton X-100 (Sigma-Aldrich, T8787) for 30 min at RT. The cells were analyzed with a 405 nm laser line at BD LSR II SORP Analyser (BD Biosciences, BD FACSDiva™ Software v8.0.1). Distributions of cell

cycle stage profiles were analyzed using FlowJo (v10, BD Biosciences).

Western Blot

To validate the targeted degradation of RAD21, CTCF and WAPL in the degron cell lines, cells were cultured on a 6-well to confluency and degradation was induced by adding 500 nM dTag-13 (Sigma-Aldrich, SML2601-1MG) or 500 μ M auxin (Sigma-Aldrich, I5148-2G) and incubating for the indicated time (0 h, 1.5 h, 2 h, 6 h, 24 h) at 37 °C, 8% CO₂. The cells were washed twice with PBS and lysed in 200 μ l RIPA buffer (150 mM sodium chloride, 1.0% NP-40 (Sigma-Aldrich, I8896-50ML), 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 50 mM Tris pH 8.0) incubating 1 min at 4 °C. Lysed cells were frozen in liquid nitrogen. For the SDS-PAGE, cell lysates were thawed on ice and 2.5 U/ml SuperNuclease (Sino Biological Inc, SSNP01) was added to digest DNA for 10 min. Protein levels were quantified using the Pierce BCA protein assay (Thermo Fisher Scientific, 23225). 10 μ g of protein extract were loaded onto a Mini-PROTEAN TGX Precast Gel (4-15% gradient, BioRad, 4561086) in 1xLaemmli buffer (BioRad, 1610747) containing 100 mM DTT. Samples were run at 180 V for 90 min in Tris-Glycine buffer (25 mM Tris pH 8.3, 192 mM Glycine, 0.1% (w/v) SDS). The protein was transferred to a Trans-Blot Turbo Mini 0.2 μ m Nitrocellulose membrane (BioRad, 1704158) using the Trans-Blot Turbo Transfer System from BioRad. The membrane was blocked for 1 h at RT (shaking) in Odyssey Blocking Buffer (PBS) (Li-Cor Biosciences, 927-40000). The membrane was incubated with 1 μ g/ml primary antibody (rabbit polyclonal anti-RAD21, abcam ab154769, rabbit polyclonal anti-WAPL, Proteintech 16370-1-AP, rabbit polyclonal anti-CTCF, Cell Signaling Technologies #2899S, rabbit anti-PK-tag (V5), abcam ab15828, and mouse monoclonal anti-Tubulin, (DM1A), Cell Signaling Technologies #3872) in Odyssey Blocking Buffer (PBS) overnight at 4 °C. The membrane was washed three times in PBS+0.1%Tween-20 for 5 min each shaking and then incubated for 1 h at RT (shaking) with secondary antibodies (IRDye 800CW Goat anti-rabbit IgG and IRDye 680RD Goat anti-Mouse IgG, dilution 1:10,000, Li-Cor BioSciences 926-32211 and 926-68070) in Odyssey Blocking Buffer (PBS). The membrane was washed three times in PBS+0.1%Tween-20 and imaged on the Odyssey infrared imaging system (Li-Cor Biosciences).

Estimation of sister chromatids

To estimate the probability of encountering sister chromatids in our dual-color experiments, we manually labeled around 1400 images randomly sampled from all movies. We assigned spots to cell masks detected using CellPose. We estimated that sister chromatids occur in approx. 3% of the cells.

Median convergent CTCF distance

To calculate the average distance between convergent CTCF within TADs, we took the list of CTCF motifs from Nora *et al.*, 2017. The list of TADs has been taken from Zuin *et al.*, 2022. By keeping only pairs of convergent CTCF motifs within TADs, the average distance is estimated to be 140,932 bp.

Analysis of piggyBac insertion site mapping

To exclude reads coming from the TetR or LacI, the two ends of paired-end reads were mapped separately to the piggyBac-TetR and LacI sequences (https://github.com/giorgettilab/Mach_et_al

_chromosome_dynamics/tree/master/sequences) using QuasR v. 1.36.0 (qAlign). Only unmapped reads were kept and mapped to the piggybac-TetO array sequence (https://github.com/giorgetti/lab/Mach_et_al_chromosome_dynamics/tree/master/sequences). Hybrid pairs with one of the read-end mapping to array were kept. The second reads from hybrid pairs were mapped to the mouse genome (build mm9) using QuasR (qAlign). Reads were then piled up in 25 bp windows using csaw v. 1.30.1 (windowCounts function). Integration sites can be identified because they correspond to local high-read coverage. Local coverage was calculated by resizing all non-zero 25 bp windows up to 525 bp (expanding by 250 bp upstream and downstream). Overlapping windows were then merged using reduce (from GenomicRanges v. 1.48.0), resulting in a set of windows w_i . The size distribution of w_i is multimodal, and only w_i from the second mode onward were kept. For each w_i we estimated the coverage c_i as the number of non-zero 25 bp windows. Only w_i where the coverage was >16 were considered. The exact positions of the integration sites were then identified with the center of w_i .

Hi-C analysis

Hi-C data were analyzed using HiC-Pro version 3.1.0 (Servant *et al.*, 2015) with the `-very-sensitive` `-end-to-end` `-reorder` options. Only unique reads were mapped to the mouse genome (build mm9). Contact maps were combined at 8 kb with iterative correction applied afterwards. (https://github.com/giorgetti/lab/Mach_et_al_chromosome_dynamics/tree/master/Hi-C).

Loop analysis

Loops were called on Hi-C map of WT sample using Mustache (Roayaei Ardakany *et al.*, 2020) version 1.0.1 on 8 kb contact map, with p-value threshold 0.1 and sparsity threshold 0.75. Pileup analysis was performed for all samples based on WT loops using coolpup.py version 0.9.2 (Flyamer *et al.*, 2020).

4C analysis

First, we trimmed the connector sequence, then we filtered out over-represented sequences. We mapped sequences to the mouse mm9 genome using a custom R script (https://github.com/giorgetti/lab/Mach_et_al_chromosome_dynamics/).

Nanopore sequencing analysis

To map nanopore sequencing reads, we first built a custom “genome” consisting of the LacO or TetO cassette flanked by approx. 2.5 kb mouse genomic sequence upstream and downstream of the target integration site. The custom genome can be found at (https://github.com/giorgetti/lab/Mach_et_al_chromosome_dynamics/tree/master/nanopore). Analysis has been performed as in Zuin *et al.*, 2022. Briefly, reads were mapped to the custom genome using minimap2 (v. 2.17-r941) with “-x map-ont” parameter. Nanopore sequencing analysis has been implemented using Snakemake workflow (v. 3.13.3). Reads were visualized using IGV (v. 2.9.4). Full workflow can be found at https://github.com/zhanyinx/Zuin_Roth_2021/tree/main/Nanopore.

Capture-Hi-C analysis

Capture-Hi-C data were processed as in Zuin *et al.*, 2022 using HiC-Pro (Servant *et al.*, 2015) (v. 2.11.4). Briefly, read pairs were mapped to the mouse genome (build mm9). Chimeric reads were recovered after recognition of the ligation site. Only unique valid pairs mapping to the target regions were used to build contact maps. Iterative correction (ICE) (M. Imakaev *et al.*, 2012) was then applied on binned data. The target regions can be found at https://github.com/zhanyinx/Zuin_Roth_2021.

Differential Capture-Hi-C maps

We accounted for differences in genomic distances due to the presence of the ectopic sequence when evaluating the structural perturbation induced by the insertion of the TetO and LacO arrays. To account for these differences, we generated distance-normalized capture-Hi-C maps where each entry corresponds to the interaction normalized by the corrected genomic distance between the interacting bins. We then calculated the WT-allele corrected ratios between distance normalized maps using the following formula:

$$ratio = \frac{2 \cdot mut - wt}{wt} \quad (4.1)$$

Where *mut* is the distance normalized heatmap of the line with TetO or LacO arrays and *wt* is the distance normalized heatmap of the wild-type. A bilinear smoothing with a window of 2 bins has been applied to the ratio maps to evaluate the structural perturbation induced by the insertion of the arrays. To quantify the effect of ectopic CTCF sites, we followed the same approach as the quantification of the arrays insertion with the exception that the ratio is calculated using the following formula

$$ratio = \frac{2 \cdot mut_+ - WT}{2 \cdot mut_- - WT} \quad (4.2)$$

Where *mut₊* and *mut₋* represent the lines with and without the ectopic CTCF (or promoter), respectively.

Conversion of simulation steps to real time

In order to convert the timescales between experiments and simulations, we measured the time needed for a bead (and 8 kb TetO array) to move for its own size ($95\text{nm} = 15\text{nm} \times \sqrt{40}$ = estimated size of a nucleosome x estimated number of nucleosomes in each 8 kb segment, assuming that chromatin inside each 8 kb behave as an ideal chain; this is in line with the previous estimations for a 3 kb segment (Giorgetti *et al.*, 2014)). To this aim, 2D movies acquired in cells with randomly inserted TetOs every 0.1 s for 10 s were analyzed with the same strategy as in **Multi-operator image analysis and localization error estimation**, now without 3D linkage. We used the MSD curve to estimate the time necessary to move for 95 nm for each line. We used the average value across 3 lines to convert the timescales between experiments and simulations.

In simulations, a bead needs approx. 1800 simulation steps to move of its own size, whereas in

experiment this time is estimated to be approx. 0.86 s. Therefore 1 s of the real time corresponds to approx. 2100 simulation steps. This approximation neglects the fact that the coarse-grained bead describes a segment of the fiber that in the cell can display internal dynamics, and the length represented by the bead can be variable. We assumed that such internal dynamics is faster and uncorrelated to the motion we want to simulate. Thus each bead is characterized only by its average size, that is 95 nm, neglecting the fluctuations around the average.

Metrics for quantifying similarity between simulations and experiments

To evaluate the similarity between experimental features (f) (here: contact duration, probability of being in the proximal state, fraction of time spent in the proximal state) we used the relative difference (RD) defined as:

$$RD = \frac{|f_{exp} - f_{sim}|}{f_{exp}} \quad (4.3)$$

To find the set of loop extrusion parameters that best reproduce the experiments in the presence of cohesin, we used the sum of the relative difference of the following features: gaussian mean of contact state, the contact duration and the fraction of time spent in the looped state in the \pm CTCF. To find the set of loop extrusion parameters that best reproduce the experiments in the absence of cohesin, we decreased the loading rate to reach extruder densities of 0.87 per Mb while keeping the extruder residence time constant (5.5 min).

Polymer modeling

Introduction

The theory of anomalous diffusion was developed studying the dynamics of loci on a polymer chain with respect to a fixed reference system. In our experimental work we measure distances between pairs of genomic locations (Figs. 3, 4, 5). We show in the following that one can extract the scaling exponent that characterizes subdiffusion from the dynamics of the distance between pairs of loci.

We first discuss the subdiffusive motion of the difference vector between two beads of a Rouse chain, that can be solved analytically and gives for short times the standard scaling exponent 1/2 as that of the position of a bead. Then we discuss the dynamics of the scalar distance between two beads, showing numerically that for short times it gives the same scaling exponent as the vector distance.

Finally, we show that in presence of excluded volume the scaling exponent is larger than the one predicted for the Rouse chain.

Subdiffusion of the difference vector in the Rouse chain

Consider a standard Gaussian chain of N links. In the continuous limit, the position $\mathbf{r}(x, t)$ of a locus on the chain evolves according to the diffusion equation

$$\xi \frac{\partial \mathbf{r}(x, t)}{\partial t} = \mathbf{f}^r(x, t) + \frac{3T}{a^2} \frac{\partial^2 \mathbf{r}(x, t)}{\partial x^2}, \quad (4.4)$$

where $x \in [0, N]$ labels the position along the chain and is assumed to be continuous, a is the distance between neighbouring beads and is thus assumed small, T is the temperature, ξ is the friction coefficient and \mathbf{f}^r is the random force with zero mean and correlations

$$\langle f_\alpha^r(x, t) \cdot f_\beta^r(x', t') \rangle = 2\xi T \delta(x - x') \delta(t - t') \delta_{\alpha\beta}. \quad (4.5)$$

Following the standard solution of the Rouse chain (see, e. g., Khokhlov *et al.*, 1994), let's write the position of a bead on the chain as a Fourier series

$$\mathbf{r}(x, t) = y_0(t) + 2 \sum_{p=1}^{\infty} y_p(t) \cos\left(\frac{\pi p x}{N}\right). \quad (4.6)$$

where $p = 0, 1, 2, \dots$. The y_p coordinates are called *Rouse Modes* and are obtained by the inverse Fourier transform

$$y_p(t) = \frac{1}{N} \int_0^N dx \cos\left(\frac{\pi p x}{N}\right) \mathbf{r}(x, t). \quad (4.7)$$

Let's briefly discuss the mean square displacement of a single bead. This can be written as

$$\langle (\mathbf{r}(x, t) - \mathbf{r}(x, 0))^2 \rangle = \frac{6T}{N\xi} t + \frac{4Na^2}{\pi^2} \sum_{p=1}^{\infty} \frac{1}{p^2} \cos^2\left(\frac{\pi p x}{N}\right) (1 - e^{-t/\tau_p}). \quad (4.8)$$

where $\tau_p \equiv \frac{N^2 a^2 \xi}{3\pi^2 T p^2}$ is the Rouse relaxation time for mode p ; in particular, τ_1 is the slowest relaxation mode.

For large times $t \gg \tau_1$ the second term in (4.8) is negligible and the mean square displacement of a bead is the same as that of the whole coil. On the other hand, when $t \ll \tau_1$ the large values of p are dominant and the sum can be approximated by an integral. Moreover, if we consider beads far from the ends of the chain, the function $\cos^2(\frac{\pi p x}{N})$ is fast oscillating and can be substituted by its mean value $\frac{1}{2}$. In this limit, the mean square displacement of a bead is approximately

$$\langle (\mathbf{r}(x, t) - \mathbf{r}(x, 0))^2 \rangle \approx \frac{2\sqrt{3}a}{\sqrt{\pi}} \sqrt{\frac{T}{\xi}} \sqrt{t}. \quad (4.9)$$

Now, we want to compute the mean square displacement of the difference vector between two beads

$$\Delta(t) = \langle (\boldsymbol{\delta}(t) - \boldsymbol{\delta}(0))^2 \rangle, \quad (4.10)$$

where $\boldsymbol{\delta}(t) = \mathbf{r}(x, t) - \mathbf{r}(y, t)$. Then, using Eq. (4.6) we have

$$\begin{aligned} \Delta(t) &= \langle \mathbf{r}(x, t) - \mathbf{r}(y, t) - \mathbf{r}(x, 0) + \mathbf{r}(y, 0) \rangle \\ &= 4 \sum_{p=1}^{\infty} \left(\cos\left(\frac{\pi p x}{N}\right) - \cos\left(\frac{\pi p y}{N}\right) \right)^2 \langle (y_p(t) - y_p(0))^2 \rangle \end{aligned} \quad (4.11)$$

But

$$\begin{aligned} \langle (y_p(t) - y_p(0))^2 \rangle &= \langle y_p^2(t) \rangle + \langle y_p^2(0) \rangle - 2 \langle y_p(t) \cdot y_p(0) \rangle \\ &= \frac{3T\tau_p}{N\xi} (1 - e^{-t/\tau_p}) \end{aligned} \quad (4.12)$$

where we used $\langle y_p(t) \cdot y_p(t') \rangle = \frac{3T\tau_p}{2N\xi} e^{-(t-t')/\tau_p}$.

Combining Eqs. (4.11) and (4.12) we get

$$\begin{aligned} \Delta(t) &= \frac{4Na^2}{\pi^2} \sum_{p=1}^{\infty} \frac{1}{p^2} \left(\cos \frac{\pi p x}{N} - \cos \frac{\pi p y}{N} \right)^2 \left(1 - e^{-t/\tau_p} \right) \\ &= \frac{4Na^2}{\pi^2} \sum_{p=1}^{\infty} \frac{4}{p^2} \sin^2 \left(\frac{\pi p}{2N} (L + 2x) \right) \sin^2 \left(\frac{\pi p}{2N} L \right) \left(1 - e^{-t/\tau_p} \right), \end{aligned} \quad (4.13)$$

where in the last step we used prosthaphaeresis formulae and defined $L = |x - y|$ as the genomic distance. In the continuous limit, with the same approximations employed to obtain eq. (4.9), we substitute the sum with an integral and the former *sin* function with its mean value $\frac{1}{2}$. Eventually,

$$\begin{aligned} \Delta(t) &= \frac{8Na^2}{\pi^2} \int_0^{\infty} \frac{1}{p^2} \sin^2 \left(\frac{\pi p}{2N} L \right) \left(1 - e^{-\frac{3\pi^2 T t}{N^2 a^2 \xi} p^2} \right) \\ &= \frac{4\sqrt{3}a}{\sqrt{\pi}} \sqrt{\frac{T}{\xi}} \sqrt{t} \left(1 - e^{-\tau/t} \right) + 2La^2 \operatorname{erfc} \left(\frac{La\sqrt{\xi}}{\sqrt{12}\sqrt{T}} \frac{1}{\sqrt{t}} \right), \end{aligned} \quad (4.14)$$

where we defined the crossover time $\tau = \frac{L^2 a^2 \xi}{12T}$. In the limit of large times $t \gg \tau$ the first term is negligible and the mean square displacement of the distance between two beads approaches

$$\Delta(t) \approx 2La^2, \quad (4.15)$$

corresponding to a plateau. In fact, this value is independent on time for times beyond τ and depends on the genomic distance L between the two sites, taking a value that is twice the steady-state mean squared distance between two beads on the chain. Noticeably, also the crossover time τ depends on

the genomic distance L , increasing quadratically with it. For long times the correlations between the two beads grow to compensate the sub-diffusion term; the correlations arise from the fact that on this time scale the two beads diffuse as a whole together with the beads laying between them. On the other hand, when $t \ll \tau$,

$$\Delta(t) \approx \frac{4\sqrt{3}a}{\sqrt{\pi}} \sqrt{\frac{T}{\xi}} \cdot \sqrt{t} \quad (4.16)$$

that is the same as the sub-diffusion of the single bead and reflect the initial regime when the two beads move independently on each other. The coefficient $\frac{4\sqrt{3}a}{\sqrt{\pi}} \sqrt{\frac{T}{\xi}}$ is exactly double the coefficient of the MSD of a single bead for short times (see eq. (4.9)), as expected.

Subdiffusion of the distance

The quantity that is directly accessible in dual colour experiments is not the difference vector but the distance $d(t) = |\boldsymbol{\delta}(t)|$ and in particular

$$\Delta'(t) = \langle (d(t) - d(0))^2 \rangle = \langle |\boldsymbol{\delta}(t)|^2 \rangle + \langle |\boldsymbol{\delta}(0)|^2 \rangle - 2\langle |\boldsymbol{\delta}(t)| \cdot |\boldsymbol{\delta}(0)| \rangle \quad (4.17)$$

which is different from

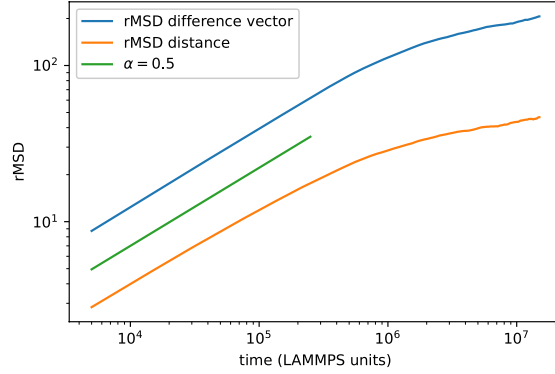
$$\Delta(t) = \langle (\boldsymbol{\delta}(t) - \boldsymbol{\delta}(0))^2 \rangle = \langle |\boldsymbol{\delta}(t)|^2 \rangle + \langle |\boldsymbol{\delta}(0)|^2 \rangle - 2\langle \boldsymbol{\delta}(t) \cdot \boldsymbol{\delta}(0) \rangle \quad (4.18)$$

because it contains the correlation between two scalars instead of a scalar product.

The difference between Eqs. (4.17) and (4.18) is the last term, that is the correlation function of either the displacement vector or the displaced distance to the initial ones. From the definition of scalar product, $\langle \boldsymbol{\delta}(t) \cdot \boldsymbol{\delta}(0) \rangle = \langle |\boldsymbol{\delta}(t)| \cdot |\boldsymbol{\delta}(0)| \cdot \cos \theta(t) \rangle$, where $\theta(t)$ is the angle between $\boldsymbol{\delta}(t)$ and $\boldsymbol{\delta}(0)$. The rotational diffusion of the direction defined by $\boldsymbol{\delta}(t)$ takes place with a diffusion coefficient $D_{rot} = 2T/\xi d^2$ (see Cugliandolo *et al.*, 2015), that is on an elementary timescale of the order of $\xi a^2 L/2T$. On the other hand, the motion of the distance $d(t)$ takes place, for $t < \tau$, on an elementary timescale $\xi a^2/2T$. Thus, one expects that for $L \gg 1$, $\cos \theta(t)$ is weakly dependent on time and the two correlation functions are different for a nearly constant quantity. In conclusion, $\Delta(t)$ and $\Delta'(t)$ are expected to show the same scaling behaviour at short times.

To substantiate these expectations, we have calculated $\Delta(t)$ and $\Delta'(t)$ for two beads ($L = |x - y| = 100$) of a Rouse chain (of length $N = 1000$). The results are displayed in Fig. S1. Both curves display the expected exponent $\approx 1/2$ in the initial linear part.

Also in the case of a bead-spring chain with excluded volume, the initial scaling of the difference vector is the same as that of the distance (see Fig. 2.2 and text below).



Supplementary Figure S1: The subdiffusion $\Delta(t)$ of the difference vector between two beads separated by 100 other beads in a Rouse chain and of that $\Delta'(t)$ of the distance between them.

Bead-spring chain with excluded volume

A scaling argument can be used to show that adding excluded volume to a Rouse chain, the scaling exponent of the sub-diffusive regime increases to $6/11 \approx 0.55$ (see Tamm and Polovnikov, 2017).

In fact, one can imagine to split the chain at time t into domains of size $g(t)$, like blobs that moves independently on each other. The genomic distance associated to $g(t)$ is $L(t)$, that is the number of monomers moving coherently in the domain; to be able to define a domain as something that moves independently on the other domains on the time scale t , its spatial size $g(t)$ must be of order of the displacement of a monomer over a time t ,

$$g(t) \sim x(t). \quad (4.19)$$

At the same time, the displacement of a single monomer has the same behaviour as that of the domain it belongs to; such a domain diffuses like a Brownian particle but with a diffusion coefficient which is inversely proportional to the number of monomers $L(t)$ (assuming that the thermal forces acting on the monomers are independent on each other), thus one has

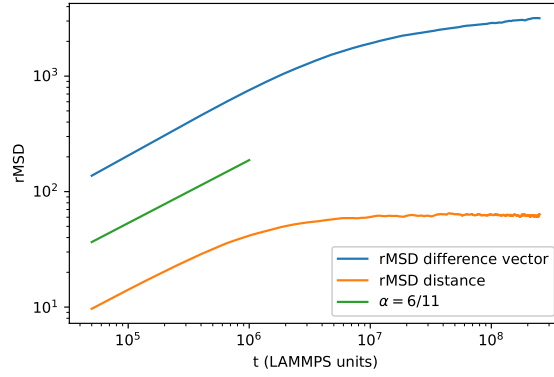
$$x^2(t) \sim \frac{t}{L(t)}. \quad (4.20)$$

For a chain with excluded volumes the size of a domain of L monomers has typical size

$$g \sim aL^{3/5} \quad (4.21)$$

and, substituting Eqs. (4.20) and (4.21) into Eq. (4.19), one obtains

$$L^{3/5}(t) \sim a \frac{t^{1/2}}{L^{1/2}(t)} \rightarrow L(t) \sim t^{5/11}. \quad (4.22)$$



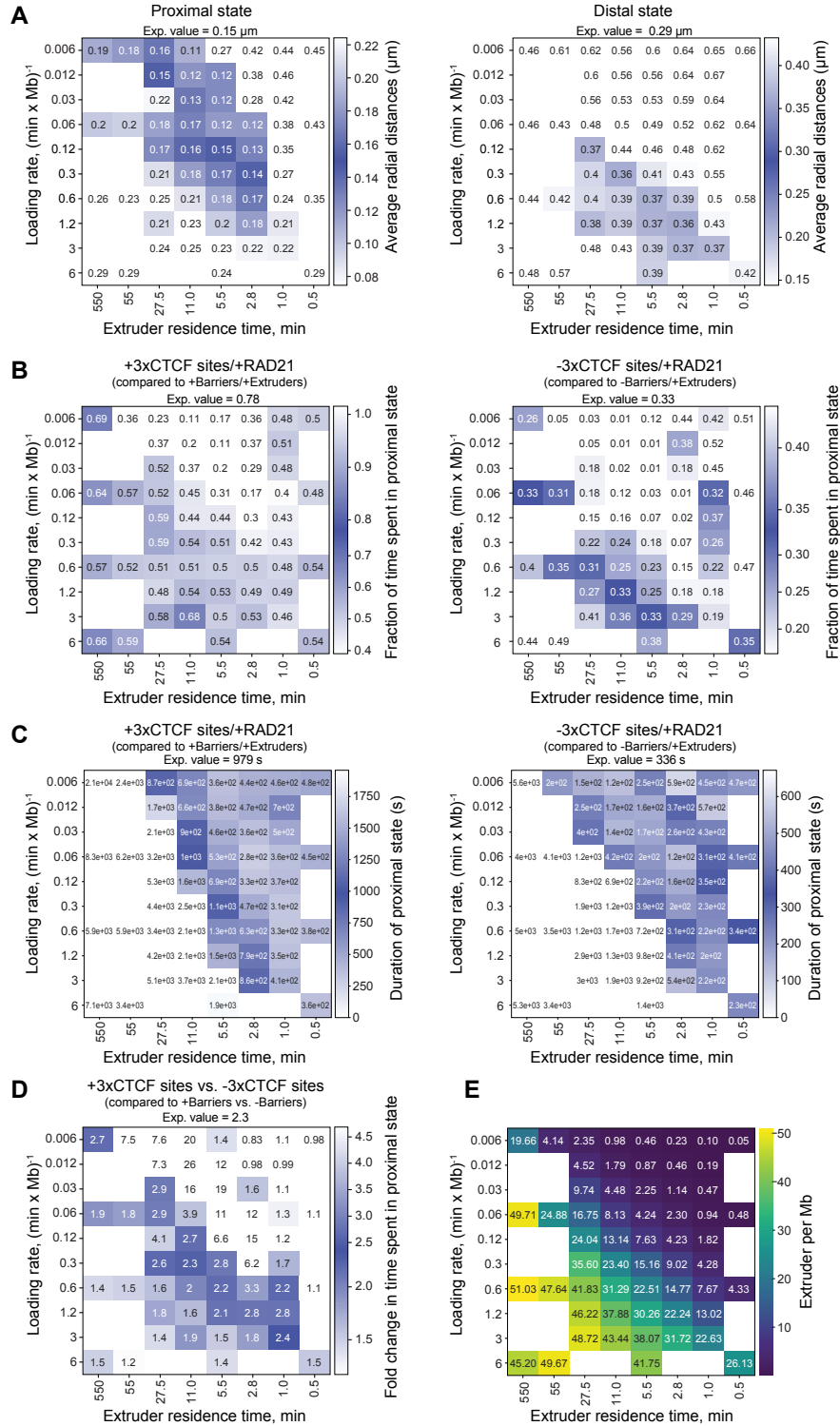
Supplementary Figure S2: The subdiffusion $\Delta(t)$ of the difference vector between two beads separated by 100 other beads in a bead-spring chain with excluded volume and of that $\Delta'(t)$ of the distance between them. The excluded volumes are implemented by the repulsive-only Lennard-Jones potential.

Finally, Eq. (4.20) can be rewritten as

$$x^2(t) \sim t^{6/11} \approx t^{0.55}. \quad (4.23)$$

Following the arguments of the previous sections, we should thus expect a power law scaling $rMSD \sim t^{6/11}$ of the pairwise distance between two beads in the case of a chain with excluded volumes, in the small times regime where the saturation is not yet occurring. This behaviour is quite evident in Fig. S2 for both the distance and difference vector curves.

Supplementary Figures



Supplementary Figure S3: Features of HMM states called on simulations across the parameter space. A, Heatmap showing the agreement of average radial distances of either the proximal state (left) or the distal state (right) called by HMM on all simulated systems of extrusion speed 1 kb/s with the proximal state or

Supplementary Figure S3: *continued from previous page:* **B**, distal state called by HMM on the experimental data. Darker shades of blue indicate better agreement with experimental values. **C**, Same as in panel A, but for the fraction of time spent in the proximal state (left) and distal state (right). **D**, Same as in panel A but for the duration of the proximal state for either the condition +Barriers/+3xCTCF sites (left) or -Barriers/-3xCTCF sites (right) in the presence of the Extruders/RAD21. **E**, Same as in panel A but for the fold change of time spent in a proximal state called by HMM on the experimental data for the comparison of the condition +Barriers/+3xCTCF sites vs. -Barriers/-3xCTCF sites in the presence of Extruders/RAD21. **F**, Extruder densities per Mb for all simulated systems with extrusion speed 1 kb/s. Color-coded for the density per Mb.

Supplementary Videos

Supplementary Video S1: Live-cell imaging of TetO arrays upon depletion of RAD21

Time course of RAD21 degradation upon induction with 500 μ M auxin in RAD21-AID-eGFP cells. TetO integrations are tagged with TetR-tdTomato (magenta) and RAD21 is tagged with eGFP (green). Green fluorescence is lost within 90 min after induction of degradation (exposure time (eGFP) = 50 ms, exposure time (tdTomato) = 50 ms, deconvolved, max. intensity projection, duration of movie 30 min, dt=10 s). *File is provided as a separate MPG file in the online version of this publication.*

Supplementary Video S2: Dynamics of LacO-TetO radial distances

Representative movie of dual-color imaging of LacO (green) and TetO (magenta) arrays flanked by 3xCTCF sites integrated on chromosome 15 at a distance of 150 kb (exposure time (eGFP) = 50 ms, exposure time (tdTomato) = 50 ms, deconvolved, max. intensity projection, duration of movie 1 h, dt=30 s). *File is provided as a separate MPG file in the online version of this publication.*

Supplementary Video S3: Cohesin and CTCF decrease average LacO-TetO radial distances

Representative movies of dual-color imaging of LacO (green) and TetO (magenta) arrays on chromosome 15 at a distance of 150 kb. Left panel: Cell line with 3xCTCF sites flanking LacO and TetO (in the presence of RAD21); middle panel: Cell line where 3xCTCF sites have been removed (in the presence of RAD21); Right panel: Cell line with 3xCTCF sites flanking the array, but where RAD21 has been degraded with 500 nM dTAG-13 (exposure time (eGFP) = 50 ms, exposure time (tdTomato) = 50 ms, deconvolved, max. intensity projection, duration of movie 1 h, dt=30 s). *File is provided as a separate MPG file in the online version of this publication.*

Supplementary Tables

Supplementary Table S1: Statistics of live-cell imaging data

Number of independent clonal lines, the number of cells imaged, and the number of tracks extracted from the movies for each condition tested in the live-cell imaging experiments. The number of tracks represent independent measurements for each condition. *File is provided as a separate XLS file in the online version of this publication.*

Supplementary Table S2: Statistical tests for duration and frequency of the HMM-called states

P-values based on two-sided Student t-tests for the comparison of duration and frequency of the proximal state between the different conditions from experimental live-cell imaging data and simulated data; based on the HMM-calls. *File is provided as a separate XLS file in the online version of this publication.*

Supplementary Table S3: Oligonucleotides

List of oligonucleotide sequences used in this study, including primer sequences used for cloning and genotyping of cell lines, gRNA sequences for Nanopore sequencing, capture probes for PiggyBac insertion site mapping and 4C primers. *File is provided as a separate XLS file in the online version of this publication.*

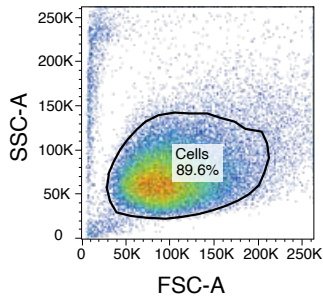
Supplementary Table S4: Spot detection and tracking parameters

List of spot detection and tracking parameters used for each live-cell imaging dataset, including names of the deepBlink model applied, spot and tracking parameters used in Fiji and motion-correction parameters applied. *File is provided as a separate XLS file in the online version of this publication.*

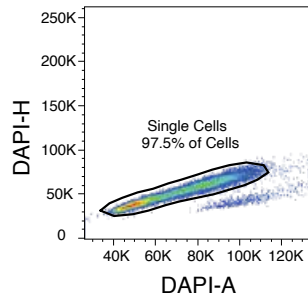
Flow cytometry gating strategy

A - Cell cycle stage analysis of Rad21-AID-eGFP cells after induction of Rad21 depletion with 500 μ M auxin

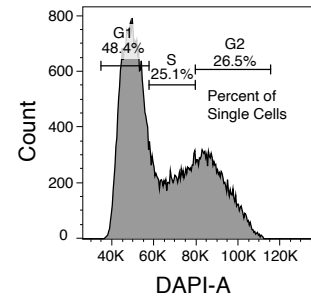
1) Discard big cells with high granularity



2) Discard doublets



3) Quantify no. of cells in each cell cycle stage based on DAPI intensity

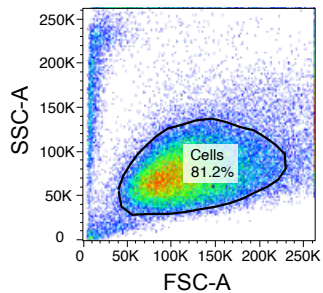


| Sample | Rad21-AID-eGFP untreated | Rad21-AID-eGFP 90min | Rad21-AID-eGFP 360min |
|----------------------------------|--------------------------|----------------------|-----------------------|
| Count of Events (Cells) | 52032 | 50946 | 51517 |
| Count of Events (Single cells) | 50710 | 49899 | 49634 |
| Percent of Parent (Single Cells) | 97.5 | 97.9 | 96.3 |
| Count of Events (G1) | 24562 | 21683 | 7688 |
| Percent of Parent (G1) | 48.4 | 43.5 | 15.5 |
| Count of Events (S) | 12709 | 12931 | 12112 |
| Percent of Parent (S) | 25.1 | 25.9 | 24.4 |
| Count of Events (G2) | 13439 | 15285 | 29834 |
| Percent of Parent (G2) | 26.5 | 30.6 | 60.1 |

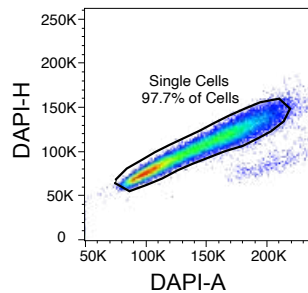
Gates for all three conditions were the same.

B - Cell cycle stage analysis of Rad21-HaloTag-FKBP cells after induction of Rad21 depletion with 500 nM dTag-13

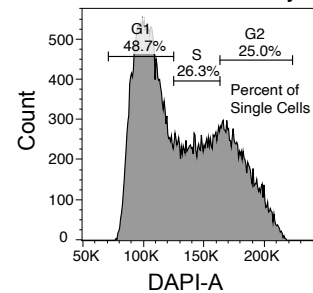
1) Discard big cells with high granularity



2) Discard doublets



3) Quantify no. of cells in each cell cycle stage based on DAPI intensity



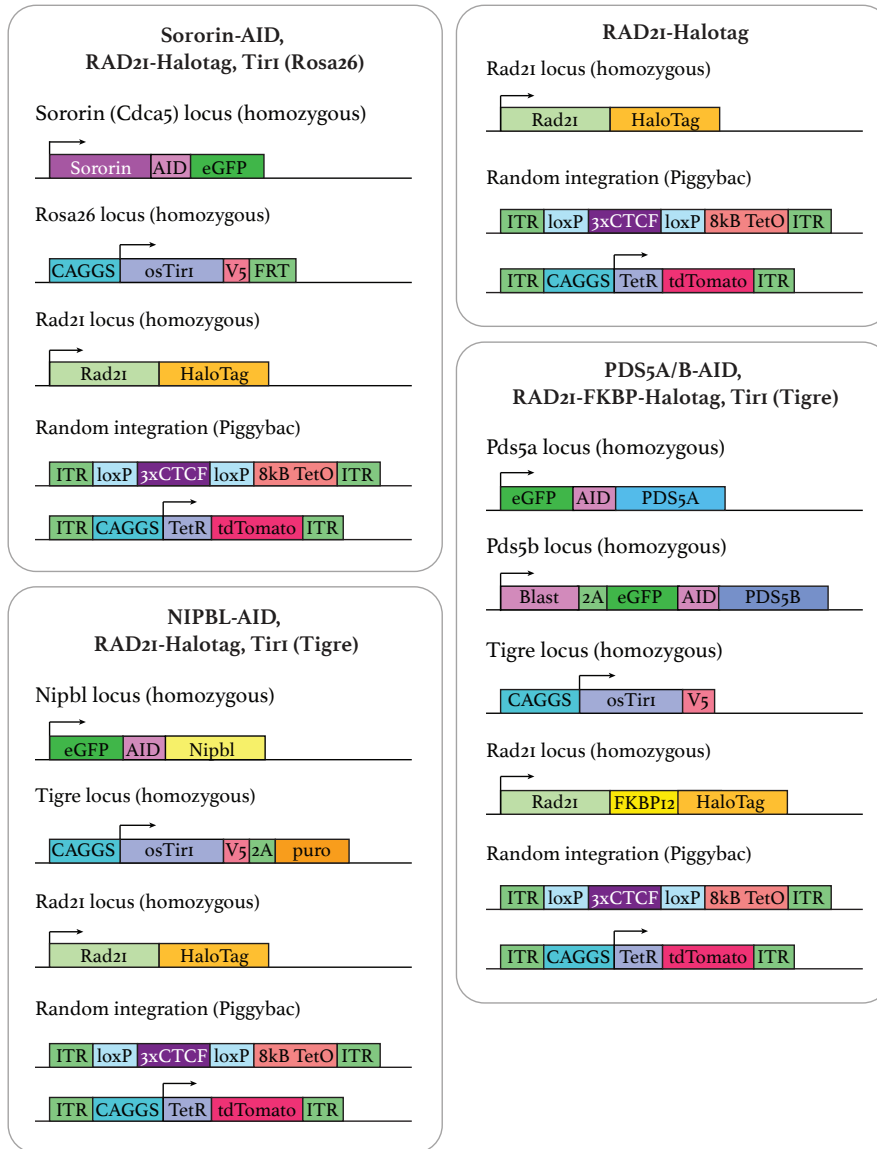
| Sample | Rad21-HaloTag-FKBP untreated | Rad21-HaloTag-FKBP 2h | Rad21-HaloTag-FKBP 6h |
|----------------------------------|------------------------------|-----------------------|-----------------------|
| Count of Events (Cells) | 49836 | 50093 | 50311 |
| Count of Events (Single cells) | 48685 | 48511 | 47344 |
| Percent of Parent (Single Cells) | 97.7 | 96.8 | 94.1 |
| Count of Events (G1) | 23708 | 21057 | 10845 |
| Percent of Parent (G1) | 48.7 | 43.4 | 22.9 |
| Count of Events (S) | 12803 | 12861 | 12478 |
| Percent of Parent (S) | 26.3 | 26.5 | 26.4 |
| Count of Events (G2) | 12174 | 14593 | 24021 |
| Percent of Parent (G2) | 25 | 30.1 | 50.7 |

Gates for all three conditions were the same.

SSC-A: Side scatter amplitude
 FSC-A: Forward scatter amplitude
 DAPI-H: DAPI height
 DAPI-A: DAPI amplitude

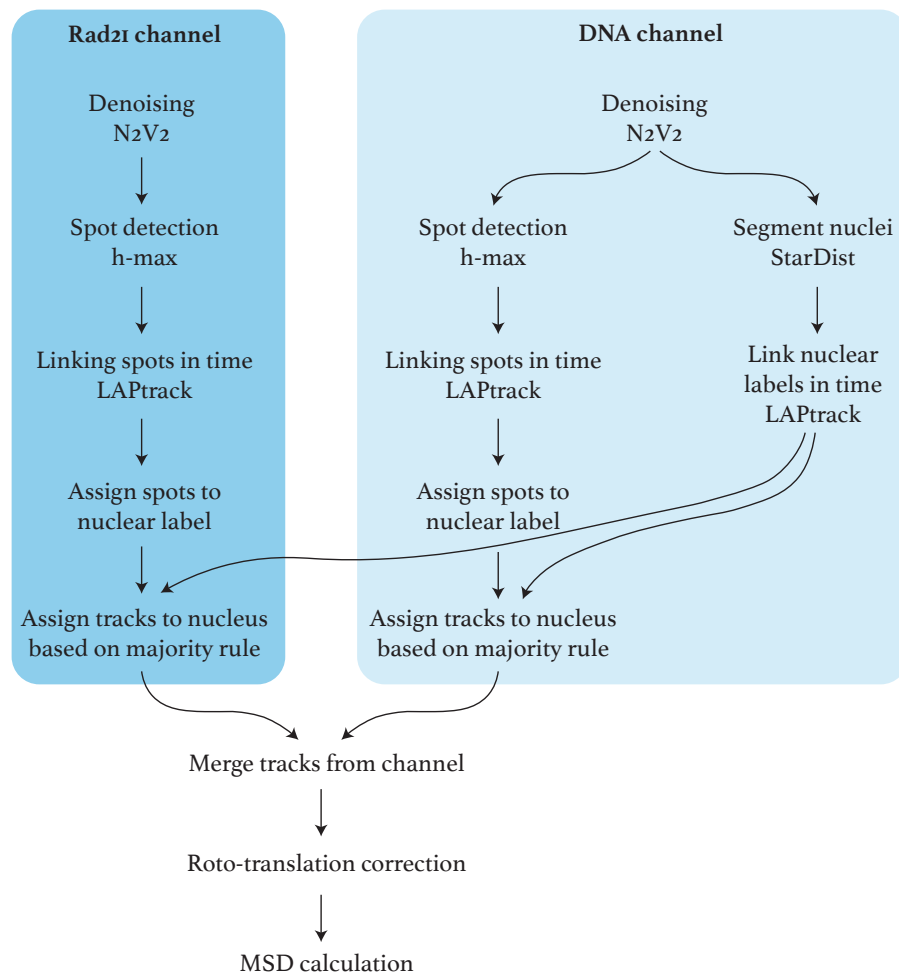
Chapter 3 – Cohesin dynamics on DNA: Cohesion or Extrusion?

Genotypes of degron cell lines



Supplementary Figure S4: Genotypes of degron cell lines used for SPT. Sororin-AID-eGFP, Rad21-HaloTag, Tir1 and Rad21-Halo cell lines were published in Nora *et al.*, 2020 and subsequently modified to also integrate the TetO and TetR-tdTomato. EGFP-AID-PDS5, Rad21-FKBP-HaloTag, Tir1 and eGFP-AID-NIPBL, Rad21-HaloTag, Tir1 were engineered by Elphège Nora (unpublished data) and were then further engineered to include random integrations of TetO and TetR-tdTomato. AID=auxin-inducible degron, eGFP=enhanced green fluorescent protein, ITR=inverted terminal repeats, V5=V5-Tag, 2A= 2A self-cleaving peptide, Puro=Puromycin resistance, CAGGS=CAGGS promoter (CMV enhancer + β -actin promoter). Schemes were adapted from schemes provided by Elphège Nora.

Overview on image analysis pipeline



Supplementary Figure S5: Overview on the image analysis pipeline. Denoising was performed using Noise2Void2 (N2V2), then spot detection was performed with h-max and spots were tracked using a linear assignment problem (LAPtrack). The nuclei were segmented based on the background fluorescence intensity in the DNA channel. This nuclear mask is then used to assign spots and tracks to a specific nucleus. The tracks from the Rad21 and TetO channel are then merged to calculate the roto-translation for correction of nuclear motion. Finally, MSDs are calculated from the tracks.

Appendix

The appendix includes the main text and figures of the publications to which I contributed during my doctoral studies. The extended manuscript files can be assessed via the links provided.

Integrative approaches to study enhancer–promoter communication

Mach, P., & Giorgetti, L.

Current Opinion in Genetics & Development, 80, 102052. (2023) <https://doi.org/10.1016/j.gde.2023.102052>

Cohesin and CTCF control the dynamics of chromosome folding

Mach, P.,* Kos, P. I.*, Zhan, Y.*, Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., & Giorgetti, L.

Nature Genetics, 54(12), 1907–1918. (2022) <https://doi.org/10.1038/s41588-022-01232-7>

*contributed equally

Nonlinear control of transcription through enhancer-promoter interactions

Zuin, J.*, Roth, G.*, Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., **Mach, P.**, Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C. van Steensel, B., Meister, P., Smallwood, S. & Giorgetti, L.

Nature, (604), 571-577. (2022) <https://doi.org/10.1038/s41586-022-04570-y>

*contributed equally

The contents of this study are not discussed in this thesis.



Integrative approaches to study enhancer–promoter communication

Pia Mach^{1,2,*} and Luca Giorgetti^{1,*}

The spatiotemporal control of gene expression in complex multicellular organisms relies on noncoding regulatory sequences such as enhancers, which activate transcription of target genes often over large genomic distances. Despite the advances in the identification and characterization of enhancers, the principles and mechanisms by which enhancers select and control their target genes remain largely unknown. Here, we review recent interdisciplinary and quantitative approaches based on emerging techniques that aim to address open questions in the field, notably how regulatory information is encoded in the DNA sequence, how this information is transferred from enhancers to promoters, and how these processes are regulated in time.

Addresses

¹ Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

² University of Basel, Basel, Switzerland

Corresponding author: Giorgetti, Luca (luca.giorgetti@fmi.ch)

* Twitter account: [@MachPia](https://twitter.com/MachPia), [@LucaGiorgetti](https://twitter.com/LucaGiorgetti)

Current Opinion in Genetics & Development 2023, 80:102052

This review comes from a themed issue on **Genome Architecture and Expression**

Edited by **Giacomo Cavalli** and **Job Dekker**

For complete overview of the section, please refer to the article collection, “[Genome Architecture and Expression \(2023\)](#)”

Available online 29 May 2023

<https://doi.org/10.1016/j.gde.2023.102052>

0959–437X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

In metazoans, the spatiotemporal control of gene expression relies not only on promoters but also, and crucially so, on enhancers. These noncoding regulatory sequences are located outside promoter regions and activate target genes often over large genomic distances, which in mammals can extend over the megabase scale [1]. Despite the intervening distance, the regulatory role of enhancers is essential for accurate tissue- and developmental time-specific gene expression patterns.

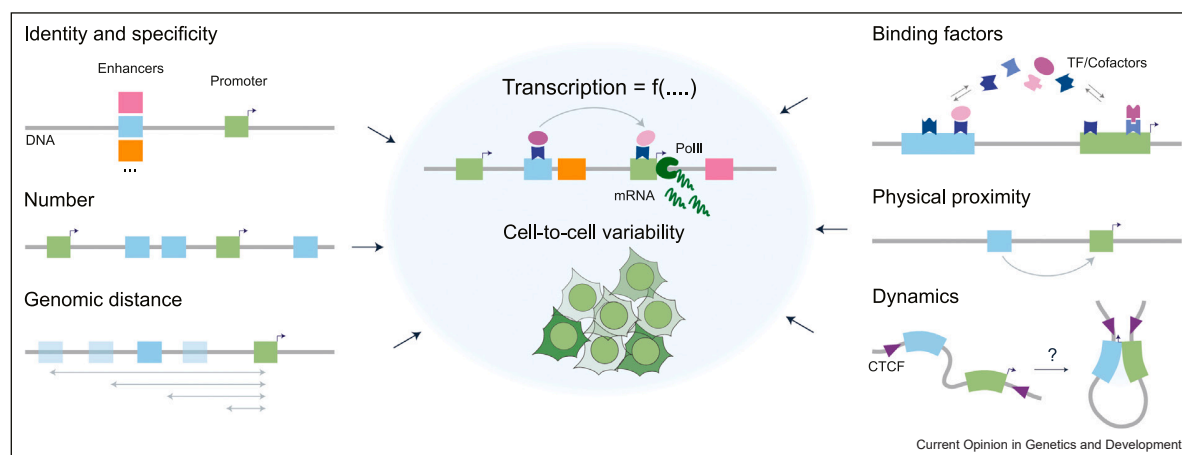
Genetic variation within these noncoding regions results in quantitative changes in gene expression and is a major driver of evolution, but also causal to developmental disorders and strongly associated with numerous human diseases [2]. However, despite our increasing ability to identify putative enhancer sequences on the basis of functional genomic annotations and genome-wide association studies, to date, it remains impossible to predict and causally explain distal genetic associations between enhancers and their target promoters from first principles. How enhancers control a promoter’s transcriptional level, and how this depends on their sequence, number, and genomic distance, remains to be understood, as well as how these dependencies relate to binding and interactions of transcription factors (TFs) and cofactors, as well as on modulations of enhancer–promoter (E–P) physical interactions (Figure 1).

Several excellent reviews have covered recent progress in enhancer biology [1,3,4]. Here, we focus specifically on emerging evidence from highly diverse approaches, which strongly suggest that a truly quantitative understanding of E–P communication requires understanding the spatiotemporal dynamics of the underlying molecular processes, which are highly stochastic and dynamic in individual living cells. We notably review technical and methodological developments bridging across biochemistry, molecular biology, and biophysics that have recently enabled exciting insights, and outline future challenges.

How are regulatory information and specificity encoded in enhancer and promoter sequences?

One key question is how an enhancer’s regulatory activity and its specificity for selected promoters are encoded in its DNA sequence. While it is well established that the activity of an enhancer depends on its ability to recruit combinations of TFs [5], it remains unclear how enhancer activity is modulated by TF binding site (TFBS) number, affinity, and arrangement (i.e. enhancer grammar) [6], and how it depends on covalent DNA or histone modifications as well as TF interactions with cofactors. The last few years have seen major advances in methods for the identification of putative regulatory sequences genome-wide [7] (some of which can be deployed at single-cell resolution [8]), the synthesis and/or mutagenesis of endogenous or ectopic sequences [9], and the massively parallel detection of their effects on transcriptional

Figure 1



Enhancer-promoter (E-P) communication and transcriptional outputs depend on a number of variables, including enhancer and promoter sequence, number, and mutual genomic distance, in a way that depends on binding of TFs and interactions with cofactors, the physical proximity between regulatory sequences and their dynamics.

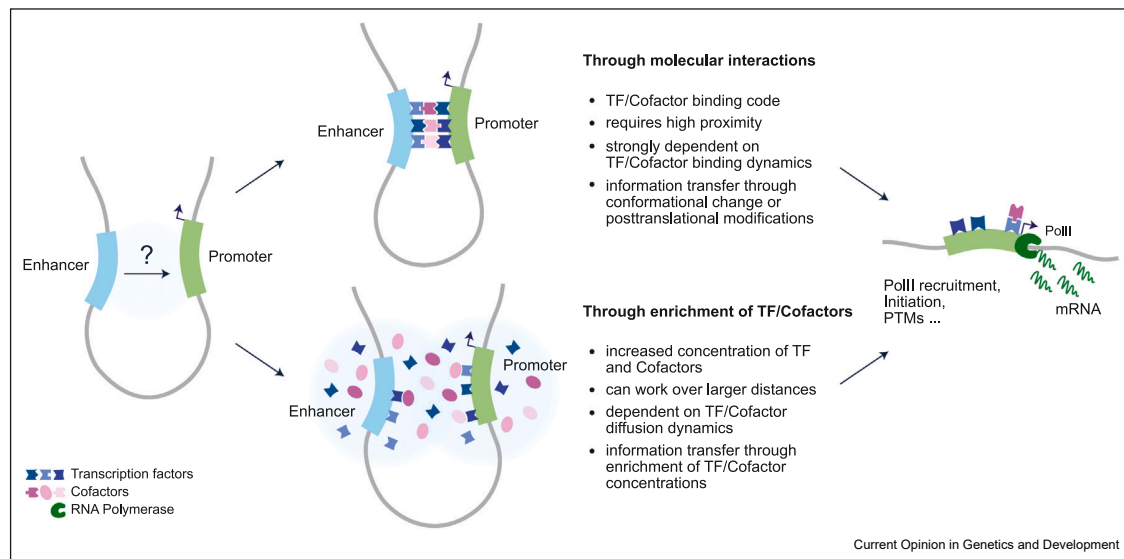
reporters [10]. Coupled with machine learning algorithms, these methods have given considerable insights into TFBS usage within enhancer regions and at least in some contexts have enabled the design of artificial enhancers with tissue-specific activity [11,12].

While such approaches have unlocked the potential to reveal quantitative sequence determinants of intrinsic enhancer activity, it remains unclear if an enhancer can activate any promoter, or rather only a compatible subset of promoters. Answering this fundamental question requires measuring the transcriptional output of large numbers of E-P combinations in parallel. Initial experiments based on transiently expressed combinatorial libraries of enhancer-reporter pairs have suggested that housekeeping and developmentally regulated promoters are activated by distinct sets of enhancers in *D. melanogaster* cells [13]. However, in mammalian cells, a rather broad compatibility was observed, with quantitative rather than qualitative differences in E-P communication [14–16]. This is in line with the earlier finding that once randomly inserted into the mouse genome, ectopic promoter sequences are activated by the (unrelated) surrounding endogenous-regulatory landscapes [17,18]. To what extent these results can be extrapolated into general principles of E-P communication in a chromosomal context is unclear. Moving the field forward will critically depend on the development of large-scale reporter assays allowing to study E-P communication in an endogenous chromatin environment, possibly taking advantage of experimental setups where genomic sequence can be engineered bottom-up [19–21] and chromatin states and/or recruitment of TFs and cofactors can be artificially modulated [22–28]. Coupled to increasingly more refined

‘explainable’ artificial intelligence approaches [29], such experiments will allow disentangling regulatory layers and unravel quantitative contributions of enhancer/promoter sequence, TFBS usage, and chromatin-mediated effects to E-P communication.

Support for a critical role of chromatin effects in E-P communication is also supported by recent studies that measured TF and nucleosome occupancy with single-molecule resolution using DNA methyltransferase footprinting followed by high-throughput sequencing. These studies revealed that nucleosome eviction following TF binding to enhancer DNA often results in indirect binding cooperativity (e.g. not mediated by direct protein–protein interactions) [30,31]. The single-molecule resolution of such assays was also key to demonstrate that CpG methylation within enhancer regions directly prevents binding of at least some TFs [32]. Coupled to long-read DNA sequencing, single-molecule footprinting methods hold great promise for revealing correlations between TFBS occupancy at distal promoter and enhancer sites [33] and downstream transcription events [34], and even inferring their temporal order [35]. It will be exciting to see how such assays can be combined in the future with atomic models of molecular interactions provided by cryo-electron microscopy (cryo-EM) [36] as well as fluorescent microscopy-based measurements of the rates and stoichiometries of such interactions in single living cells [37,38] to gain insight into the mechanistic bases of sequence determinants of E-P communication. Ultimately, studying E-P pairs in a genomic context will also be fundamental to reveal how their communication is influenced by the intervening genomic distance and their numbers [39]. This will lead to an

Figure 2



Two alternative (and not mutually exclusive) potential mechanisms of information transfer from enhancers to promoters. Information might be passed on to the promoter through direct interactions between TFs and cofactors, or through local enrichment of TFs and cofactors in the vicinity of the promoter. Both mechanisms must result in the transfer of information being used by the promoter to enhance PolII recruitment, post-transcriptional modifications, or events leading to transcriptional initiation or elongation (PTM=post-translational modifications).

understanding of how E–P interactions are modulated by additional regulatory elements and chromosome structure (see below).

How is regulatory information conveyed from enhancers to promoters?

The exchange of regulatory information between an enhancer and a promoter is thought to require at least some degree of physical proximity in the three-dimensional (3D) space of the cell nucleus. E–P communication is indeed generally restricted within topologically associating domains (TAD) [40,41], which are established as the loop-extrusion activity of the cohesin complex is arrested by CTCF molecules bound to DNA in a specific orientation [42]. Compartment-like interactions as well as cohesin/CTCF-independent tethering elements also contribute to establishing physical associations between active regulatory sequences [43–45]. The notion that proximity is required for conveying regulatory information is also supported by the fact that E–P interaction probabilities measured with cross-linking-based chromosome conformation capture (3C) correlate quantitatively with the degree of promoter activity [46–48] and that cohesin-mediated CTCF loops can often modulate E–P communication [46,49,50], although the effect of TAD boundaries and CTCF sites might depend on the actual intrinsic activity of the enhancer [46,51] or compatibility with the promoter.

However, what level of ‘proximity’ is required and what molecular processes actually mediate regulatory exchange remains unclear. Spatial proximity is usually defined as the ability to form ligation products in 3C methods, but it is unclear at what spatial distance two genomic loci can be cross-linked and ligated, and whether this distance depends on local chromatin composition (e.g. active or repressive chromatin) [52]. Measurements of proximity-induced DNA methylation have shown that at least a fraction of cross-linking events detected by 3C methods occur in a range of a few nanometers [53]. Absolute quantification of physical distances between regulatory sequences may be enabled by optical methods, but is complicated by experimental uncertainty on distance measurements in fluorescence microscopy (approx. 50 nm in single-dye sequential FISH [54–56]), which currently exceeds the scale of molecular interactions [57]. Indeed, although measurements of physical proximity based on 3C and optical methods often correlate well, discrepancies have been observed [58]. Recent developments in super-resolution microscopy start to bridge this gap achieving a localization precision in the range of 1 nm [59] and promise to yield more precise estimations of molecular proximity in the near future. However, since the mechanisms and factors that mediate E–P communication remain unclear, so do the physical scales that are the most relevant to study. If activation is mediated by direct TF-cofactor–RNA polymerase II (PolII) interactions, these are

expected to take place in the few tens-of-nanometer range (as a reference, a single human Mediator/pre-initiation complex spans $\sim 25 \times 10 \times 10$ nm based on recent cryo-EM structures [60]). If instead communication occurs through macromolecular complexes, phase-separated droplets nucleated by low-complexity protein domains [61,62], or local diffusion of (post-translationally modified) transcriptional coactivators [63], it is possible that distances involved are much larger (in the range of hundreds of nm) (Figure 2). Both mechanisms, irrespective of whether they co-occur or are mutually exclusive, must result in transmission of information to the promoter that ultimately leads to PolIII licensing into productive transcription [5].

Identifying cofactors involved in E–P communication (and in transcriptional regulation in general) has proven difficult because contrary to TFs, coactivators and corepressors do not bind DNA in a sequence-specific manner and no specific ‘interaction code’ with TFs has been identified so far [1]. The last few years have nonetheless seen exciting progress in the identification and characterization of transcriptional cofactors. Genetic screens and new methods for the multiplexed recruitment of cofactors [64,65] have revealed large numbers of protein domains with either activating or repressive effects on transcription. The large-scale, unbiased design of these assays has, for example, enabled the discovery of previously unknown amino acid compositional biases in protein fragments that show coactivator activity and a distinctive role of post-translational SUMOylation of transcriptional repressors [65]. Another step ahead in the study of cofactors is the recent development of inducible degradation methods and availability of small-molecule inhibitors, which complement gene knockout approaches by making it possible to study the acute effects of temporary depletion of essential cofactors [66]. Degradation of Mediator subunits was notably key to characterizing their cell-type-specific roles in transcriptional regulation [67] and a role in facilitating E–P interactions [68,69]. Interestingly, depletion and chemical inhibition have shown that chromatin remodeler cofactors, such as SWI/SNF (SWItch/Sucrose Non-Fermentable), not only regulate promoter expression by modulating TFBS accessibility in a TF-specific manner [70,71], but can also regulate chromosome interactions by modulating accessibility of CTCF binding sites [70].

A better characterization of cofactors will likely enable a more functional, rather than purely physical understanding of E–P contacts. This would allow one to tag cofactors specifically involved in long-range communication with enzymes that promote proximity-mediated chemical modifications (e.g. through proximity-mediated biotinylation followed by mass spectrometry [72] or fluorescent complementation assays [73]). This could also account for the temporal aspect of such interactions.

One interesting step in this direction was recently provided by proximity-assisted photoactivation [74], in which energy transfer between two rhodamine dyes allows the fluorescent detection of tagged molecules separated by a distance of few nanometers at single-molecule resolution. Finally, the exciting perspective to visualize physical proximity between genomic locations together with the factors that mediate functional communication might come from spatial sequencing techniques [75], especially combined with methods to fluorescently detect chromatin-binding proteins *in situ* [76]. Recent advances in correlative light and electron microscopy [77] also open up the possibility that functional interactions might become observable at nearly atomic resolution in a not-too-distant future.

How do stochastic molecular processes at enhancer–promoter interfaces mediate transcriptional regulation?

Irrespective of the exact nature of the biochemical reactions and molecular players mediating the transfer of regulatory information, many studies have shown that the physical distances of E–P pairs vary substantially from one cell to another within the same cell population or tissue [78–80]. This suggests that if E–P communication requires physical proximity, it can generally only occur in a subset of cells at any given time. But it is unclear how often these windows of opportunity are created and how long they last in time within individual cells. This ultimately depends on how fast chromosome structure changes, especially within TADs where most E–P interactions occur in mammalian cells.

Since TADs are established by the loop-extrusion activity of cohesin, the dynamics of chromosome structure inside a TAD is expected to be linked to the kinetics of loop extrusion and of CTCF binding. A series of recent exciting studies based on quantitative microscopy of reconstituted protein complexes demonstrated that mammalian cohesin indeed extrudes loops of DNA and can be blocked by CTCF, and even measured rates of extrusion *in vitro* (~ 1 kb/s) [81–83]. Single-molecule tracking and analysis of protein mobility *in vivo* [84,85] additionally suggested that cohesin resides on DNA for 5–20 min on average. Together, these studies suggested that the conformation of the chromatin fiber within a TAD might vary considerably over the timescale of minutes in living cells under the action of cohesin-mediated extrusion.

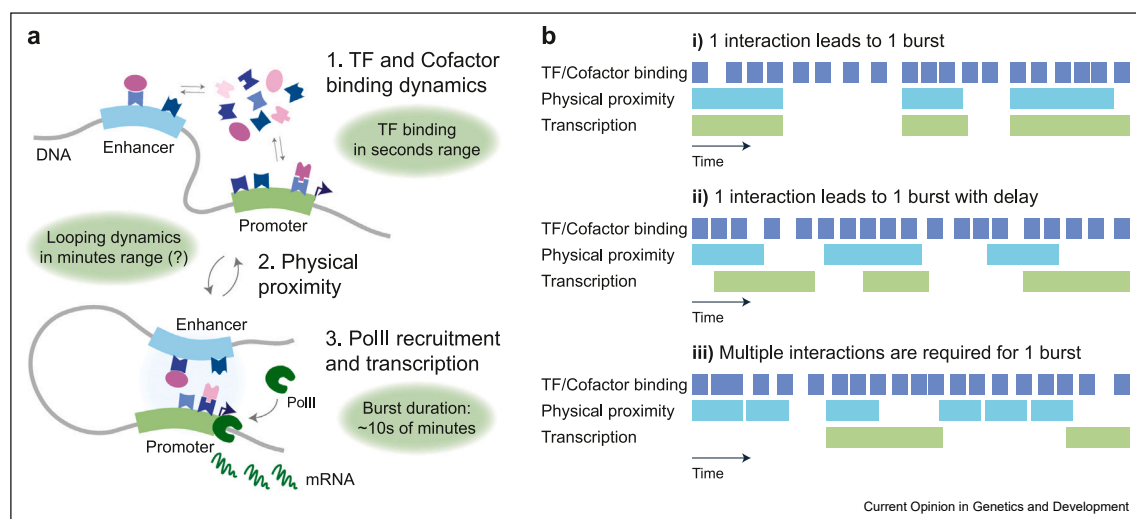
Analysis of the motion of genomic locations [86–89] or chromatin-bound proteins [90] initially suggested that chromatin motion is subdiffusive and that promoters or enhancers explore the surrounding nuclear space in a way that might depend on their activity [86,87]. Only very recently, live-cell imaging studies provided insights

into how CTCF and cohesin control the looping dynamics of chromosomal sequences located within or close to TAD boundaries. Analysis of the dynamics of pairs of genomic loci in mouse embryonic stem cells revealed that cohesin-mediated loops between pairs of convergent CTCF sites last ~5–30 min on average [91,92], which is a relatively small fraction of the cell cycle duration (~12 hours in this cell type). One study estimated that in the absence of CTCF sites, two sequences within a TAD only spend around 5 min on average within ~150 nm from each other (i.e. the upper range of estimated E–P communication distances, see above) [92]. Together with other recent studies of chromatin motion [93,94], these new results suggest that windows of opportunities for regulatory exchange might be transient events occurring as enhancers and promoters explore the nuclear space in the context of a highly dynamic chromatin fiber.

Interestingly, these experimental estimates for the timescales of CTCF looping and physical proximity represent an intermediate level between those of nucleosome remodeling and TF binding/unbinding, which typically occur on the second timescale, and those of promoter bursting, which in mammalian cells takes place over tens of minutes (Figure 3a) [95]. How these substantially disconnected timescales (at least in mammals; this might not be the case in budding yeast [96]) are integrated to elicit the correct transcription levels and

cell-to-cell and temporal variability remains to be understood. Given the fast binding/unbinding kinetics of many TFs and the potentially longer timescales of E–P proximity, it is possible that time-averaged TFBS occupancies within enhancer and promoter regions are sufficient to enable the exchange of regulatory information throughout the duration of a contact. But even if this is the case, does a burst of transcription initiate every time that an enhancer is close enough to its target promoter that their transiently bound TFs can interact (Figure 3b, i)? Or does this happen with a delay, possibly due to the kinetics of intermediate regulatory steps (e.g. assembly or interactions with cofactors) (Figure 3b, ii)? Or rather does a burst of transcription only initiate after several contacts, for example, because of cooperativity in one or more sequential events occurring at the promoter (Figure 3b, iii)? Is every E–P pair actually expected to behave the same, or do alternative modes of signal integration [97] result in different behaviors at different E–P pairs? And how is this related to bursting kinetics at the promoter? Although the 1-contact-1-burst scenario has been observed in live *Drosophila* embryos, it is likely that it was favored by ectopically induced, unusually stable E–P interactions [98]. Live-cell imaging experiments in mammalian cells suggested a total absence of correlation between E–P and transcription bursts [99], but optical resolution limits discussed above urge caution on the interpretation of these results. Theoretical arguments that are compatible with

Figure 3



Temporal dynamics of E–P communication. **(a)** Transcription initiation and the resulting transcriptional bursts are preceded by multiple processes, notably fast TF and cofactor binding (typically in the second to tens-of-seconds range) and looping of the chromatin fiber that establishes physical proximity, recently suggested to occur on the order of minutes. **(b)** While TF-binding dynamics are fast and occur frequently, possibly enabling continuous occupancy of TFBS at the enhancer and the promoter, E–P proximity is presumably longer and occurs less frequently, leading to potential models of E–P communication in time: i) one interaction might lead to one event of transcription initiation, ii) possibly with a delay, or alternatively more complicated scenarios iii) where one initiation event can only be achieved by multiple consecutive interaction events.

scenarios ii) and iii) have been recently evoked to explain striking nonlinearities in how transcription levels depend on E–P contact probabilities [46,100], although none has been formally proven. The rapid advance of super-resolution microscopy techniques that are compatible with live-cell approaches [101,102] will likely provide measurements of E–P proximity with increased accuracy together with single-molecule-resolved binding/unbinding of TFs and cofactors [96], thus allowing for a more precise assessment of the temporal order of events at E–P interfaces over timescales ranging from a fraction of a second to minutes and hours.

We would finally like to note that mathematical and physical models have been instrumental to push forward our understanding of how stochastic and dynamic regulatory processes in single cells convert regulatory inputs into transcriptional outputs. To cite just a few, models have provided mechanistic bases for the processes that shape chromosome organization and dynamics [42] as well as alternative hypotheses on how transcription bursts are generated [97] and connected to E–P interactions [46,100]. Further, they have allowed to quantify how regulatory information is processed by E–P modules into cell states [103]. We expect that the role of models will become even more prominent in the future as the data in the field become more and more quantitative.

Conclusions and outlook

Understanding how genomic sequence determines gene expression programs and their dynamics during development and homeostasis relies on our ability to understand E–P communication from a quantitative point of view, that is, to a level that allows us to predict how sequence mutations translate into changes of transcription levels and their cell-to-cell variability. Over the last few years, we have witnessed spectacular progress in this direction, thanks to the fast-paced rate of technological development and the advent of quantitative methods for the measurements of biological processes. It will be exciting to see how these methods can be coupled to another emerging family of approaches relying on the ‘bottom-up’ generation of regulatory landscapes with controllable levels of complexity [20,21,46,47]. By reducing the number of biological degrees of freedom to an amenable amount, such genome engineering approaches have the potential to disentangle regulatory layers and push the field forward toward the identification of mechanistic regulatory principles of E–P communication. We very much look forward to the next phase of experimental and theoretical studies, which holds great promise to open exciting insights into the molecular mechanisms of long-range transcriptional regulation.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Ilya Flyamer for critical reading of the paper. P.M. was supported by a Marie Skłodowska-Curie Innovative Training Network (Grant no. 813327 ‘ChromDesign’). Research in the Giorgetti laboratory is funded by the Novartis Foundation, the European Research Council (Grant no. 759366, ‘BioMcTre’), Marie Skłodowska-Curie Innovative Training Networks (Grant nos. 813327 ‘ChromDesign’ and 813282 ‘PEP-NET’) under the European Union’s Horizon 2020 research and innovation program, and the Swiss National Science Foundation (Grant no. 310030_192642).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Kim S, Wysocka J: **Deciphering the multi-scale, quantitative cis-regulatory code.** *Mol Cell* 2023, **83**:373-392.
2. Claringbould A, Zaugg JB: **Enhancers in disease: molecular basis and emerging treatment strategies.** *Trends Mol Med* 2021, **27**:1060-1073.
3. Karpinska MA, Oudelaar AM: **The role of loop extrusion in enhancer-mediated gene activation.** *Curr Opin Genet Dev* 2023, **79**:102022.
4. Pachano T, Haro E, Rada-Iglesias A: **Enhancer-gene specificity in development and disease.** *Development* 2022, **149**:dev186536.
5. Field A, Adelman K: **Evaluating enhancer function and transcription.** *Annu Rev Biochem* 2020, **89**:213-234.
6. Jindal GA, Farley EK: **Enhancer grammar in development, evolution, and disease: dependencies and interplay.** *Dev Cell* 2021, **56**:575-587.
7. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, Jackson D, Leith A, Schreiber J, Noble WS, et al.: **A genome-wide framework for mapping gene regulation via cellular genetic screens.** *Cell* 2019, **176**:377-390 .e19.
8. Clark SJ, Argelaguet R, Kapourani C-A, Stubbs TM, Lee HJ, Alda-Catalinas C, Krueger F, Sanguinetti G, Kelsey G, Marioni JC, et al.: **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells.** *Nat Commun* 2018, **9**:781.
9. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J: **Saturation editing of genomic regions by multiplex homology-directed repair.** *Nature* 2014, **513**:120-123.
10. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science* 2013, **339**:1074-1077.

11. de Almeida BP, Reiter F, Pagani M, Stark A: **DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers.** *Nat Genet* 2022, **54**:613-624.
- Together with Ref. [12], this study provides a striking example of deep-learning guided strategy for enhancer grammar inference. Learning was done on STARR-seq libraries and generated models were then used to inform the design of synthetic enhancer sequences that were experimentally tested for transcriptional activity in *Drosophila* cells.
12. Taskiran II, Spanier KI, Christiaens V, Mauduit D, Aerts S: **Cell type directed design of synthetic enhancers.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.07.26.501466>.
- Together with Ref. [11], this study provides a striking example of deep-learning guided strategy for enhancer grammar inference. This study employed deep learning approaches to learn transcription levels from endogenous DNA sequence in *Drosophila*. Models were then used to inform the design of short (~50 bp) synthetic enhancers whose functionality can be modulated in strength and cell-type specificity.
13. Haberle V, Arnold CD, Pagani M, Rath M, Schernhuber K, Stark A: **Transcriptional cofactors display specificity for distinct types of core promoters.** *Nature* 2019, **570**:122-126.
14. Sahu B, Hartonen T, Pihlajamaa P, Wei B, Dave K, Zhu F, Kaasinen E, Lidschreiber K, Lidschreiber M, Daub CO, et al.: **Sequence determinants of human gene regulatory elements.** *Nat Genet* 2022, **54**:283-294.
- This paper reports massive parallel reporter assays with fully synthetic enhancer and promoter sequences. They found that spacing between TFBS, TFBS motif, and orientation, as well as their combination, is a weak determinant of the activity of single TFs, but that many TFs interact in an additive manner. This study also argues against a model of E-P communication that is mediated by specific TF-TF interactions.
15. Martinez-Ara M, Comoglio F, van Arensbergen J, van Steensel B: **Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome.** *Mol Cell* 2022, **82**:2519-2531 e6.
- Here, E-P compatibility rules in mammalian cells were investigated with a high-throughput combinatorial reporter assay that samples thousands of E-P combinations from a 3-Mb region in the mouse genome, revealing quantitative differences in compatibility levels. Interestingly, comparing the compatibility of individual E-P pairs to their contact probability in the mouse genome revealed no correlation, suggesting that genome topology and E-P compatibility are independent mechanisms of transcriptional regulation.
16. Bergman DT, Jones TR, Liu V, Ray J, Jagoda E, Siraj L, Kang HY, Nasser J, Kane M, Rios A, et al.: **Compatibility rules of human enhancer and promoter sequences.** *Nature* 2022, **607**:176-184.
- Combinations of 1'000 × 1'000 human enhancer and promoter fragments were assayed in massively parallel reporter assays. Most of the transcriptional activity of E-P pairs could be explained by the product of individual enhancer and promoter intrinsic activities, and broad levels of E-P compatibility were detected, with only mild sequence specificity.
17. Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F: **Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor.** *Nat Genet* 2011, **43**:379-386.
18. Anderson E, Devenney PS, Hill RE, Lettice LA: **Mapping the Shh long-range regulatory domain.** *Development* 2014, **141**:3934-3943.
19. Galupa R, Alvarez-Canales G, Borst NO, Fuqua T, Gandara L, Misunou N, Richter K, Alves MRP, Karumbi E, Perkins ML, et al.: **Enhancer architecture and chromatin accessibility constrain phenotypic space during *Drosophila* development.** *Dev Cell* 2023, **58**:51-62 e4..
20. Brosh R, Coelho C, Ribeiro-dos-Santos AM, Ellis G, Hogan MS, Ashe HJ, Somogyi N, Ordoñez R, Luther RD, Huang E, et al.: **Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus.** *Mol Cell* 2023, **7**:1140-1152 e7.
21. Blayney JW, Kassouf M, Francis H, Camellato BR, Stolper R, Higgs DR, Mitchell L, Boeke J: **Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.06.20.496856>
22. Policarpi C, Munafò M, Tsagkris S, Carlini V, Hackett JA: **Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.09.04.506519>
23. Alerasool N, Leng H, Lin Z-Y, Gingras A-C, Taipale M: **Identification and functional characterization of transcriptional activators in human cells.** *Mol Cell* 2022, **82**:677-695 e7.
24. Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Boija A, Henninger JE, Li CH, Stejskal K, Lin G, et al.: **Differential cofactor dependencies define distinct types of human enhancers.** *Nature* 2022, **606**:406-413.
25. Mukund AX, Tycko J, Allen SJ, Robinson SA, Andrews C, Ludwig CH, Spees K, Bassik MC, Bintu L: **High-throughput functional characterization of combinations of transcriptional activators and repressors.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.12.20.521091>
26. Naqvi S, Kim S, Hoskens H, Matthews HS, Spritz RA, Klein OD, Hallgrímsson B, Swigut T, Claes P, Pritchard JK, et al.: **Precise modulation of transcription factor levels reveals drivers of dosage sensitivity.** *Nat Genet* 2023, **55**:841-851.
27. Noviello G, Gjaltema RAF, Schulz EG: **CasTuner: a degranon and CRISPR/Cas-based toolkit for analog tuning of endogenous gene expression.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.10.05.511019>.
- The method developed here allows the fine-tuning of endogenous gene expression by targeting dCas9 fused to FKBP degranon domain to endogenous promoters. By titrating the small molecule dTAG-13 to induce degradation of the FKBP-fusion protein, expression levels at targeted promoters can be tuned with a high homogeneity across a cell population.
28. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, et al.: **The interaction landscape between transcription factors and the nucleosome.** *Nature* 2018, **562**:76-81.
29. Novakovskiy G, Dexter N, Libbrecht MW, Wasserman WW, Mostafavi S: **Obtaining genetics insights from deep learning via explainable artificial intelligence.** *Nat Rev Genet* 2023, **24**:125-137.
30. Sönmez C, Kleinendorst R, Imanci D, Barzaghi G, Villacorta L, Schübeler D, Benes V, Molina N, Krebs AR: **Molecular co-occupancy identifies transcription factor binding cooperativity in vivo.** *Mol Cell* 2021, **81**:255-267 e6..
31. Rao S, Ahmad K, Ramachandran S: **Cooperative binding between distant transcription factors is a hallmark of active enhancers.** *Mol Cell* 2021, **81**:1651-1665 e4..
32. Kreibich E, Kleinendorst R, Barzaghi G, Kaspar S, Krebs AR: **Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation.** *Mol Cell* 2023, **83**:787-802 e9.
- Dual-enzyme single-molecule footprinting performed on endogenous mouse regulatory sequences demonstrated that although the majority of enhancers is not sensitive to DNA methylation, CpG methylation can directly inhibit the binding of several TFs whose TFBS define a class of methylation-sensitive regulatory elements.
33. Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA: **Single-molecule regulatory architectures captured by chromatin fiber sequencing.** *Science* 2020, **368**:1449-1454.
34. Krebs AR, Imanci D, Hoerner L, Gaidatzis D, Burger L, Schübeler D: **Genome-wide single-molecule footprinting reveals high RNA Polymerase II turnover at paused promoters.** *Mol Cell* 2017, **67**:411-422 e4..
35. Battaglia S, Dong K, Wu J, Chen Z, Najm FJ, Zhang Y, Moore MM, Hecht V, Shores N, Bernstein BE: **Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements.** *Nat Genet* 2022, **54**:1504-1513.
- This study uses single-molecule methylation footprinting coupled with Cas9-mediated sequence enrichment followed by Nanopore sequencing to profile open chromatin regions across immune loci of human

8 Genome Architecture and Expression

cells. Long read length (~100 kb) generated allows interrogation of correlation between activities of distal regulatory elements, such as enhancers and promoters

36. Michael AK, Thomä NH: **Reading the chromatinized genome.** *Cell* 2021, **184**:3599-3611.
 37. Li J, Dong A, Saydaminova K, Chang H, Wang G, Ochiai H, Yamamoto T, Pertsinidis A: **Single-molecule nanoscopy elucidates RNA Polymerase II transcription at single genes in live cells.** *Cell* 2019, **178**:491-506 e428.
 38. Balzarotti F, Eilers Y, Gwosch KC, Gynnå AH, Westphal V, Stefani FD, Elf J, Hell SW: **Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes.** *Science* 2017, **355**:606-612.
 39. Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al.: **Enhancer redundancy provides phenotypic robustness in mammalian development.** *Nature* 2018, **554**:239-243.
 40. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions.** *Nature* 2012, **485**:376-380.
 41. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al.: **Spatial partitioning of the regulatory landscape of the X-inactivation centre.** *Nature* 2012, **485**:381-385.
 42. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA: **Formation of chromosomal domains by loop extrusion.** *Cell Rep* 2016, **15**:2038-2049.
 43. Hsieh T-HS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R: **Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1.** *Nat Genet* 2022, **54**:1919-1932.
 44. Batut PJ, Bing XY, Sisco Z, Raimundo J, Levo M, Levine MS: **Genome organization controls transcriptional dynamics during development.** *Science* 2022, **375**:566-570.
 45. Chen L-F, Long HK, Park M, Swigut T, Boettiger AN, Wysocka J: **Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus.** *Mol Cell* 2023, **83**:1446-1461.e6.
 46. Zuin J, Roth G, Zhan Y, Cramard J, Redolfi J, Piskadlo E, Mach P, Kryzhanovska M, Tihanyi G, Kohler H, et al.: **Nonlinear control of transcription through enhancer-promoter interactions.** *Nature* 2022, **604**:571-577.
 47. Rinzema NJ, Sofiadis K, Tjalsma SJD, Verstegen MJAM, Oz Y, Valdes-Quezada C, Felder A-K, Filipovska T, van der Elst S, de Andrade dos Ramos Z, 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* 2022, **29**:563-574.
 48. Kane L, Williamson I, Flyamer IM, Kumar Y, Hill RE, Lettice LA, Bickmore WA: **Cohesin is required for long-range enhancer action at the Shh locus.** *Nat Struct Mol Biol* 2022, **29**:891-897.
 49. Chakraborty S, Kopitchinski N, Zuo Z, Eraso A, Awasthi P, Chari R, Mitra A, Tobias IC, Moorthy SD, Dale RK, et al.: **Enhancer-promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness.** *Nat Genet* 2023, **55**:280-290.
 50. Huang H, Zhu Q, Jussila A, Han Y, Bintu B, Kern C, Conte M, Zhang Y, Bianco S, Chiariello AM, et al.: **CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains.** *Nat Genet* 2021, **53**:1064-1074.
 51. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al.: **Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations.** *Nat Genet* 2019, **51**:1664-1669.
 52. Gavrilov A, Razin SV, Cavalli G: **In vivo formaldehyde cross-linking: it is time for black box analysis.** *Brief Funct Genom* 2015, **14**:163-165.
 53. Redolfi J, Zhan Y, Valdes-Quezada C, Kryzhanovska M, Guerreiro I, Iesmantavicius V, Pollex T, Grand RS, Mulugeta E, Kind J, et al.: **DamC reveals principles of chromatin folding in vivo without crosslinking and ligation.** *Nat Struct Mol Biol* 2019, **26**:471-480.
 54. Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X: **Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells.** *Science* 2018, **362**:eaau1783.
 55. Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN: **Visualizing DNA folding and RNA in embryos at single-cell resolution.** *Nature* 2019, **568**:49-54.
 56. Gizzi AMC, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, Houbroun C, Ogiyama Y, Papadopoulos GL, Cavalli G, et al.: **Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms.** *Mol Cell* 2019, **74**:212-222 e5..
 57. Brandão HB, Gabriele M, Hansen AS: **Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging.** *Curr Opin Cell Biol* 2021, **70**:18-26.
 58. Acuña LIG, Flyamer I, Boyle S, Friman ET, Bickmore WA: **Enhancer-promoter contact frequencies and physical proximity.** *bioRxiv* 2023, <https://doi.org/10.1101/2023.03.29.534720>
 59. Ostersehl LM, Jans DC, Wittek A, Keller-Findeisen J, Inamdar K, Sahl SJ, Hell SW, Jakobs S: **DNA-PAINT MINFLUX nanoscopy.** *Nat Methods* 2022, **19**:1072-1075.
 60. Rengachari S, Schilbach S, Aibara S, Dienemann C, Cramer P: **Structure of the human Mediator-RNA polymerase II pre-initiation complex.** *Nature* 2021, **594**:129-133.
 61. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA, Phase A: **Separation model for transcriptional control.** *Cell* 2017, **169**:13-23.
 62. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al.: **Imaging dynamic and selective low-complexity domain interactions that control gene transcription.** *Science* 2018, **361**:eaar2555.
 63. Karr JP, Ferrie JJ, Tjian R, Darzacq X: **The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication.** *Genes Dev* 2022, **36**:7-16.
 64. Tycko J, DelRosso N, Hess GT, Aradhana, Banerjee A, Mukund A, Van MV, Ego BK, Yao D, Spees K, et al.: **High-throughput discovery and characterization of human transcriptional effectors.** *Cell* 2020, **183**:2020-2035 e16..
 65. DelRosso N, Tycko J, Suzuki P, Andrews C, Aradhana, Mukund A, Liongson I, Ludwig C, Spees K, Fordyce P, et al.: **Large-scale mapping and mutagenesis of human transcriptional effector domains.** *Nature* 2023, **616**:365-372.
- Highly parallel assays where protein fragments are recruited to a reporter promoter and their sequences are identified by high-throughput sequencing revealed amino acid sequence that can activate or repress transcription. Sequence analysis revealed composition biases of activators and repressors and a widespread role of SUMOylation in transcriptional repressors.
66. de Wit E, Nora EP: **New insights into genome folding by loop extrusion from inducible degron technologies.** *Nat Rev Genet* 2022, **24**:73-85.
 67. Jaeger MG, Schwalb B, Mackowiak SD, Velychko T, Hanzl A, Imrichova H, Brand M, Agerer B, Chorn S, Nabet B, et al.: **Selective Mediator dependence of cell-type-specifying transcription.** *Nat Genet* 2020, **52**:719-727.
 68. Haarhuis JHI, van der Weide RH, Blomen VA, Flach KD, Teunissen H, Willems L, Brummelkamp TR, Rowland BD, de Wit E: **A Mediator-cohesin axis controls heterochromatin domain formation.** *Nat Commun* 2022, **13**:754.

69. Ramasamy S, Aljahani A, Karpinska M, Cao N, Cruz N, Oudelaar AM: **The Mediator complex regulates enhancer-promoter interactions.** *bioRxiv* 2022, <https://doi.org/10.1101/2022.06.15.496245>
70. Barisic D, Stadler MB, Iurlaro M, Schübeler D: **Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors.** *Nature* 2019, **569**:136-140.
71. Martin BJE, Ablondi EF, Goglia C, Adelman K: **Global identification of direct SWI/SNF targets reveals compensation by EP400.** *bioRxiv* 2023, <https://doi.org/10.1101/2023.03.07.531379>
72. Göös H, Kinnunen M, Salokas K, Tan Z, Liu X, Yadav L, Zhang Q, Wei G-H, Varjosalo M: **Human transcription factor protein interaction networks.** *Nat Commun* 2022, **13**:766.
73. Tebo AG, Gautier A: **A split fluorescent reporter with rapid and reversible complementation.** *Nat Commun* 2019, **10**:2822.
74. Graham TG, Ferrie JJ, Dailey GM, Tjian R, Darzacq X: **Detecting molecular interactions in live-cell single-molecule imaging with proximity-assisted photoactivation (PAPA).** *eLife* 2022, **11**:e76870.
- Extending the detection range of proximity-mediated biotinylation, protein complementation assays and Förster resonance energy transfer, this new technique allows detecting single protein-protein interactions in living cells by fluorescence imaging of a receiver fluorophore that gets activated when in physical proximity with a donor fluorophore.
75. Nguyen HQ, Chatteraj S, Castillo D, Nguyen SC, Nir G, Lioutas A, Hershberg EA, Martins NMC, Reginato PL, Hannan M, et al.: **3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing.** *Nat Methods* 2020, **17**:822-832.
76. Lu T, Ang CE, Zhuang X: **Spatially resolved epigenomic profiling of single cells in complex tissues.** *Cell* 2022, **185**:4448-4464 e17..
77. Müller TG, Zila V, Peters K, Schifferdecker S, Stanic M, Lucic B, Laketa V, Lusic M, Müller B, Kräusslich H-G: **HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells.** *eLife* 2021, **10**:e64776.
78. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.** *Nature* 2013, **502**:59-64.
79. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, Tiana G, Heard E: **Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription.** *Cell* 2014, **157**:950-963.
80. Finn EH, Misteli T: **Molecular basis and biological function of variability in spatial genome organization.** *Science* 2019, **365**:eaaw9498.
81. Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters J-M: **DNA loop extrusion by human cohesin.** *Science* 2019, **366**:1338-1345.
82. Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H: **Human cohesin compacts DNA by loop extrusion.** *Science* 2019, **366**:1345-1349.
83. Golfier S, Quail T, Kimura H, Brugués J: **Cohesin and condensin extrude loops in a cell-cycle dependent manner.** *eLife* 2020, **9**:e53885.
84. Wutz G, Ladurner R St, Hilaire BG, Stocsits RR, Nagasaka K, Pignard B, Sanborn A, Tang W, Várnai C, Ivanov MP, et al.: **ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin-STAG1 from WAPL.** *eLife* 2020, **9**:e52091.
85. Cattoglio C, Pustova I, Walther N, Ho JJ, Hantsche-Grininger M, Inouye CJ, Hossain MJ, Dailey GM, Ellenberg J, Darzacq X, et al.: **Determining cellular CTCF and cohesin abundances to constrain 3D genome models.** *eLife* 2019, **8**:e40164.
86. Gu B, Swigut T, Spencley A, Bauer MR, Chung M, Meyer T, Wysocka J: **Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements.** *Science* 2018, **359**:1050-1055.
87. Germier T, Kocanova S, Walther N, Bancaud A, Shaban HA, Sellou H, Politi AZ, Ellenberg J, Gallardo F, Bystricky K: **Real-time imaging of a single gene reveals transcription-initiated local confinement.** *Biophys J* 2017, **113**:1383-1394.
88. Masui O, Bonnet I, Le Baccon P, Brito I, Pollex T, Murphy N, Hupé P, Barillot E, Belmont AS, Heard E: **Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation.** *Cell* 2011, **145**:447-458.
89. Khanna N, Zhang Y, Lucas JS, Dudko OK, Murre C: **Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions.** *Nat Commun* 2019, **10**:2771.
90. Nozaki T, Imai R, Tanbo M, Nagashima R, Tamura S, Tani T, Joti Y, Tomita M, Hibino K, Kanemaki MT, et al.: **Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging.** *Mol Cell* 2017, **67**:282-293 e7.
91. Gabriele M, Brandão HB, Grosse-Holz S, Jha A, Dailey GM, Cattoglio C, Hsieh T-HS, Mirny L, Zechner C, Hansen AS: **Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging.** *Science* 2022, **376**:496-501.
- Together with Ref. [92], this study presents the first measurements of cohesin-mediated CTCF looping in single living cells. CTCF-anchored loops between sites separated by ~500 kb at an endogenous locus were found to last ~10–30 min, for a total of ~3–6% of time spent in the looped state. These results suggest that CTCF loops are rare and transient compared to the duration of the cell cycle and that the conformation of chromatin inside a TAD is highly dynamic.
92. Mach P, Kos PI, Zhan Y, Cramard J, Gaudin S, Tünnermann J, Marchi E, Eglinger J, Zuin J, Kryzhanovska M, et al.: **Cohesin and CTCF control the dynamics of chromosome folding.** *Nat Genet* 2022, **54**:1907-1918.
- Together with Ref. [91], this study presents the first measurements of cohesin-mediated CTCF looping in single living cells. CTCF-anchored loops spanning 150 kb within an ectopic location were found to last ~5–15 min, for a total of ~20–30% of time spent in the looped state. These results are in excellent agreement with Ref. [85] considering the different genomic distance and CTCF sites and also suggest that the conformation of chromatin inside a TAD is highly dynamic.
93. Keizer VIP, Grosse-Holz S, Woringer M, Zambon L, Aizel K, Bongaerts M, Delille F, Kolar-Znika L, Scolari VF, Hoffmann S, et al.: **Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics.** *Science* 2022, **377**:489-495.
- Using controlled magnetic forces to manipulate the motion of a genetic locus in living mouse stem cells, the authors showed that chromosome mobility in living cells is less influenced by nuclear constraints than previously thought. This enabling technology can potentially be used in the future to study the mechanics of other processes in combination with the motion of a single locus in living cells.
94. Brückner DB, Chen H, Barinov L, Zoller B, Gregor T: **Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome.** *bioRxiv* 2023, <https://doi.org/10.1101/2023.01.18.524527>.
- Live-cell imaging of ectopic E–P pairs in *Drosophila* embryos provides the first measurements of chromosome looping as a function of the intervening genomic distance. Genomic loci were found to undergo fast subdiffusive dynamics, while being constrained to a compact 3D organization of the chromatin fiber. Together with Refs. [85,86], the results point to a striking similarity of chromatin dynamics across organisms.
95. Coulon A, Chow CC, Singer RH, Larson DR: **Eukaryotic transcriptional dynamics: from single molecules to cell populations.** *Nat Rev Genet* 2013, **14**:572-584.
96. Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, Ferguson ML, Lenstra TL: **Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting.** *EMBO J* 2019, **38**:e100809.
97. Lammers NC, Kim YJ, Zhao J, Garcia HG: **A matter of time: using dynamics and theory to uncover mechanisms of transcriptional bursting.** *Curr Opin Cell Biol* 2020, **67**:147-157.
- Here, the authors propose different mathematical models of how the integrations of time scales of molecular processes involved in transcriptional regulation (TF/Cofactor binding, E–P contacts and transcriptional bursting) might result in transcriptional bursting.

10 Genome Architecture and Expression

98. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T: **Dynamic interplay between enhancer-promoter topology and gene activity.** *Nat Genet* 2018, **50**:1296-1303.
99. Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, Huang B, Lomvardas S, Weiner OD: **Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity.** *eLife* 2019, **8**:e41769.
100. Xiao JY, Hafner A, Boettiger AN: **How subtle changes in 3D structure can create large changes in transcription.** *eLife* 2021, **10**:e64320.
101. Deguchi T, Iwanski MK, Schentarra E-M, Heidebrecht C, Schmidt L, Heck J, Weihs T, Schnorrenberg S, Hoess P, Liu S, *et al.*: **Direct observation of motor protein stepping in living cells using MINFLUX.** *Science* 2023, **379**:1010-1015.
MINFLUX nanoscopy was used to track the dynamics of kinesin-1 stepping on microtubules in living cells at the single-molecule level. This is one of the first examples of how single proteins can be followed with nanometer spatial and millisecond temporal resolution in live-cell imaging.
102. Cao B, Coelho S, Li J, Wang G, Pertsinidis A: **Volumetric interferometric lattice light-sheet imaging.** *Nat Biotechnol* 2021, **39**:1385-1393.
103. Tkačik G, Gregor T: **The many bits of positional information.** *Development* 2021, **148**:dev176065.

Cohesin and CTCF control the dynamics of chromosome folding

Received: 4 April 2022

Accepted: 19 October 2022

Published online: 5 December 2022

 Check for updates

Pia Mach^{1,2,9}, Pavel I. Kos^{1,9}, Yinxiu Zhan^{1,9}, Julie Cramard¹, Simon Gaudin^{1,3,4}, Jana Tünnermann^{1,2}, Edoardo Marchi^{5,6}, Jan Eglinger¹, Jessica Zuin¹, Mariya Kryzhanovska¹, Sebastien Smallwood¹, Laurent Gelman¹, Gregory Roth¹, Elphège P. Nora^{7,8}, Guido Tiana^{5,6} & Luca Giorgetti¹✉

In mammals, interactions between sequences within topologically associating domains enable control of gene expression across large genomic distances. Yet it is unknown how frequently such contacts occur, how long they last and how they depend on the dynamics of chromosome folding and loop extrusion activity of cohesin. By imaging chromosomal locations at high spatial and temporal resolution in living cells, we show that interactions within topologically associating domains are transient and occur frequently during the course of a cell cycle. Interactions become more frequent and longer in the presence of convergent CTCF sites, resulting in suppression of variability in chromosome folding across time. Supported by physical models of chromosome dynamics, our data suggest that CTCF-anchored loops last around 10 min. Our results show that long-range transcriptional regulation might rely on transient physical proximity, and that cohesin and CTCF stabilize highly dynamic chromosome structures, facilitating selected subsets of chromosomal interactions.

In mammalian cells, interactions between chromosomal sequences play important roles in fundamental processes such as DNA replication¹, repair² and transcriptional regulation by distal enhancers³. Chromosome conformation capture (3C) methods, which measure physical proximity between genomic sequences in fixed cells, revealed that chromosomal contacts are organized into submegabase domains of preferential interactions known as topologically associating domains (TADs)^{4,5} whose boundaries can functionally insulate regulatory sequences³. TADs mainly arise from nested interactions between convergently oriented binding sites of the DNA-binding protein CTCF, which are established as chromatin-bound CTCF arrests the loop extrusion activity of the cohesin complex^{6–10}.

Determining the timing and duration of chromosomal interactions within TADs and their relationship with CTCF and cohesin is key to understanding how enhancers communicate with promoters^{11,12}. Single-cell analyses of chromosome structure in fixed cells^{4,13–15},

chromosome tracing experiments^{16–19}, in vitro^{9,10,20} and live-cell²¹ measurements of CTCF and cohesin dynamics, and polymer simulations^{6,15,22}, as well as live-cell imaging of chromosomal locations and nascent RNA^{23,24}, all suggested that TADs and CTCF loops are dynamic structures whose temporal evolution might be governed by the kinetics of loop extrusion²⁵. Recent live-cell measurements of a CTCF loop connecting two opposite TAD boundaries in mouse embryonic stem cells (mESCs) provided direct evidence that this is the case, and revealed that cohesin-mediated loops between CTCF sites located 500 kilobases (kb) away last 10–30 min (ref. 26). However, it is still unclear if contacts between sequences separated by genomic distances where enhancers and promoters interact within the same TAD occur on the timescale of seconds, minutes or hours. We also have little knowledge on whether and how rates and durations of such contacts are modulated by loop extrusion. We finally do not know if cohesin increases chromosome mobility and thus favors the encounters between genomic sequences

¹Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. ²University of Basel, Basel, Switzerland. ³École Normale Supérieure de Lyon, Lyon, France. ⁴Université Claude Bernard Lyon I, Lyon, France. ⁵Università degli Studi di Milano, Milan, Italy. ⁶INFN, Milan, Italy. ⁷Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA. ⁸Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA. ⁹These authors contributed equally: Pia Mach, Pavel I. Kos, Yinxiu Zhan. ✉e-mail: luca.giorgetti@fmi.ch

by reeling them into loops, or if instead it provides constraints that decrease mobility and prolong the duration of such encounters. Both scenarios have been suggested to be possible theoretically^{27,28}, but it is unclear which effect dominates in living cells.

Here we use live-cell fluorescence microscopy to measure chromosome dynamics and its dependence on cohesin and CTCF in mESCs. By combining two live-cell imaging strategies with polymer simulations, we reveal that loops extruded by cohesin constrain global chromosome motion, while also increasing the temporal frequencies and durations of physical encounters between sequences inside the same TAD. Convergent CTCF sites substantially stabilize contacts through cohesin-mediated CTCF-anchored loops that last around 5–15 min on average. Our results support the notion that chromosome structure within single TADs is highly dynamic during the span of a cell cycle and thus that long-range transcriptional regulation might rely on transient physical proximity between genomic sequences. They also reveal how contact dynamics and the temporal variability in chromosome folding are modulated by cohesin and CTCF in single living cells and provide a quantitative framework for understanding the role of folding dynamics in fundamental biological processes.

Results

Cohesin decreases chromosome mobility independently of CTCF

To study how cohesin and CTCF influence the global dynamics of the chromatin fiber independently of local chromatin state and structural differences, we examined the dynamic properties of large numbers of random genomic locations in living cells. We generated clonal mESC lines carrying multiple random integrations of an array of ~140 repeats of the bacterial Tet operator sequence (TetO) using piggyBac transposition²⁹. These can be visualized upon binding of Tet repressor (TetR) fused to the red fluorescent protein tdTomato. To compare the motion of genomic locations that either block or allow the loop extrusion activity of cohesin, the TetO array was adjacent to three CTCF motifs (3 × CTCF) that could be removed by Cre-assisted recombination (Fig. 1a). Motifs were selected based on high CTCF enrichment in chromatin immunoprecipitation followed by sequencing (ChIP-seq) and each was confirmed to be bound by CTCF in nearly 100% of alleles at any time in mESCs using dual-enzyme single-molecule footprinting³⁰ (R. Grand and D. Schübeler, personal communication), thus providing a close experimental representative of an ‘impermeable’ loop extrusion barrier.

3 × CTCF-TetO sequences were introduced in mESCs that stably expressed *OsTir1* and where the endogenous *Rad21*, *Wapl* or *Ctcf* genes were targeted with an auxin-inducible degron (AID) peptide fused to eGFP^{31,32}. This resulted in several mESC clones (three per degron condition) with different sets of genomic insertions of the 3 × CTCF-TetO

cassette, where over 95% of any of the AID-tagged proteins could be rapidly depleted upon addition of auxin (Fig. 1b and Extended Data Fig. 1a). This allowed us to study chromosome dynamics following acute depletion of factors affecting cohesin-mediated chromosome structure (Extended Data Fig. 1b) at previously reported time points (90 min for RAD21 (ref. ³³), 6 h for CTCF³¹ and 24 h for WAPL³²) that minimize secondary effects such as defects in cell-cycle progression (Extended Data Fig. 1c).

Mapping TetO insertion sites revealed 10–20 insertions per cell line, with on average 1–2 heterozygous insertions per chromosome without any strong bias towards active or inactive chromatin (Extended Data Fig. 1d,e). Insertions were on average 10 kb away from the nearest endogenous CTCF binding sites (Extended Data Fig. 1f). 4C sequencing (4C-seq) confirmed that insertion of 3 × CTCF-TetO cassettes often led to the formation of ectopic interactions with endogenous CTCF sites, which were lost upon removal of 3 × CTCF sites or depletion of RAD21 (Extended Data Fig. 1g,h).

To measure the dynamics of 3 × CTCF-TetO insertions, we acquired three-dimensional (3D) movies (one z-stack of 10 μm every 10 s for 30 min) using highly inclined and laminated optical sheet microscopy³⁴ (Fig. 1b and Supplementary Video 1). This resulted in ~270 cells per condition with over 8,000 trajectories from three clonal lines imaged with 3–4 biological replicates per condition. Detection and localization of TetO arrays as subdiffraction fluorescent signals³⁵ enabled reconstruction of trajectories of individual genomic insertions (Fig. 1c and Methods). We then studied their mean squared displacement (MSD) as a function of time after correcting each trajectory for the confounding effect of cell movement, which we inferred from the collective displacement of all insertions in each nucleus (Fig. 1d, Extended Data Fig. 2a and Methods). Independently of the degron background, in untreated cells, genomic locations underwent on average a subdiffusive motion whose anomalous exponent (−0.6) and generalized diffusion coefficients (D) ($-1.2 \times 10^{-2} \mu\text{m}^2 \text{s}^{-\alpha}$) were in line with previous studies of specific genomic loci^{36,37} (Fig. 1d and Extended Data Fig. 2b). The MSD of radial distances (radial MSD) between insertions within the same nuclei showed the same scaling although statistics were less robust for long time intervals due to the shorter trajectories that could be built based on pairwise distances (Fig. 1d). Interestingly, removal of 3 × CTCF sites (Extended Data Fig. 1g) or degradation of CTCF (6-h auxin treatment) did not have a significant impact on MSD averaged over all genomic locations nor on its distribution across trajectories and cells (Fig. 1e,f, results for single clones in Extended Data Fig. 2c, P values in Extended Data Fig. 2e).

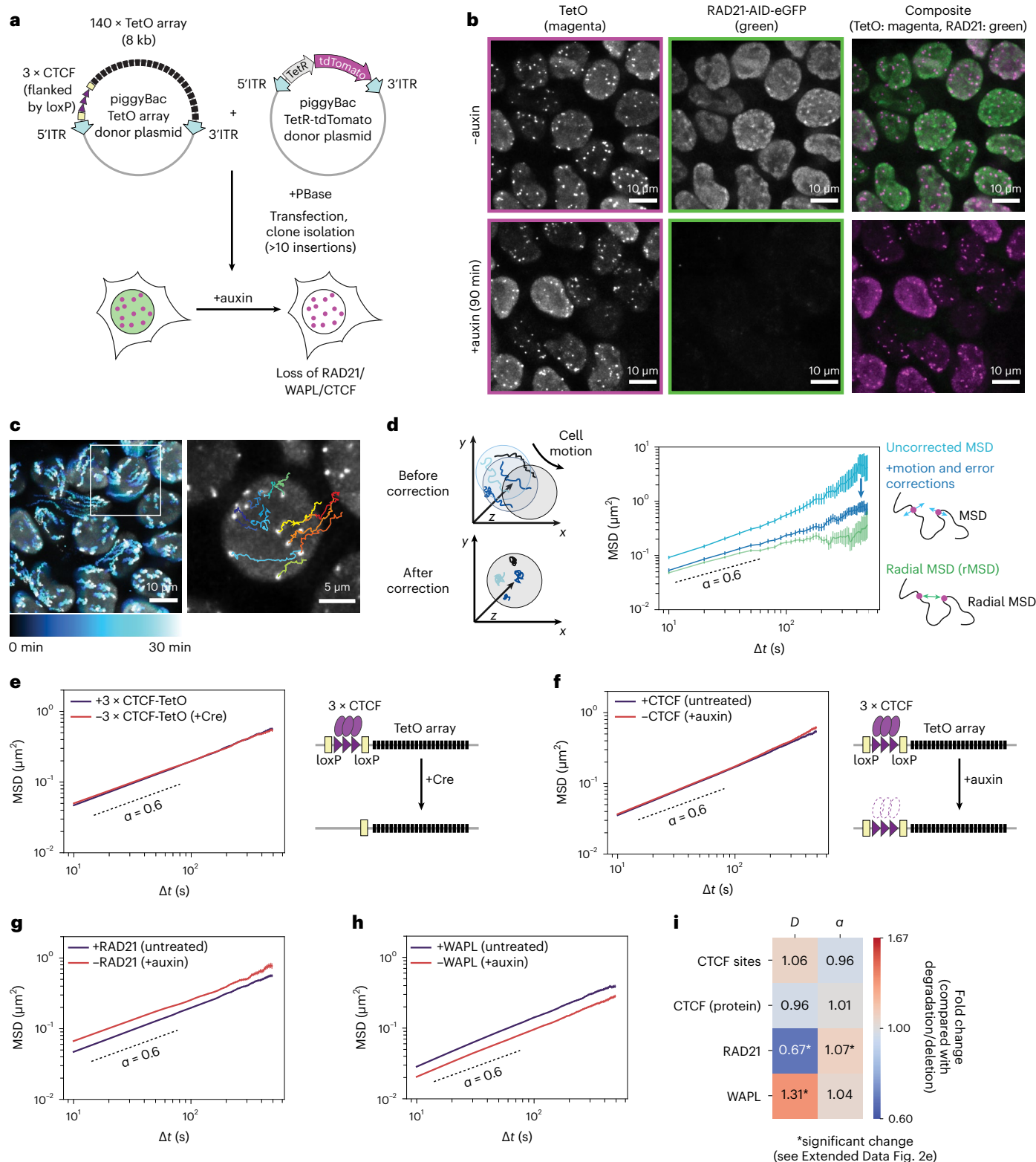
By contrast, acute depletion of RAD21 (90-min auxin treatment) led to a significant increase in mobility both in the presence (Fig. 1g) and absence of 3 × CTCF sites (Extended Data Fig. 2d), with only a very minor impact on anomalous exponents (Extended Data Fig. 2b,e,

Fig. 1 | Cohesin slows down chromosome dynamics in living cells. **a**, Clonal mESC lines containing random TetO arrays flanked by 3 × CTCF motifs and expressing TetR-tdTomato. Constructs were integrated using piggyBac transposition in mESCs allowing auxin-inducible degradation of GFP-tagged RAD21, WAPL or CTCF. ITR, inverted terminal repeats. **b**, Representative images of RAD21-AID-eGFP cells containing 3 × CTCF-TetO imaged before or after 90 min of auxin treatment (exposure time eGFP and tdTomato: 50 ms, deconvolved, maximum intensity projection, bicubic interpolation, $n = 3$ replicates). **c**, Left, time series of TetR-tdTomato signal over 30 min (maximum intensity projection, time interval $\text{dt} = 10$ s, color-coded for intensity changes over time). Right, magnification with overlay of TetR-tdTomato signal with reconstructed trajectories of individual TetO arrays. **d**, Left, cell motion is approximated as the average roto-translational motion of TetO signals within the same nucleus. Right, MSD averaged over trajectories within one nucleus (mean \pm s.e.m.) before (cyan, $n = 77$) and after (blue, $n = 77$) cell motion and localization error correction. Green, radial MSD of pairs of operator arrays within the same nucleus (mean \pm s.e.m., $n = 491$ pairs). **e**, Left, MSD (mean \pm s.e.m.) in mESC lines before

(blue, 310 cells, 13,537 trajectories) or after (red, 271 cells, 11,082 trajectories) Cre-mediated removal of 3 × CTCF sites. Three replicates per cell line and three lines per condition were analyzed and merged here and in all following MSD graphs. P values (two-sided Student's t -test) for all panels shown in Extended Data Fig. 2e. Right, schematic representation of Cre-mediated removal of CTCF sites. **f**, Left, same as in **e** but in mESC lines with 3 × CTCF-TetO arrays, before (blue, 323 cells, 9,829 trajectories) or after (red, 365 cells, 12,495 trajectories) CTCF degradation (6 h of auxin treatment). Right, schematic representation of auxin-induced CTCF degradation. **g**, MSD (mean \pm s.e.m.) of 3 × CTCF-TetO insertions before (blue, 310 cells, 13,537 trajectories) or after (red, 240 cells, 8,788 trajectories) RAD21 degradation (90 min of auxin). **h**, MSD (mean \pm s.e.m.) of 3 × CTCF-TetO before (blue, 336 cells, 6,687 trajectories) or after (red, 350 cells, 6,717 trajectories) WAPL degradation (24 h of auxin). **i**, Fold changes in generalized diffusion coefficients (D) and scaling exponents (α) in untreated cells compared with cells where degradation of CTCF, RAD21 and WAPL or removal of CTCF motifs (3 × CTCF) occurred.

P values in Extended Data Fig. 2e). In the presence of wild-type levels of RAD21, generalized diffusion coefficients were on average ~30% lower than in depleted cells, where RAD21 levels were low enough to prevent formation of cohesin-mediated structures (compare with Extended Data Fig. 1b). This outcome was consistent across three clonal cell lines with different TetO insertion sites and the small differences in the magnitude of the effect were likely due to location-dependent effects (Extended Data Fig. 2f). Importantly, the effect was specific for

RAD21 degradation as we did not observe any changes in MSD behavior in control cell lines expressing *OsTir1* but no AID-tag (Extended Data Fig. 2g). In addition, depletion of WAPL (24-h auxin treatment), which results in higher levels of DNA-bound cohesin³², caused a substantial decrease in chromosome mobility (Fig. 1h and Extended Data Fig. 2e). Together, these results indicate that increasing levels of DNA-bound cohesin decrease chromosome mobility, with only very minor effects (if any) mediated by the presence of even strong CTCF motifs (Fig. 1i).



Loop extrusion can explain reduced chromosome dynamics

We next used polymer simulations to determine if loop extrusion alone could explain the observed global reduction in chromosome dynamics in the presence of cohesin and minimal effects from CTCF. We simulated the dynamics of a polymer with excluded volume, with or without loop extrusion and extrusion barriers whose linear arrangement and orientation were sampled from endogenous CTCF sites (Fig. 2a and Extended Data Fig. 3a). To mimic random insertion of $3 \times$ CTCF sites, we also simulated the same polymers with additional loop extrusion barriers separated by 800 kb which were inserted at random positions in the polymer (magnified area in Fig. 2a). To emphasize their potential effects on chromosome dynamics, all barriers in the simulations were impermeable to loop extruders. Every monomer represented 8 kb of chromatin, corresponding to the genomic size of the TetO array. Simulation steps were approximated to real-time units by matching the time needed for a monomer to move by its own diameter with the time required by the TetO array to move by its estimated mean physical size (Methods). We sampled an extremely large range of extruder residence times and loading rates (4 orders of magnitude each) centered around a residence time of ~30 min and extruder densities of ~20 per Mb (in line with previous measurements^{38,39}), and using two extrusion speeds corresponding to in vivo and in vitro estimates (-0.1 kb s^{-1} and -1 kb s^{-1} , respectively)^{20,38} (Fig. 2b and Extended Data Fig. 3b).

In the absence of loop extrusion, the polymer underwent subdiffusive behavior with anomalous exponent of ~0.6 (Fig. 2c), as expected from simple polymers with excluded volume^{40–42} (see Supplementary Information) and compatible with our experimental results on random TetO insertions (Fig. 1e). Strikingly, in line with experimentally measured effects of RAD21 (Fig. 1g), introduction of loop extrusion led to lower generalized diffusion coefficients and minor effects on anomalous exponents, independently of loading rate and residence time (Fig. 2c,d), extrusion speed (Extended Data Fig. 3c) or the presence of extrusion barriers (Fig. 2e,f). Interestingly, for extruder residence times of 5.5–11 min and unloading rates corresponding to extruder linear densities of ~20 per Mb, the predicted decrease in generalized diffusion coefficients was in quantitative agreement with the experimentally observed value of ~30% (Fig. 2d,f; extruder densities as in Fig. 2b; compare with Fig. 1g). Also, consistently with WAPL depletion experiments (Fig. 1h), increasing extruder residence times systematically resulted in larger reductions in generalized diffusion coefficients (Fig. 2d,f and Extended Data Fig. 3c).

Importantly, addition of barriers in the presence of loop extrusion led to substantially smaller changes in polymer dynamics compared with the effect of loop extrusion itself even when probed directly on the barriers (Fig. 2g,h and Extended Data Figs. 3c and 4a,b), in agreement with our experimental finding that CTCF degradation had no strong effect on MSDs of TetO insertions (Fig. 1e,f and Extended Data Fig. 2c,e). Similarly, insertion of additional barriers had little impact on MSD (Fig. 2i,j), thus recapitulating the negligible effect of removal of $3 \times$ CTCF sites (Fig. 1e). Polymer simulations thus strongly support the notion that the observed decrease in chromosome mobility and lack

of effects from CTCF is a macroscopic manifestation of the physical constraints imposed by cohesin in living cells.

Cohesin and CTCF constrain the dynamics of sequences in cis

We next asked how cohesin and CTCF impact the reciprocal motion of two genomic sequences located on the same DNA molecule. To this aim, we simulated the dynamics of a polymer carrying two convergent impermeable extrusion barriers mimicking strong CTCF motifs separated by ~150 kb (Fig. 3a). This is comparable to median distances between convergent CTCF sites within TADs genome-wide in mESCs (141 kb, Methods) and also to the estimated average separation between enhancers and promoters in human cells (~160 kb)⁴³. Simulations performed with extrusion parameters recapitulating the dynamic effects of RAD21 depletion (black square in Fig. 2f) predicted that radial MSDs should be lowest in the presence of loop extrusion and barriers (Fig. 3b) due to the formation of transient loops anchored by the barriers (Fig. 3c). Similar to MSDs (Fig. 2), radial MSDs should increase upon removal of extrusion barriers and become maximal when loop extrusion is also removed (Fig. 3b).

Importantly, simulations also predicted that scaling exponents of radial MSD curves should be considerably smaller (~0.2) than those we previously observed for TetO arrays separated by several Mb or located on different chromosomes (~0.6, Fig. 1d). This is because correlations in the motion of two monomers are stronger when they are located closer along the polymer. Indeed, simulations predicted that scaling exponents fitted from radial MSD curves at short times should increase with increasing genomic distance and approach 0.6 for loci separated by several Mb (consistent with radial MSDs of randomly inserted TetO arrays) (Extended Data Fig. 5a) before saturating to stationary values at longer times. This holds true also without loop extrusion (theoretical analysis in Supplementary Information and simulations in Extended Data Fig. 5b).

To test these predictions, we turned to a live-cell imaging approach allowing us to measure the radial dynamics of two sequences located within the same TAD, in the presence and absence of cohesin and/or strong CTCF sites. We engineered mESCs carrying targeted integrations of two orthogonal operator arrays: ~140 \times TetO and 120 \times LacO separated by 150 kb (Fig. 4a), which could be visualized upon binding of TetR-tdTomato and a weak DNA-binding variant of LacI fused to eGFP (LacI^{**}-eGFP)⁴⁴. To minimize confounding effects from additional regulatory sequences such as active genes or enhancers, we targeted the arrays into a 560-kb 'neutral' TAD on chromosome 15 where we previously removed internal CTCF sites³ (Fig. 4a). The two operator arrays were directly adjacent to excisable $3 \times$ CTCF site cassettes arranged in a convergent orientation (Fig. 4a). Cell lines were verified by Nanopore Cas9-targeted sequencing (nCATS)⁴⁵ to contain a single copy of each targeting cassette (Extended Data Fig. 5c). We additionally targeted the endogenous *Rad21* locus with a C-terminal HaloTag-FKBP fusion allowing the inducible degradation of RAD21 upon treatment with dTAG-13 (ref. ⁴⁶) as confirmed by severely decreased protein levels (>95% after 2-h treatment, Extended Data Fig. 5d).

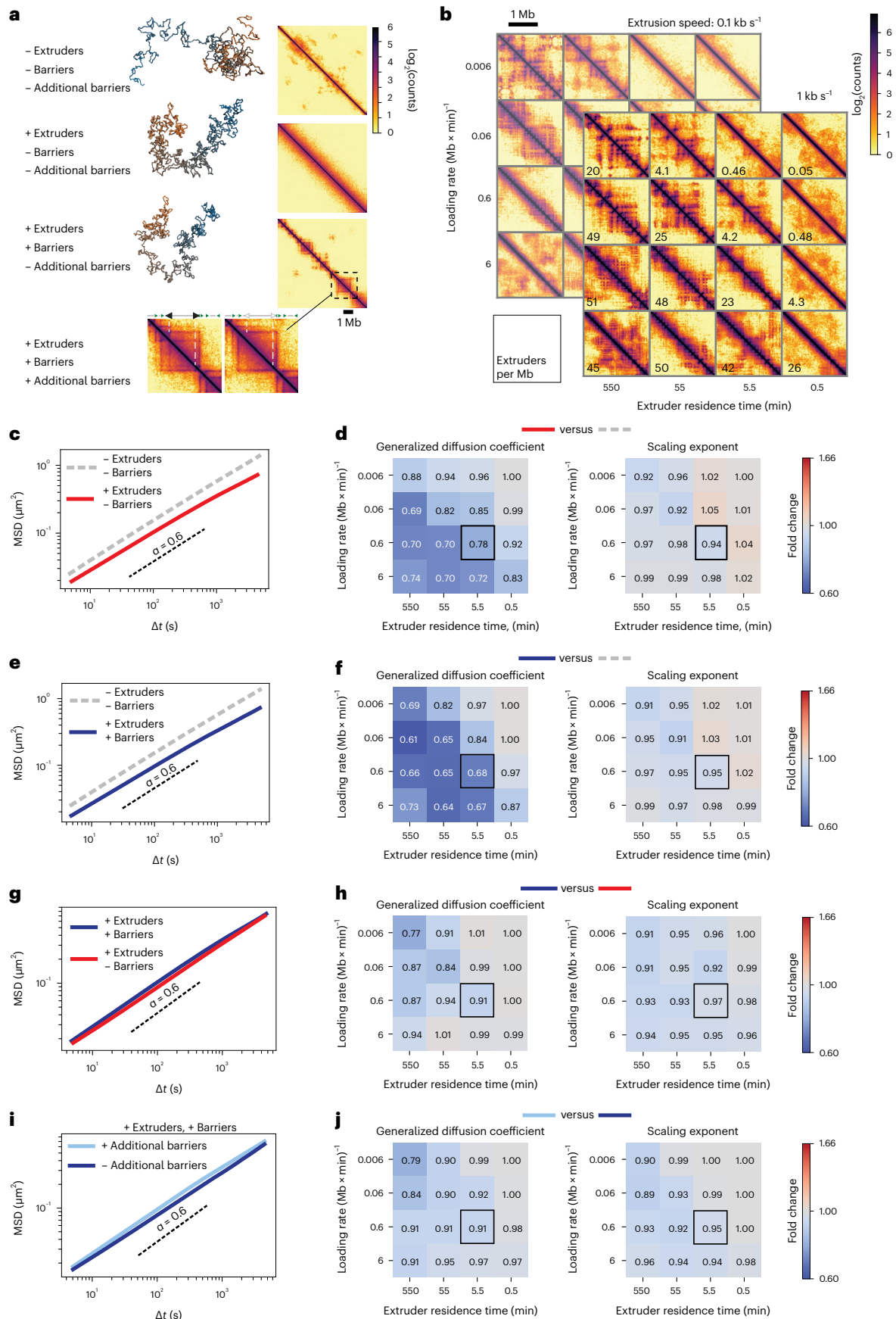
Fig. 2 | Loop extrusion generally slows down polymer motion. a.

Representative snapshots of conformations and simulated contact maps for a polymer model with excluded volume and increasingly complex models with loop extruders, extrusion barriers sampled from CTCF motifs within 9 Mb on chromosome 15 (Chr15:7–16 Mb) and additional randomly distributed extrusion barriers. For the system with additional barriers, the contact map is presented aside with magnification of the contact map of the system without additional barriers to highlight the differences. **b**, Simulated contact maps (with loop extrusion and extrusion barriers) for polymers with two extrusion speeds (1 kb s^{-1} and 0.1 kb s^{-1}) and different combinations of extruder loading rates and residence times. The resulting linear densities of extruders (number per Mb) are shown in the bottom left corner of each contact map. **c**, Effect of extruders. MSDs of polymers with (red line) or without (gray dashed line) loop extruders in the

absence of extrusion barriers (loading rate $0.6 \text{ (Mb} \times \text{min)}^{-1}$ and residence time 5.5 min, corresponds to black square in panel **d**). Black dashed curve represents $\alpha = 0.6$ as an eye guide. **d**, Effect of extruders. Ratios of generalized diffusion coefficients and anomalous exponents between the two conditions shown in panel **c**. Black square, set of parameters whose corresponding MSDs are shown in panel **c**. **e**, MSDs of polymers with (blue line) or without (gray dashed line) both extruders and barriers. Same parameters as in panel **c**. **f**, Same as panel **d** for cases illustrated in panel **e**. **g**, MSDs of polymers with loop extruders in the presence (blue) or absence (red) of extrusion barriers. Same parameters as in panels **c** and **e**. **h**, Same as panels **d** and **f** but for cases illustrated in panel **g**. **i**, MSDs of polymers either with (light blue) or without (red) additional randomly inserted extrusion barriers. Same parameters as in panels **c**, **e**, **g**. **j**, Same as panels **d**, **f** and **h** but for cases illustrated in panel **i**.

Capture-C with tiled oligonucleotides revealed that integration of operator arrays themselves did not lead to detectable changes in chromosome structure (Extended Data Fig. 5e). Convergent 3 × CTCF

sites, however, led to the formation of a new CTCF-mediated interaction within the TAD (2.8× increase in contact probability after correcting the confounding contribution of the wild-type allele) (Fig. 4b), which



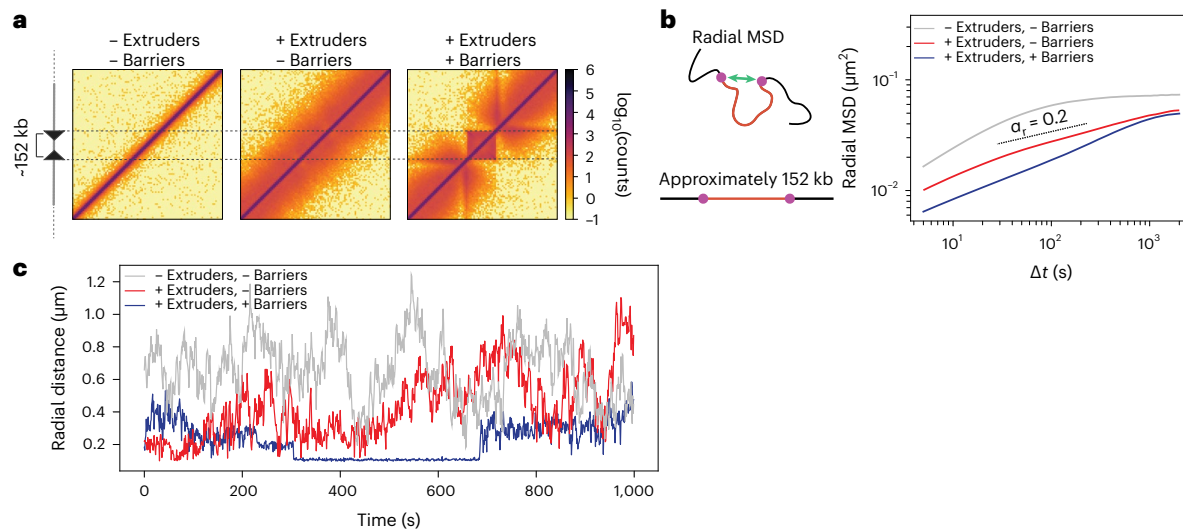


Fig. 3 | Convergent CTCF sites further constrain polymer dynamics.

a, Simulated contact maps of a region spanning the equivalent of 800 kb for a polymer chain without loop extrusion, with loop extruders and with convergent extrusion barriers separated by the equivalent of 152 kb. **b**, Radial MSD of the two monomers separated by the equivalent of 152 kb in the three conditions from

panel **a**. Dashed line is an exponent of 0.2 as a guide to the eye (α_r indicates the slope of radial MSDs). Loop extrusion parameters as in Fig. 2c. **c**, Representative examples of distances between the two monomers in simulations with or without loop extrusion and extrusion barriers. The flat stretch in the trajectory with extrusion and barriers corresponds to a loop anchored by the two barriers.

was lost upon RAD21 depletion along with all other CTCF-mediated interactions across the locus (Extended Data Fig. 5f).

We imaged cells for 3 h every 30 s in three dimensions (Fig. 4c and Supplementary Video 2), either in the presence or absence of RAD21, and measured distances between the two arrays over time (Fig. 4d, $n = 3-7$ biological replicates for each condition, on average 220 cells per condition, Supplementary Table 1 and Methods). Doublet signals corresponding to replicated alleles occurred in a very minor fraction (3%, Methods) of trajectories, compatible with the late-replication profile of the ‘neutral’ TAD and the cell-cycle distribution (Extended Data Figs. 5g and 6a). In these cases, only trajectories that were initially closest across channels were considered. After correction of chromatic aberrations (Methods and Extended Data Fig. 6b), we estimated our experimental uncertainty on radial distances to be ~ 130 nm by measuring pairwise distances in control cells where multiple TetO insertions were simultaneously bound by both TetR-tdTomato and TetR-eGFP (Extended Data Fig. 6c–e).

In agreement with model predictions for locations separated by 150 kb, radial MSDs of the two arrays showed scaling exponents close to 0.2, much smaller than those observed with randomly inserted TetO arrays (Extended Data Fig. 7a,b). Also in line with model predictions

(Extended Data Fig. 5a), presence of RAD21 and $3 \times$ CTCF sites led to the most constrained radial mobility, whereas RAD21 degradation and deletion of CTCF sites resulted in the least constrained motion (Extended Data Fig. 7c). These measurements thus verified the model prediction that genomic sequences located at short distances (150 kb) experience stronger physical constraints than sequences located at larger genomic distances²⁶ (Fig. 1 and Extended Data Fig. 7b), and that loop extrusion provides constraints that are further reinforced by convergent CTCF sites.

Consistently with their more constrained radial MSD behavior, we finally observed that distances between TetO and LacO signals were smallest in the presence of convergent CTCF sites and cohesin. In these conditions, distances between TetO and LacO arrays tended to remain close to the ~ 130 -nm experimental uncertainty with only occasional fluctuations toward larger values in the course of the 3 h of imaging (Fig. 4d,e). Removal of $3 \times$ CTCF sites led to increased radial distances and variability within single trajectories, which were further increased upon degradation of RAD21, irrespective of the presence or absence of CTCF sites (Fig. 4d,e and Supplementary Video 3). Thus, constraints imposed by extruding cohesin and convergent CTCF sites reduce not only average physical distances between sequences but also

Fig. 4 | Cohesin and CTCF reduce variability in chromosome folding dynamics.

a, Top, insertion of TetO and LacO arrays separated by 150 kb within a ‘neutral’ TAD on chromosome 15 in mESCs. Flanking $3 \times$ CTCF sites can be excised by Cre and FLP recombinases. Arrays are visualized by binding of LacI*-eGFP and TetR-tdTomato, respectively. Bottom, tiled Capture-C map (6.4-kb resolution) and genomic datasets in mESCs in a region in 2.6 Mb surrounding the engineered TAD. Capture-C was performed in cells where arrays were flanked by $3 \times$ CTCF sites. Dashed lines, positions of LacO and TetO insertions. **b**, Capture-C maps in mESC lines with (left) or without (middle) $3 \times$ CTCF sites flanking TetO and LacO arrays, and differential map (right, $+3 \times$ CTCF versus $-3 \times$ CTCF, Methods) highlighting interactions formed between convergent $3 \times$ CTCF sites (arrows). **c**, Top, representative fluorescence microscopy images of mESCs with $3 \times$ CTCF-LacO and TetO- $3 \times$ CTCF insertions. Bottom, magnified view with time series overlay of LacI*-eGFP and TetR-tdTomato signals (exposure time 50 ms, deconvolved, maximum intensity projection, bicubic interpolation). **d**, Representative trajectories of TetO-LacO radial distances with or without

convergent $3 \times$ CTCF sites, either before or after degradation of RAD21 (2 h of dTag-13) ($dt = 30$ s). **e**, Distribution of TetO-LacO radial distances in the four experimental conditions ($+3 \times$ CTCF sites/ $+RAD21$: $n = 152$ cells, 4 pooled replicates; $-3 \times$ CTCF sites/ $+RAD21$: $n = 214$ cells, 4 pooled replicates; $+3 \times$ CTCF sites/ $-RAD21$: $n = 248$ cells, 7 pooled replicates; $-3 \times$ CTCF sites/ $-RAD21$: $n = 277$ cells, 6 pooled replicates). **f**, Distributions of variance over mean within single trajectories across the four experimental conditions (no. of cells as in panel **e**). Boxes, lower and upper quartiles (Q1 and Q3, respectively). Whiskers denote $1.5 \times$ interquartile region (IQR) below Q1 and above Q3. P values are calculated using two-sided Kolmogorov-Smirnov test. NS, not significant; $***P < 0.01$; $****P < 0.0001$. Exact P values can be found in Supplementary Table 2. Outliers are not shown. **g**, Distribution of jump step size (changes in TetO-LacO radial distance) across increasing time intervals for the four experimental conditions (no. of cells as in panel **e**). Boxes, lower and upper quartiles (Q1 and Q3, respectively). Whiskers, $1.5 \times$ IQR below Q1 and above Q3. Outliers are not shown.

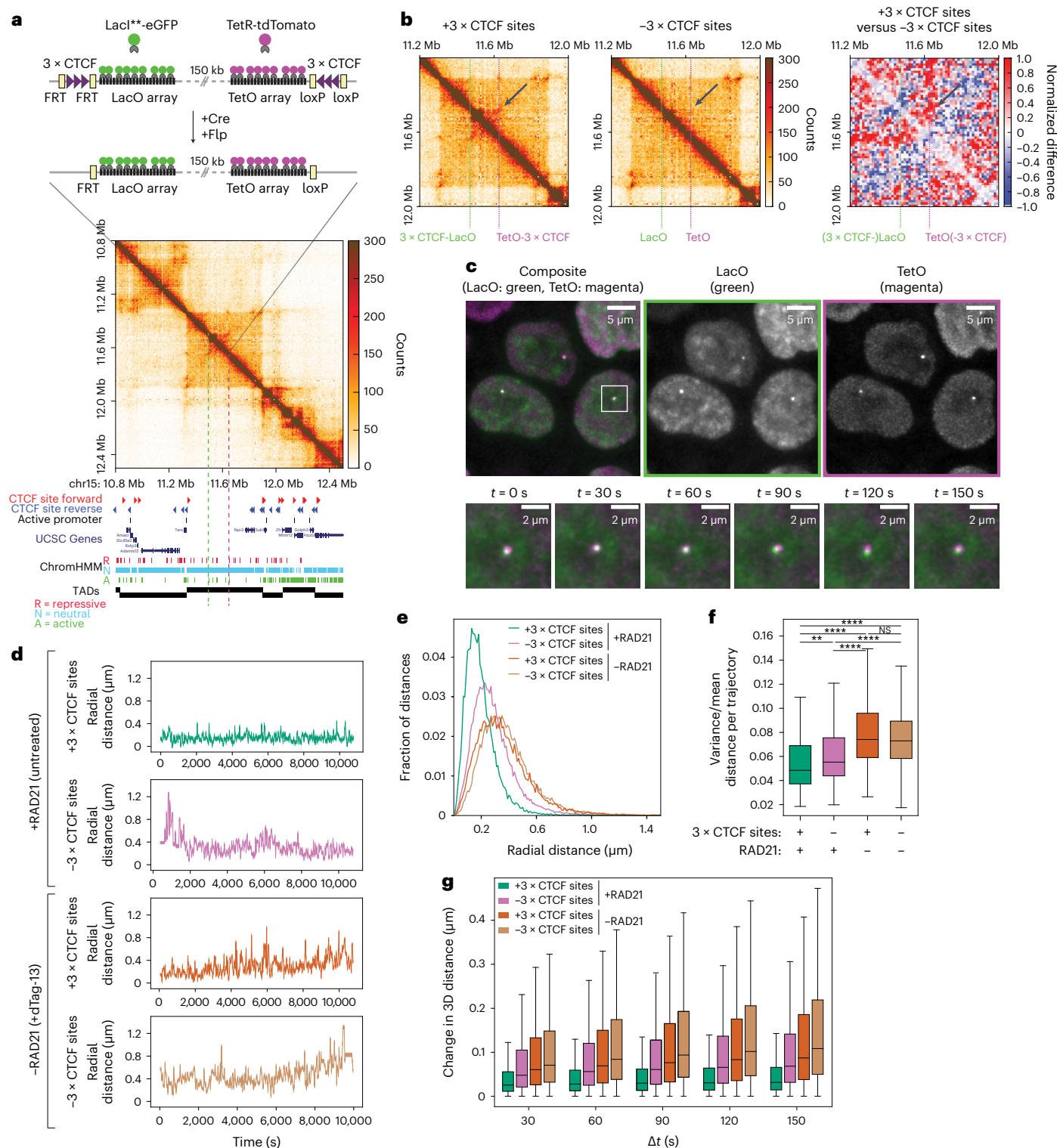
their variability in time (Fig. 4f), also supported by analysis of distance changes (jumps) as a function of time (Fig. 4g).

Finally, to test whether the effects of cohesin on chromosome motion would be different in the presence of active transcription at nearby locations, we measured looping dynamics this time in a parental mESC line before the removal of resistance cassettes. In this line, both the TetO and LacO arrays were immediately flanked by mouse *Pgk1* promoters⁴⁷ driving the transcription of resistance genes (Extended Data Fig. 6f,g). In line with previous studies^{48,49}, we found that active transcription led to slightly decreased radial MSD. Cohesin depletion

resulted in similar amounts of increased radial mobility irrespective of the presence or absence of active promoters (Extended Data Fig. 7c).

Chromosomal contacts are transient

We next set off to quantify changes in distances over time and determine whether despite the experimental uncertainty on 3D distances (Extended Data Fig. 6e) we could observe transitions between two states: a ‘proximal’ state with small radial distances (presumably including cohesin-mediated loops between convergent CTCF sites), and a generic ‘distal’ state with larger spatial distances corresponding to



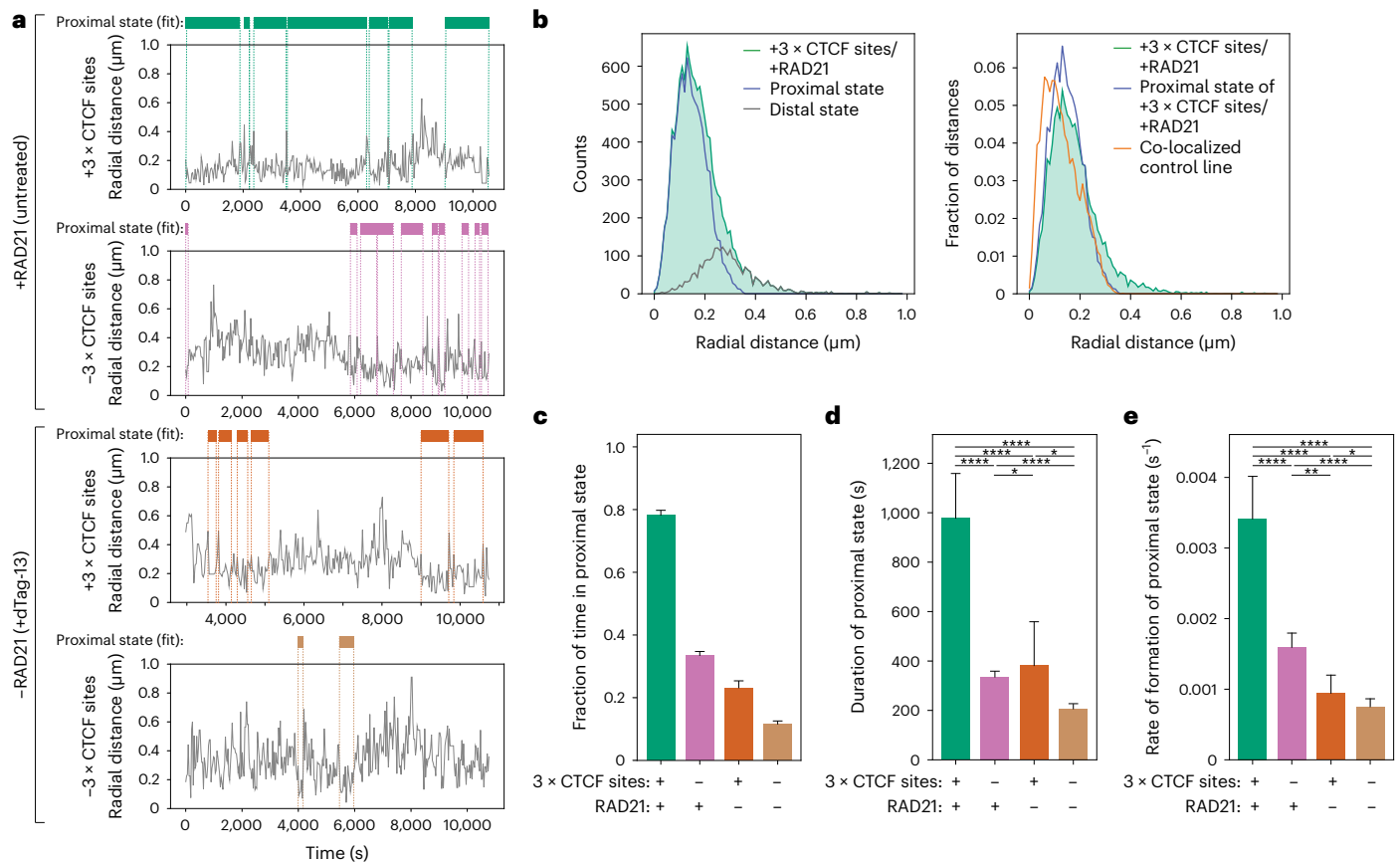


Fig. 5 | Cohesin and CTCF control contact dynamics inside a TAD.

a, Representative trajectories of radial distance (gray) and occurrences of the proximal state called by HMM (colored bars). The HMM was fitted on data with convergent 3 × CTCF sites and RAD21 (top left) to find the proximal state which was then imposed on the other three samples. **b**, Left, radial distance distribution in cells with convergent 3 × CTCF sites and RAD21 overlaid with those of proximal and distal states called by HMM on the same sample. Right, same as in the left panel but normalized and with the additional display of the distance distribution from a control cell line where TetO and LacO signals perfectly co-localize. **c**, Fraction of time spent in the proximal state called by HMM in the four experimental conditions (no. of replicates as indicated in

Fig. 4e). Shown are averages across experimental conditions; error bars represent bootstrapped ($n = 10,000$) standard deviations. **d**, Average durations of proximal states (mean \pm 95% confidence interval (CI), $n = 680$ (−3 × CTCF/+RAD21); $n = 287$ (+3 × CTCF/+RAD21); $n = 268$ (−3 × CTCF/−RAD21); $n = 114$ (+3 × CTCF/−RAD21)). P values (two-sided Kolmogorov–Smirnov): * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. Exact P values can be found in Supplementary Table 2. **e**, Average rates of contact formation—time elapsed between the end of a proximal state and the beginning of the next (mean \pm 95% CI, $n = 726$ (−3 × CTCF/+RAD21); $n = 323$ (+3 × CTCF/+RAD21); $n = 268$ (−3 × CTCF/−RAD21); $n = 138$ (+3 × CTCF/−RAD21)). P values as in panel **d**.

other configurations of the chromatin fiber. This was motivated by the expectation that any polymer with site-specific attractive interactions, such as those mediated by cohesin at convergent CTCF sites, should in principle result in two-state thermodynamic behavior. We thus fitted a two-state hidden Markov model (HMM) on the ensemble of trajectories obtained in cells where both convergent 3 × CTCF sites and RAD21 were present (Fig. 5a,b). Interestingly, distances in the proximal state inferred by HMM largely overlapped with those detected on perfectly colocalizing signals in control experiments where TetR-eGFP and TetR-tdTomato were bound to the same set of randomly inserted TetO arrays (149 versus 130 nm on average, respectively) (Fig. 5b and Extended Data Fig. 6e). The proximal state thus corresponds to configurations of the chromatin fiber where the two arrays were in very close physical proximity, also including (but not restricted to) cohesin-mediated loops between CTCF sites. For simplicity, we refer to the proximal state interchangeably as ‘contact’, without implying a direct molecular interaction between the two DNA fibers. Radial distances in the distal state (288 nm on average) instead were similar to those measured in cells where both CTCF sites had been removed (291 nm) (Extended Data Fig. 8a,b). Thus, the distal state largely overlapped with chromosome conformations where specific cohesin-mediated CTCF loops were lost.

We next fitted the HMM to all experimental conditions while keeping the same proximal state as in cells with 3 × CTCF sites and RAD21 (Fig. 5a). This showed that in the presence of RAD21, the LacO and TetO arrays spent ~78% of the time in contact (that is, in the proximal state) when 3 × CTCF sites were present. This was 2.3× higher than the 33% of time they spent in contact when the 3 × CTCF sites were removed (Fig. 5c), in agreement with the corresponding 2.8-fold difference in contact probability inferred from Capture-C (Fig. 4b). The fraction of time spent in contact decreased markedly upon depletion of RAD21 to ~23% in the presence of 3 × CTCF sites and 11% in the absence (Fig. 5c). Both the average duration of contacts and their rate of formation were maximal in the presence of RAD21 and 3 × CTCF sites, where they lasted around 16 min and reformed every 5 min on average (Fig. 5d,e). Contacts became substantially shorter (6 min) and rarer (one every 10 min) when 3 × CTCF sites were removed, and even more so upon RAD21 depletion (lasting 2 min and occurring every 22 min on average). Interestingly, these results were not affected by the presence of actively transcribed promoters in the immediately flanking regions (Extended Data Fig. 8b–f), in line with the lack of changes in contact probability measured in Capture-C (Extended Data Fig. 6g). Thus, both cohesin and CTCF impact both the duration and the probability of

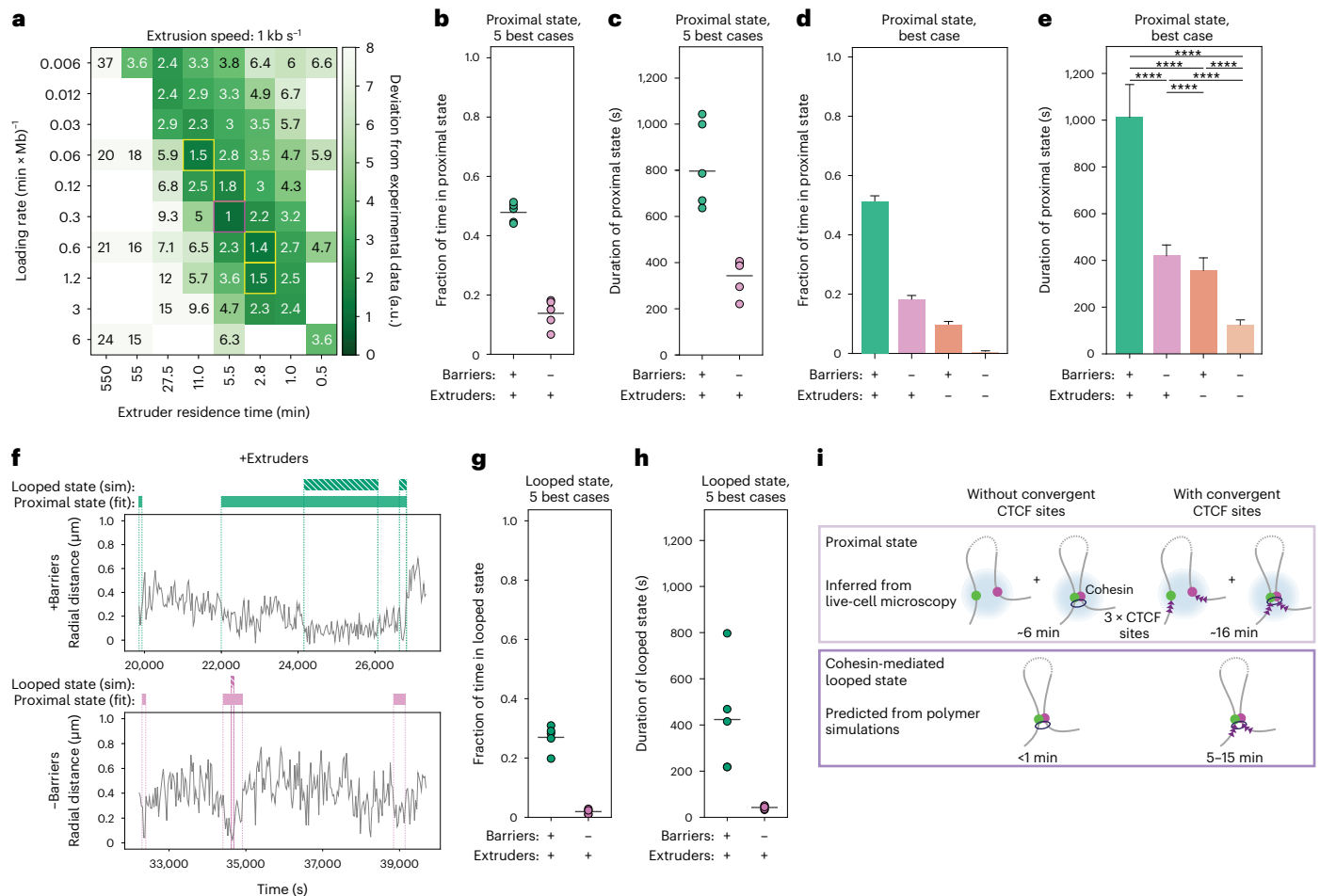


Fig. 6 | Estimation of frequency and duration of cohesin-mediated CTCF loops. **a**, Levels of agreement between simulations and experimental data as a function of loop extrusion parameters (here shown with extrusion speed 1 kb s^{-1}). The score represents the deviations of the distance, duration and fraction of time spent in the proximal state with those experimentally observed in the presence of RAD21 with or without $3 \times$ CTCF sites (Methods). Magenta square, parameter set maximizing the agreement with experimental values. Yellow squares, four additional second-best parameter sets. **b**, Fraction of time spent in the proximal state called by HMM on simulations with the five best-matching parameters (magenta and yellow squares in panel **a** for +Extruder case, Methods). **c**, Average duration (mean \pm 95% CI) of proximal state called by HMM on simulations with the five best-matching parameters. **d**, Fraction of time spent in the proximal state called by HMM on simulations (over $n = 15,880$ time points) for the best-matching parameter set in the presence of extruders (+) or low levels (-) of extruders, either with or without extrusion barriers. Shown are averages across experimental

conditions; error bars represent bootstrapped ($n = 10,000$) standard deviations. **e**, Average duration of the proximal state (mean \pm 95% CI, over $n = 15,880$ time points) either in the presence of extruders (+) or low levels of extruders (-), either with or without extrusion barriers. Two-sided Kolmogorov-Smirnov P values can be found in Supplementary Table 2. **f**, Representative trajectories of radial distances (gray), contact states called by HMM (full bar) and looped states in the underlying polymer conformations (striped bars) from +Extruders/+Barriers (top) and +Extruders/-Barriers simulations (bottom) with best-matching parameters (magenta square in panel **a**). **g**, Fraction of time spent in the looped state based on simulations with the five best-matching parameters. **h**, Average duration of the looped state based on simulations with the five best-matching parameters (mean \pm 95% CI). **i**, Scheme summarizing the durations of proximal and looped states in the presence and absence of $3 \times$ CTCF sites. a.u., arbitrary unit; sim, simulation.

formation of chromosomal contact events between loci separated by 150 kb within an 'empty' TAD.

To understand if these results could be rationalized in terms of loop extrusion, we compared them with polymer simulations with convergent impermeable loop extrusion barriers separated by ~ 150 kb. Simulations were performed using loop extrusion parameters spanning a finer-grained 25-fold range around experimentally realistic values that reproduced the dynamic effect of RAD21 degradation (compare with Fig. 2d, black square) and with both in vitro and in vivo estimates of extrusion speeds^{20,38}. In a large region of the parameter space, distances between convergent barriers were bimodally distributed, supporting the expectation that the polymer can be approximated as a two-state system (Extended Data Fig. 9a). To allow direct comparison with experimental distance-based HMM states, we applied random

errors matching experimental uncertainty levels to radial distances generated by the models (Extended Data Fig. 9b). We called proximal and distal states using the same HMM strategy as with experimental data. Importantly, for a large number of parameter combinations, distances in the proximal state largely overlapped with the corresponding distribution observed experimentally in the presence of convergent CTCF sites and cohesin (Extended Data Fig. 9c and Supplementary Fig. 1a).

We then compared the distance, duration and fraction of time spent in the proximal state with those experimentally observed in the presence of RAD21 with or without $3 \times$ CTCF sites. We found that their similarity was maximal for extruder densities ranging from 8 to 32 per Mb (Supplementary Fig. 1e) and residence times of 2.8–11 min, with extrusion speeds of both 0.1 and 1 kb s^{-1} , all of which

were in the range of previous estimations of experimental values^{20,21,38,50} (Fig. 6a and Extended Data Fig. 9d). Considering the five best-matching scenarios (red- and yellow-marked values in Fig. 6a), the two locations spent 45–55% of the time in the proximal state with an average contact duration of around 10–17 min, which reduced to 18% and 8 min in the absence of extrusion barriers (Fig. 6b,c). Similar to the effects observed experimentally upon depletion of RAD21, decreasing extruder densities (for example, by decreasing loading rates) led to decreased fractions of time and shorter durations of the proximal state (Fig. 6d,e, shown for the best case, general trends in Supplementary Fig. 1d–d). Thus, the duration and the fraction of time spent in the proximal state, and most importantly how these quantities change upon removing cohesin and/or CTCF sites, can be understood in terms of a simple loop extrusion model.

The HMM-based proximal state likely provides an overestimation of the duration of underlying CTCF-CTCF loops mediated by stalled cohesins, since it also contains a fraction of CTCF-independent proximity events that cannot be distinguished from loops. To estimate the duration and times the two loci spent in a cohesin-mediated CTCF-CTCF looped conformation, we quantified occurrences in the simulated polymer where the two monomers formed the base of an extruded loop (Fig. 6f and Methods). As expected, these events were rarer and shorter than contacts detected by HMM on polymer simulations (Fig. 6b,c), with two monomers spending ~20–31% of time at a loop base for 5–15 min on average in the presence of extrusion barriers (Fig. 6g,h). Finally, transient cohesin-dependent loops that are not stabilized by CTCF sites should occur much more rarely (1–3% of the time) and lasted less than a minute on average (Fig. 6g,h). Comparison of polymer simulations with HMM states thus suggests that the dynamics of chromosome contacts detected at a range of 150 nm are generated by faster and rarer cohesin-mediated CTCF loops (Fig. 6i).

Discussion

Our study provides quantitative measurements of chromosome folding dynamics in living cells and reveals how they are controlled by cohesin and CTCF. Two experimental strategies allow us to minimize biological variation from specific regulatory and structural genomic contexts and enable direct comparison with polymer models. By studying large numbers of random genomic locations, we average over local differences in chromosome mobility and reveal the global dynamic effects of cohesin. By visualizing and manipulating two locations within a ‘neutral’ genomic environment, we unravel how cohesin and CTCF impact chromosome looping within a single TAD. We show that although higher extrusion speeds could in principle result in increased chromosome motion (Extended Data Fig. 9e,f), physiological extrusion rates rather generate transient constraints that decrease chromosome dynamics, in line with previous measurements of histone mobility⁴⁹. Similar to previous reports⁵¹, we observe that constraints introduced by cohesin reduce spatial distances between genomic sequences in *cis* and increase the chances that they interact. We now, however, reveal that this entails an increase in both the rate of formation and the duration of contacts. Convergent oriented high-affinity CTCF motifs lead to higher contact frequencies and substantially longer contact durations, somewhat similar to the effect of insulator elements in *Drosophila*²³. Comparison with polymer simulations reveals that in mESCs this can be understood in terms of stalling of loop-extruding cohesins. This observation also suggests that asymmetries in contact patterns established by CTCF motifs genome-wide might also lead to temporal asymmetries in physical interactions, notably between regulatory sequences. We additionally observe that constraints introduced by cohesin and CTCF sites lead to reduced temporal variability in physical distances, arguing that loop extrusion increases the reproducibility of chromosome folding at selected genomic sites.

Our study also provides estimates of the frequency and duration of chromosomal contacts at genomic-length scales that represent

enhancer–promoter communication genome-wide. In our study, contacts are defined by physical distances (~150 nm) that might be comparable to those where signals arise in 3C methods¹¹. For sequences separated by 150 kb, such contacts assemble and disassemble over minutes. This provides many opportunities in a single cell-cycle for regulatory sequences in a TAD to contact each other, and suggests that long-range regulation by distal enhancers might rely on transient interactions. We note that despite accurate correction of chromatic aberrations, shorter-range and thus potentially faster proximity events remain inaccessible in our experimental set-up⁵². Estimates based on comparison with polymer simulations further suggest that cohesin-mediated interactions between convergent CTCF sites might last around 5–15 min on average and at least for sequences located 150 kb apart occur around 27% of the time. This is in good agreement with recent estimates of the duration of a 500-kb loop in mESCs (10–30 min on average)²⁶, which, however, occurs more rarely (3.5–6% of the time). This is in line with the predictions from polymer simulations that increasing the genomic distance between convergent CTCF sites should substantially decrease the frequency of CTCF-mediated interactions, but not their duration (Extended Data Fig. 10). Taken together, our data establish firm quantitative bases for understanding the dynamics of chromosome folding within TADs and provide temporal constraints for mechanistic models of chromosome structure and its impact on fundamental biological processes such as long-range transcriptional regulation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-022-01232-7>.

References

1. Marchal, C., Sima, J. & Gilbert, D. M. Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **20**, 721–737 (2019).
2. Arnould, C. et al. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660–665 (2021).
3. Zuin, J. et al. Nonlinear control of transcription through enhancer–promoter interactions. *Nature* **604**, 571–577 (2022).
4. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
7. Nora, E. P. et al. Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.* **11**, 5612 (2020).
8. Li, Y. et al. The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 472–476 (2020).
9. Davidson, I. F. et al. CTCF is a DNA-tension-dependent barrier to cohesin-mediated DNA loop extrusion. Preprint at bioRxiv <https://doi.org/10.1101/2022.09.08.507093> (2022).
10. Zhang, H. et al. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. Preprint at bioRxiv <https://doi.org/10.1101/2022.09.15.508177> (2022).
11. 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).
12. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

13. Amano, T. et al. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
14. Finn, E. H. et al. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* **176**, 1502–1515.e10 (2019).
15. Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
16. Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
17. Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
18. Huang, H. et al. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).
19. Liu, M. et al. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).
20. Davidson, I. F. et al. DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
21. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* <https://doi.org/10.7554/eLife.25776> (2017).
22. Barbieri, M. et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl Acad. Sci. USA* **109**, 16173–16178 (2012).
23. Chen, H. et al. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
24. Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
25. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
26. Gabriele, M. et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
27. Lawrimore, J. et al. ChromoShake: a chromosome dynamics simulator reveals that chromatin loops stiffen centromeric chromatin. *Mol. Biol. Cell* **27**, 153–166 (2016).
28. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).
29. Cadiñanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
30. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422.e4 (2017).
31. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
32. Liu, N. Q. et al. WAPL maintains a cohesin loading cycle to preserve cell-type specific distal gene regulation. *Nat. Genet.* **53**, 100–109 (2021).
33. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
34. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
35. Eichenberger, B. T., Zhan, Y., Rempfler, M., Giorgetti, L. & Chao, J. A. deepBlink: threshold-independent detection and localization of diffraction-limited spots. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkab546> (2021).
36. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050–1055 (2018).
37. Khanna, N., Zhang, Y., Lucas, J. S., Dudko, O. K. & Murre, C. Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nat. Commun.* **10**, 2771 (2019).
38. Cattoglio, C. et al. Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife* **8**, e40164 (2019).
39. Holzmann, J. et al. Absolute quantification of cohesin, CTCF and their regulators in human cells. *eLife* **8**, e46269 (2019).
40. Tamm, M. V. & Polovnikov, K. Dynamics of polymers: classic results and recent developments. Order, disorder and criticality: advanced problems of phase transition theory, 113–172 (2018). https://www.worldscientific.com/doi/abs/10.1142/9789813232105_0003
41. Dünweg, B. & Kremer, K. Molecular dynamics simulation of a polymer chain in solution. *J. Chem. Phys.* **99**, 6983–6997 (1993).
42. Salari, H., Stefano, M. D. & Jost, D. Spatial organization of chromosomes leads to heterogeneous chromatin motion and drives the liquid- or gel-like dynamical behavior of chromatin. *Genome Res.* **32**, 28–43 (2022).
43. Jung, I. et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* **51**, 1442–1449 (2019).
44. Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C. & Taddei, A. Tight protein–DNA interactions favor gene silencing. *Genes Dev.* **25**, 1365–1370 (2011).
45. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
46. Nabet, B. et al. The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
47. Adra, C. N., Boer, P. H. & McBurney, M. W. Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter. *Gene* **60**, 65–74 (1987).
48. Germier, T. et al. Real-time imaging of a single gene reveals transcription-initiated local confinement. *Biophys. J.* **113**, 1383–1394 (2017).
49. Nozaki, T. et al. Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell* **67**, 282–293.e7 (2017).
50. Wutz, G. et al. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin^{STAG1} from WAPL. *eLife* **9**, e52091 (2020).
51. Szabo, Q. et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).
52. Brandão, H. B., Gabriele, M. & Hansen, A. S. Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* **70**, 18–26 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Culture of mESC lines

All cell lines are based on the E14Tg2a parental mESC line (karyotype 19, XY, 129/Ola isogenic background; E14 for brevity). E14 CTCF-AID-eGFP (clone ENS2.9.1) was published by Nora et al.³¹. E14 WAPL-AID-eGFP and E14 RAD21-AID-eGFP were published by Liu et al.³². The latter were kindly provided by Elzo de Wit (Netherlands Cancer Institute). All cell lines for the dual-array imaging approach are based on the double-CTCF knockout cell line described by Zuin et al.³. Cells were cultured on gelatin-coated culture plates in Glasgow Minimum Essential Medium (Sigma-Aldrich, G5154) supplemented with 15% fetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μM β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U ml⁻¹ leukemia inhibitory factor (Miltenyi Biotec, premium grade) in 8% CO₂ at 37 °C. Cells were tested for mycoplasma contamination regularly and no contamination was detected. For Hi-C, Capture-Hi-C, 4C-seq, western blot and imaging experiments, cells were cultured in standard E14 medium supplemented with 2i (1 μM MEK inhibitor PDO35901 (Axon, 1408) and 3 μM GSK3 inhibitor CHIR 99021 (Axon, 1386)). For live-cell imaging experiments, cells were cultured in Fluorobrite DMEM (Gibco, A1896701) supplemented with 15% fetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 μM β -mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U ml⁻¹ leukemia inhibitory factor (Miltenyi Biotec, premium grade) and with 2i inhibitors (1 μM MEK inhibitor PDO35901 (Axon, 1408) and 3 μM GSK3 inhibitor CHIR 99021 (Axon, 1386)).

Generation of mESC lines carrying random integrations of TetO arrays

To generate clonal cell lines carrying random integrations of the TetO array in the degro cell lines (E14 Rad-AID-eGFP, E14 CTCF-AID-eGFP and E14 WAPL-AID-eGFP), 0.5×10^6 cells were transfected with 2 μg of PB-3 \times CTCF-TetO vector, 200 ng of PB-TetR-tdTomato and 200 ng of pBroad3_hypBase_IRES_tagRFpT (ref.⁵³) with Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's recommendations. Cells were cultured in standard E14 medium for 5 d and subsequently sorted by FACS for fluorescence emission at 581 nm (tdTomato) on 96-well plates to isolate clonal lines. Sorted cells were kept for 2 d in standard E14 medium supplemented with 100 μg μl^{-1} puromycin (InvivoGen, ant-pm-1) and 10 μM ROCK inhibitor (STEMCELL Technologies, Y-27632). At 10 d after sorting, the plates were duplicated by detaching with accutase (Sigma-Aldrich, A6964) and re-seeding in full E14 culture medium. One-third of the cells were replated onto Corning High-Content Imaging Glass Bottom Microplates (96-well, Corning, 4580). At 2 d after re-seeding, clonal lines were screened by microscopy for >10 insertions of TetO per cell and a good signal-to-noise ratio (SNR). Selected clones were expanded and genotyped by PCR for the absence of random integration of the piggyBac itself. Primers used for genotyping are listed in Supplementary Table 3.

Generation of dual-array (TetO-LacO) mESC line

Integration of the TetO array into the genomic locus on chr15:11,647,372: the vector containing the guide RNA (gRNA) sequence was available from a previous study (PX459-chr15_gRNA/Cas9). The gRNA sequence can be found in Supplementary Table 3. E14 mESCs already containing a double-knockout for CTCF sites (clone D6 in ref.³) were transfected with the targeting vector pMK-3 \times CTCF-TetO-Rox-PuroR-Rox and the gRNA vector PX459-chr15_gRNA/Cas9 using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024 KT). Then, 2×10^6 cells were nucleofected with 1 μg of TetO targeting vector and 1 μg of PX459-chr15_gRNA/Cas9

as described above and treated with 1 μg ml⁻¹ puromycin (InvivoGen, ant-pr-1) 48 h after transfection for 3 d to select cells for insertion of the TetO cassette. Cells were then cultured in standard E14 medium for an additional 7 d and subsequently sorted by FACS on 96-well plates as described above to isolate clonal lines. At 10 d after sorting, the plates were duplicated by detaching with accutase (Sigma-Aldrich, A6964) and re-seeding in full E14 culture medium. Genomic DNA was extracted on-plate by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl and 1 mg ml⁻¹ proteinase K (Macherey-Nagel, 740506)) and 0.05 mg ml⁻¹ RNase A (Thermo Fisher Scientific, EN0531) and subsequent isopropanol precipitation. Individual cell lines were analyzed by genotyping PCR to determine heterozygous insertion of the TetO cassette. Cell lines showing the corrected genotype were selected and expanded. Primers used for genotyping are listed in Supplementary Table 3. Targeted nanopore sequencing with Cas9-guided adapter ligation⁴⁵ was performed on expanded clones to confirm single-copy insertion of the TetO cassette. Clone 2G5 was used for further engineering. Integration of the LacO array into the genomic locus on chr15:11,496,908: the gRNA sequence for the CRISPR-Cas9 knock-in of the LacO cassette was designed using the online tool https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE and purchased from Microsynth AG. The gRNA sequence can be found in Supplementary Table 3. The gRNA sequence was cloned into the PX330 plasmid (Addgene, no. 58778) using the BsaI restriction site. The clonal line carrying the TetO cassette (clone 2G5) was transfected with the targeting vector pUC19-ITR-NeoR-ITR-3 \times CTCF-LacO and the gRNA vector pX330-chr15_LacO_gRNA/Cas9 using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024 KT). A total of 2×10^6 cells were harvested using accutase (Sigma Aldrich, A6964) and resuspended in 100 μl transfection solution (82 μl primary solution, 18 μl supplement, 15 μg targeting vector and 5 μg of gRNA vector) and transferred to a single Nucleocuvette (Lonza). Nucleofection was performed using the protocol CG110. Transfected cells were directly seeded in pre-warmed E14 standard medium. At 48 h after transfection, 250 μg ml⁻¹ G418 (InvivoGen, ant-gn-1) was added to the medium for 3 d to select cells for insertion of the LacO cassette. Cells were sorted and genotyped as described for the TetO integration. Primers used for genotyping are listed in Supplementary Table 3. Cell lines showing the corrected genotype were selected and expanded. Expanded clones were transiently transfected with 200 ng of PB-TetR-tdTomato and 200 ng of PB-LacI-eGFP using Lipofectamine3000 according to the manufacturer's instructions (Thermo Fisher Scientific, L3000008) and 2 d after transfection validated for heterozygous insertion of the LacO cassette on the same allele as the TetO by microscopy. Targeted nanopore sequencing with Cas9-guided adapter ligation⁴⁵ was performed on correct clones to confirm single-copy insertion of the LacO cassette. Clone 1F11 was used for further engineering. To visualize the operator arrays in live-cell imaging and remove the puromycin resistance gene used for selection during integration, 0.5×10^6 E14 TetO-LacO cells (clone 1F11) were transfected with 200 ng of PB-TetR-tdTomato, 200 ng of PB-LacI-eGFP and 200 ng of pBroad3_hypBase_IRES_tagRFpT (ref.⁵⁴) with Lipofectamine3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. At 7 d after transfection the cells were sorted (as described previously) for fluorescence emission at 507 nm (eGFP) and 581 nm (tdTomato). Sorted cells were cultured and genotyped as described for the random TetO integration. Primers used for genotyping are listed in Supplementary Table 3. Cell lines showing the corrected genotyping pattern were selected and expanded and a good and comparable SNR was selected for by microscopy. Clones 1B4 (+PuroR) and 2C10 (-PuroR) were used for further engineering.

Live-cell imaging

First, 35-mm glass-bottom dishes (Mattek, P35G-1.5-14-C) were coated with 1–2 μg ml⁻¹ Laminin (Sigma-Aldrich, L2020) in PBS at

37 °C overnight. Cells (1×10^6) were seeded in Fluorobrite medium (as described above) 24 h before imaging. For targeted degradation of RAD21, WAPL or CTCF in the degron cell lines, the medium was exchanged to medium containing 500 μ M auxin (Sigma-Aldrich, I5148-2G) at the respective time required for complete degradation of the protein target before imaging (RAD21: 90 min, WAPL: 24 h, CTCF: 6 h). For targeted depletion of RAD21 using the FKBP degron system (dual-array cell lines), cells were cultured in Fluorobrite medium containing 500 nM dTAG-13 (Sigma-Aldrich, SML2601-IMG) 2 h before imaging. For fixed cell measurements to estimate the localization error, 1×10^6 cells were seeded onto Mattek dishes and incubated for 24 h at 37 °C, 8% CO₂. The medium was removed and the cells were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15710) in PBS for 10 min at room temperature. The cells were washed three times in PBS and Fluorobrite medium was added to the Mattek dish to achieve comparable background fluorescence levels. Cells were imaged with a Nikon Eclipse Ti-E inverted widefield microscope equipped with a Total Internal Reflection Microscopy iLAS2 module (Roper Scientific), a Perfect Focus System (Nikon) and motorized Z-Piezo stage (ASI) using a CFI APO TIRF 100 \times 1.49 NA oil immersion objective (Nikon). The microscope was operating in highly inclined and laminated optical sheet mode³⁴. Excitation sources were a 48-nm, 200-mW Topptica iBEAM SMART laser and a 561-nm, 200-mW Coherent Sapphire laser. Images were collected on two precisely aligned back-illuminated Evolve 512 Delta EMCCD cameras with a pixel size of $16 \times 16 \mu\text{m}^2$ (Photometrics). Cells were maintained at 37 °C and 8% CO₂ using an enclosed microscope environmental control set-up (The BOX and The CUBE, Life Science Instruments). Before the acquisition of movies for the dual-array set-up, TetraSpeck Microspheres, 0.1- μ m beads (Thermo Fisher Scientific, T7279), were imaged to allow for correction of chromatic aberrations during image processing and analysis. Movies for measurement of random TetO integrations in degron cell lines were acquired every 10 s (exposure time: 50 ms) in 34 z-planes (10- μ m stack, distance between consecutive z planes = 300 nm) with the Visiview software (Visiview 4.4.0.12, Visitron). Images for measurement of cell lines with the dual-array set-up were acquired every 30 s, with an exposure time of 50 ms, respectively, each in a sequential mode with 21 z-planes (6- μ m stack, dz = 300 nm). For the measurement of the time it takes the operator arrays to displace by their own size, images were acquired continuously on a single focal plane over 10 s every 0.1 s with exposure times of 50 ms.

Image processing

Raw images were deconvolved using the Huygens Remote Manager and a classical maximum likelihood estimation algorithm with a theoretical point-spread function. The initial SNRs were estimated from the images and images were deconvolved until one of the following stopping criteria was reached: the maximum number of iterations was performed (for random integrations: 20 cycles, for tdTomato and eGFP; in dual-color set-up: 15 cycles for tdTomato signal, 5 cycles for GFP signal) or a quality change criterion below 0.001 was returned. Representative image series shown in the main figures were deconvolved as described above, adjusted to display the same brightness and contrast, and interpolated using a bicubic interpolation. Movies were corrected for bleaching over time using an exponential fit. The two-dimensional (2D) projection of intensity changes over time was created using the Temporal Color Code in Fiji v.2.0. (https://github.com/fiji/fiji/blob/master/plugins/Scripts/Image/Hyperstacks/Temporal-Color_Code.ijm).

Spot detection and localization of multi operator data

Our field of view typically contains approximately 25 mESC nuclei. Despite the fact that our mESC lines are clonal, background nuclear fluorescence intensities in each cell can vary substantially. This poses challenges to conventional threshold-dependent algorithms for spot detection and localization which perform unevenly across cells with

different background intensities. To overcome these limitations, we implemented a two-step procedure for 3D spot detection and localization. To detect spots, we used deepBlink v.0.1.1 (ref.³⁵), a convolutional neural network-based spot detection and localization algorithm in two dimensions, which has been shown to be able to deal with different background intensities and to detect spots in a threshold-independent manner. To enhance our detection efficiency, we employed custom models trained on a combination of the following datasets: smFISH and SunTag datasets provided by deepBlink and in-house manually curated live-cell imaging images. To detect 3D spots, we applied deepBlink to all z-stacks separately followed by linkage of the spots across z-stacks using Trackpy³³. The precise 3D coordinates of the spots were then determined using 3D Gaussian fitting using a voxel of size $6 \times 6 \times 4$ pixels centered at the spot in the brightest z-stack. deepBlink models can be found at https://github.com/zhanyinx/SPT_analysis/tree/main/models. The parameters and models used for each cell line can be found in Supplementary Table 4. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

Tracking and cell motion correction of multi operator data

3D spots coordinates are fed into TrackMate for tracking using linear assignment problem (LAP) tracker. Each track is assigned to manually annotated cell masks (from max z-projection of frame 93) using a custom script (https://github.com/zhanyinx/SPT_analysis/blob/main/source/spot_detection_tracking/assign_cellids.py), which uses the majority rule. Motion correction is then performed using a roto-translation model. Specifically, for each pair of consecutive time frames, a set of matching spots in every cell is determined by solving the LAP using the Euclidean distance between spots as a measure of distance. Only spots that match across two consecutive frames are then used to estimate the roto-translation model which is then applied to correct for nuclear motion (six matching spots on average across all time frames, trajectories and movies, with a minimum of four spots per pair of time frames). All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

MSD analysis of multi operator data

Tracks with fewer than ten spots are filtered out for follow-up analysis. To calculate the MSD, we first calculate the time-averaged MSD for each trajectory. We then calculate the ensemble average (across trajectories) MSD by pooling all replicates. The ensemble average is done in log space. We corrected the localization error effect on the MSD curve by estimating the standard deviation of the error distribution using fixed images as described by Kepten et al.⁵⁵. To calculate the scaling (α) and the generalized diffusion coefficient (D) of each MSD curve, we fitted the ensemble average of the log-time average MSD between 10 and 100 s. To test the significance of differences between conditions, we fitted α and diffusion coefficient for each cell. The P value is calculated using Student's t -test (two-sided). Since we are always comparing two conditions whose cell-cycle profiles are similar, we ignore the effect of sister chromatids. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/. The specific Fiji and relative plug-ins can be found at https://github.com/giorgetti/Mach_et_al_chromosome_dynamics/tree/master/Fiji.

Chromatic aberration correction of dual-color data

To correct for chromatic aberration we took 3D image stacks of TetraSpeck Microspheres, 0.1- μ m beads (Thermo Fisher Scientific, T7279), adsorbed on MatTek dishes in $1 \times$ PBS at the beginning of every imaging session and used them to correct the corresponding set of movies. After detecting signals from single beads in each channel using deepBlink and determining their 3D location by Gaussian fitting, we first identified spots that are shared across channels by solving the LAP using the Euclidean distance between spots. We then used the common set of bead signals to compute a 3D roto-translation that we

finally applied to xyz positions. This procedure corrects for x , y and z aberrations simultaneously. The same transformations accurately corrected chromatic aberrations in actual experiments in double-labeled mESCs ('Control TetO' in Extended Data Fig. 6g), with the exception of a small residual systematic shift (approximately 40 nm) along the z axis ('TetO-LacO case' in Extended Data Fig. 6g), which is likely due to 3D image anisotropies that cannot be measured using '2D' bead images.

Tracking and MSD analysis of dual-color data

To increase the ability to detect longer tracks, we used an in-house script to stitch multiple tracks belonging to the same cell (https://github.com/zhanyinx/SPT_analysis/blob/main/source/dual_channel_analysis/utis.py, stitch function). In short, if two tracks from the same cell overlap more than 50% in time, the shortest one is filtered out. We called cell masks using CellPose⁵⁶ on the max z -projection of the middle frame of the movie using the GFP channel and used these masks to define cell identity. For tracks with overlaps lower than 50%, the overlapping part of the tracks are randomly removed from one of the two tracks. The resulting tracks are stitched if the distance across the time gap is smaller than 1.6 μm . To match tracks across channels, we used the following measure to calculate the distance between tracks across channels:

$$\frac{\sum_{i=1}^3 \langle (x_{1i}(t) - x_{2i}(t))^2 \rangle_{t \in T_1 \cap T_2}}{\sqrt{\text{len}(t \in T_1 \cap T_2)}}$$

Where x_i are the coordinates from channel 1 and x_j are the coordinates from channel 2, T_1 contains all the time frames from channel 1 and T_2 contains all the time frames from channel 2, and len is a function that returns the length of an array. We solved the LAP using the distance measure above to match tracks across channels. Tracks with average distances across channels higher than 1 μm are filtered out. Matched tracks with lower than 25 time points are filtered out. For each matched pair of tracks, we calculate the pairwise distance using the Euclidean distance in three dimensions. We define noisy pairwise distance using the ratio of the pairwise distance in three dimensions and two dimensions. In particular, we defined as noisy the top 5% of this ratio and filtered them out. To calculate the radial MSD, we first calculate the time-averaged radial MSD for each pairwise distance 'trajectory'. We then calculate the ensemble average (across trajectories) of the log of time-averaged radial MSD. We corrected for the radial localization uncertainty by estimating the standard deviation of the error distribution using fixed images as described by Kepten et al.⁵⁵. To calculate the scaling (α) and the generalized diffusion coefficient (D) of each MSD curve, we fitted the ensemble average time average MSD between 30 and 300 s. Since we are always comparing two conditions whose cell-cycle profiles are similar, we ignore the effect of sister chromatids. All scripts used for the analysis can be found at https://github.com/zhanyinx/SPT_analysis/.

Estimation of experimental uncertainty on radial distance

To estimate our uncertainty in detecting distances across channels, we used a cell line with multiple integration of TetO arrays that can be tagged with TetR-eGFP and TetR-tdTomato. Spot detection is done as for our dual-color lines. We corrected for chromatic aberration using TetraSpeck Microspheres, 0.1- μm beads (Thermo Fisher Scientific, T7279), and then matched spots across channels by solving the LAP using `scipy.optimize.linear_sum_assignment` function with the Euclidean distance between spots as a measure of distance. Spots across channels with distances higher than a threshold are filtered out to avoid mismatches. We used a threshold of 300 nm for matching the spots registration. We applied a second round of chromatic aberration correction using the set of registered points themselves. The resolution

limit (uncertainty) is then estimated as the average distance between registered spots which corresponds to 130 ± 70 nm.

HMM for detection of the proximal state

To detect the proximal state in a threshold-independent manner, we used an HMM with two hidden states ('proximal' and 'distal'). We used a Gaussian model for the emission probabilities. Only distance trajectories with less than 20% missing values at any time point are kept. Missing values are filled with the first preceding time point with distance value. To more reliably detect the proximal state, we used all the trajectories from the experimental condition with both cohesin and CTCF sites to train an HMM. We then re-trained an HMM model for each experimental condition by using the proximal state (Gaussian mean and standard deviation) from the experimental condition with both cohesin and CTCF sites. Finally, we applied the experimental condition-specific HMM to every trajectory to estimate the contact duration and rate of contact formation for all the experimental conditions. The HMM model training can be found as a jupyter notebook (https://github.com/zhanyinx/SPT_analysis/blob/main/notebooks/HMM_experimental_data.ipynb). We modified the `hmmlearn` library to allow fixing proximal state during HMM training. The modified `hmmlearn` library can be found at <https://github.com/zhanyinx/hmmlearn>.

Simulations

Polymer simulations were performed using LAMMPS⁵⁷. We chose Langevin dynamics with the NVT thermostat. Arbitrary units were set such that thermal energy $k_B T = 1$, where k_B is the Boltzmann constant and T is room temperature, corresponding to 300 K. For every set of parameters, we performed ten independent runs. A run consists of an equilibration part of 10^7 simulation steps and a production part of 10^8 simulation steps. For subsequent analysis and calculation of contact maps, we recorded the data every 10^4 simulation steps. In simulations for Fig. 2, the chain length was 1,125 beads. In simulations for Figs. 5 and 6, the chain length was 1,000 beads. We used PyMOL software (v.2.3.3) to represent snapshots of polymer chain in Fig. 2a. Examples of initial conformations and simulation parameters can be found at https://github.com/gjorgettilab/Mach_et_al_chromosome_dynamics, in the polymer simulations section.

To simulate the loop extrusion process, we developed and embedded in LAMMPS a package called 'USER-LE'. The loop extrusion model contains extruders and barriers on the polymer. An extruder is represented as an additional sliding bond, which extrudes the loop in a two-sided manner. It can be loaded to the polymer between (i) and ($i + 2$) beads with a certain probability only when the bead ($i + 1$) is unoccupied by another extruder and is not a barrier. Each extruder can be unloaded from polymer with a certain probability. Every bead can be occupied by only one extruder. Extruders cannot pass through each other. When extruders meet each other on the polymer, they stall until one of them is released. Every extruder attempts to make an extruding step every N simulation steps.

In addition to 'neutral' polymer beads, there are three types of barriers blocking loops coming from the left, from the right and from any direction. These barriers mimic CTCF sites, for which one can define a probability for the loop extruder to go through (the same probability for all barriers). To launch loop extrusion, one should define three fixes with LAMMPS syntax: loading, unloading and loop extrusion. Loading: frequency in number of steps to try to load extruders, types of beads, max distance to create, type of the bond (extruder) to be created, probability to create, seed for pseudorandom generator of numbers, new type of the first beads and new type for the second bead. Unloading: frequency in number of steps to try to unload extruders, type of the bond (extruder), min distance to release bond, probability to release bond, seed for pseudorandom number generator. Loop extrusion: frequency in number of steps to try to move extruders, neutral polymer type, left barrier type, right barrier type, probability to go

through the barrier, type of the bond (extruder) and type of two-sided barrier (optional).

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. No randomization was performed as the study did not require sample allocation into different groups. Live-cell imaging experiments were performed in 3–7 biological replicates and all replicates showed consistent results. For Capture-C, Hi-C, 4C-seq, piggyBac insertion site mapping and Nanopore sequencing with Cas9-guided adapter ligation, one biological replicate was performed. For flow cytometry measurements two biological replicates were performed. Western blot analysis and genotyping PCR with subsequent agarose gel electrophoresis were performed with 1–2 biological and 2 technical replicates. Blinding was not possible for data collection in live-cell imaging experiments, as data acquisition required identification of the sample for further processing. Data analysis for live-cell imaging, Capture-C, Hi-C, 4C-seq and piggyBac insertion site mapping were performed in a blinded manner. Blinding was not necessary for the other experiments since the results are quantitative and did not require subjective judgment or interpretation. Whenever Student's *t*-test was used, we formally verified the normality of distributions but assumed variance equality.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All Capture-C, Hi-C, 4C-seq and integration site mapping sequencing fastq files generated in this study have been uploaded to the Gene Expression Omnibus (GEO) under accession [GSE197238](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197238). The following public database was used: BSgenome.Mmusculus.UCSC.mm9 (<https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Mmusculus.UCSC.mm9.html>). The trajectories from imaging data can be found at <https://doi.org/10.5281/zenodo.6627715>. Source data are provided with this paper.

Code availability

Custom codes generated in this study are available at: https://github.com/zhanyinx/SPT_analysis/ (image analysis); https://github.com/giorgetti/Mach_et_al_chromosome_dynamics/ (4C-seq, Hi-C, nanopore, simulation analysis); https://github.com/polly-code/lammms_le (repository with loop extrusion module for the LAMMPS); and <https://github.com/zhanyinx/hmmllearn> (the modified version of hmmllearn).

References

- Allan, D. B., Caswell, T., Keim, N. C., van der Wel, C. M. & Verweij, R. W. *soft-matter/trackpy: Trackpy v0.5.0* (2021).
- Redolfi, J. et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019).
- Kepten, E., Bronshtein, I. & Garini, Y. Improved estimation of anomalous diffusion exponents in single-particle tracking experiments. *Phys. Rev. E* **87**, 052713 (2013).
- Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

- Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).

Acknowledgements

We thank all members of the Giorgetti laboratory for help with labeling images for training the deepBlink spot-detection algorithm and feedback on the manuscript; A. S. Hansen, L. Mirny and members of their laboratories (in particular S. Grosse-Holz) for useful discussions of the results; E. de Wit for sharing the WAPL-AID-eGFP cells; H. Kohler for assistance with flow cytometry experiments; and R. Grand and D. Schübeler for sharing single-molecule footprinting data. P.M. was supported by a Marie Skłodowska-Curie Innovative Training Network (grant no. 813327 'ChromDesign'). J.T. was supported by a Marie Skłodowska-Curie Innovative Training Network (grant no. 813282 'PEP-NET'). J.Z. was supported by a Marie Skłodowska-Curie grant (no. 748091 – 3DQuant). Research in the Giorgetti laboratory is funded by the Novartis Foundation, the European Research Council (grant no. 759366, 'BioMeTre'), Marie Skłodowska-Curie Innovative Training Networks (grant nos. 813327 'ChromDesign' and 813282 'PEP-NET') under the European Union's Horizon 2020 research and innovation program, and the Swiss National Science Foundation (grant no. 310030_192642).

Author contributions

P.M. and P.I.K. conceived the study with L. Giorgetti and with the help of discussions with E.P.N. P.M. performed the experiments with help from J.T., J.C., S.G., J.Z. and M.K. P.I.K. developed the code and performed and analyzed polymer simulations with input from E.M. and G.T. Y.Z. and P.I.K. performed image and trajectory analysis with help from E.M., G. R. and G.T. L. Gelman and J.E. contributed to setting up and provided assistance with microscopy and image analysis. S.S. performed piggyBac insertion site mapping and provided assistance with high-throughput sequencing experiments. P.I.K. and Y.Z. analyzed sequencing data. L. Giorgetti wrote the paper with P.M., P.I.K., Y.Z. and G.T. and with input from all the authors.

Competing interests

All authors declare no competing interests.

Additional information

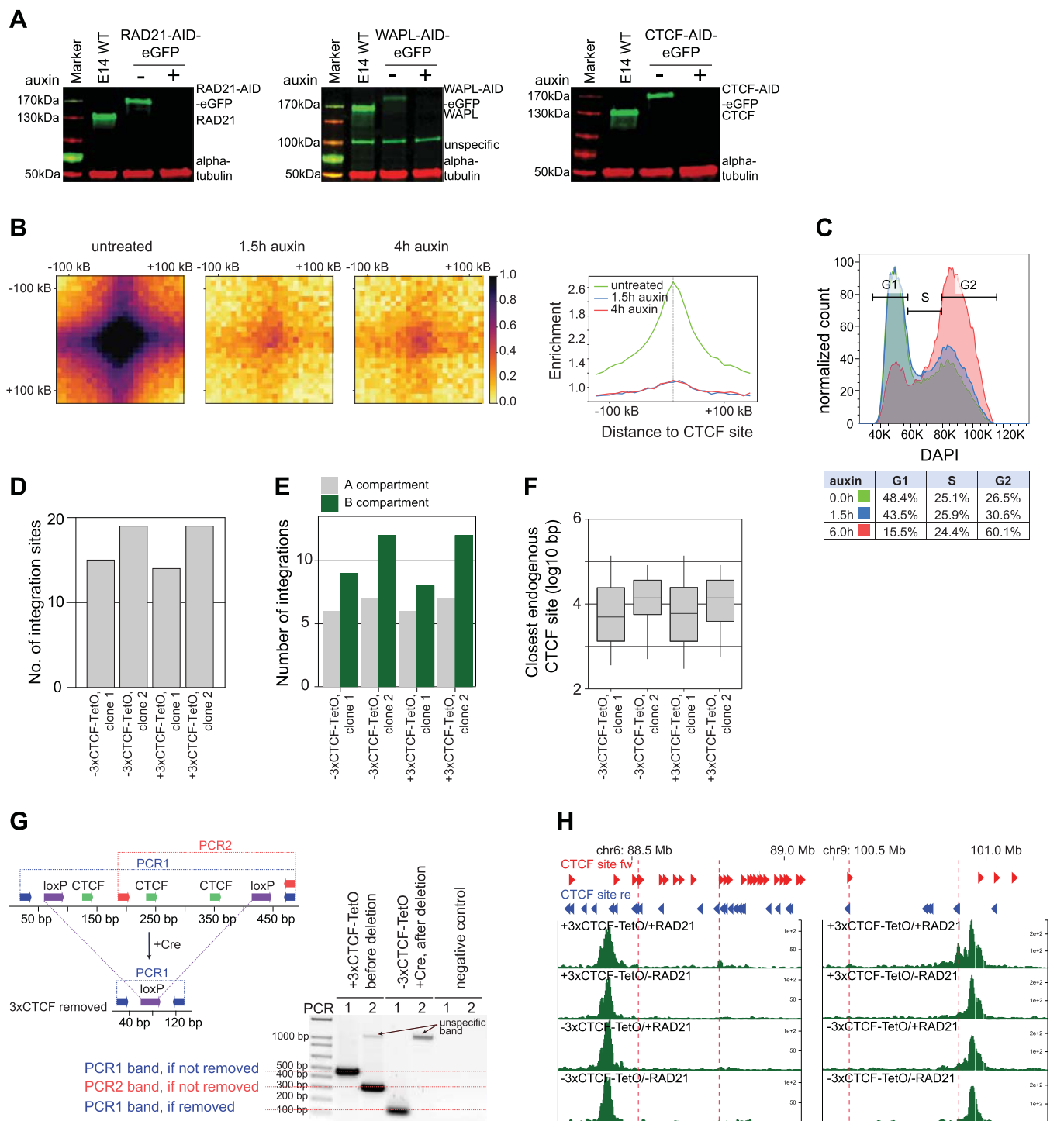
Extended data is available for this paper at <https://doi.org/10.1038/s41588-022-01232-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01232-7>.

Correspondence and requests for materials should be addressed to Luca Giorgetti.

Peer review information *Nature Genetics* thanks Marcelo Nollmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

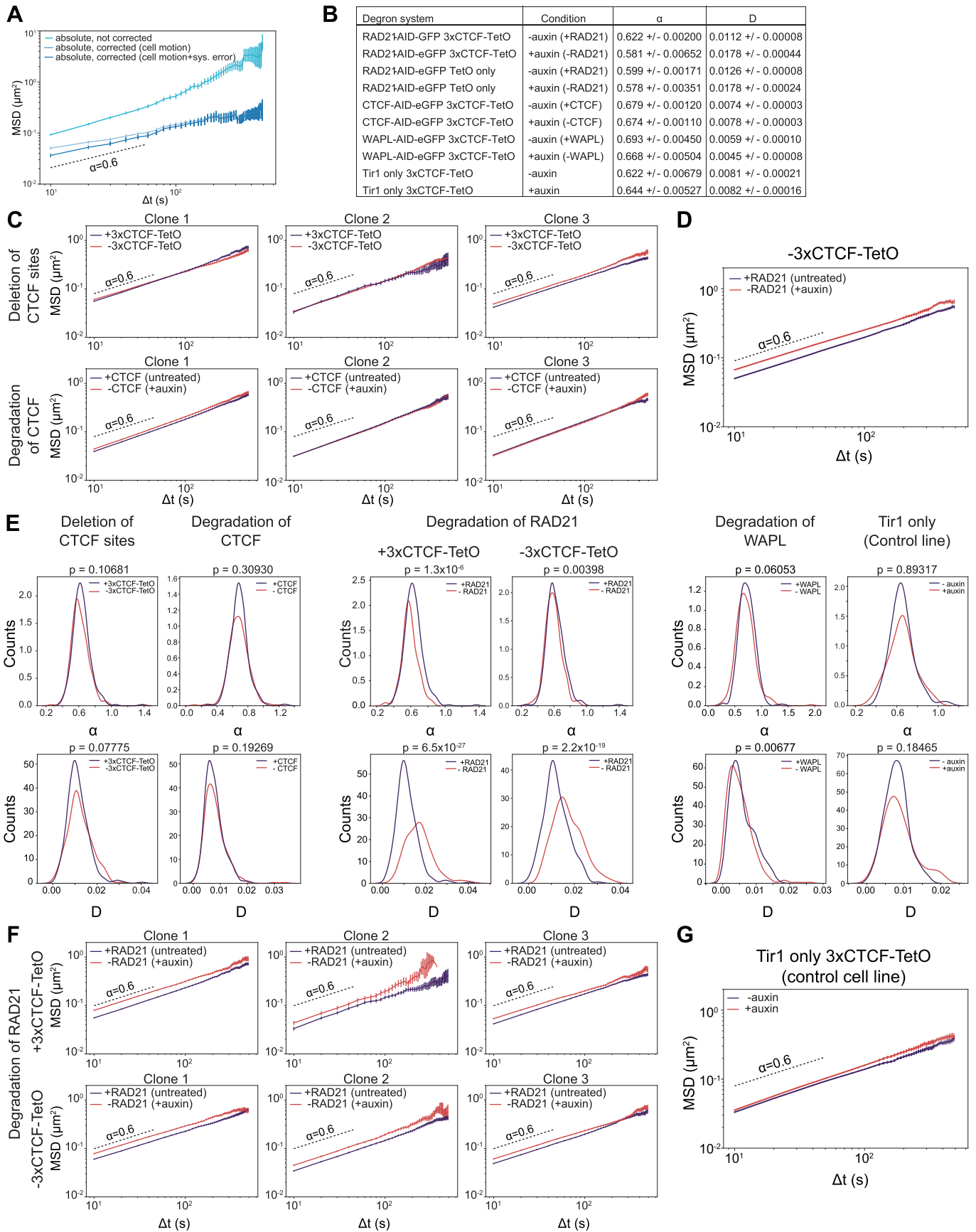
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Chromosome structure is altered upon degradation of factors involved in loop extrusion.

A. Western Blots showing degradation of RAD21, WAPL and CTCF upon 1.5 h, 24 h and 6 h, respectively. Loading control: α -tubulin, $n = 1-2$ replicates for each cell line. **B.** Left: Average enrichment in Hi-C read counts at CTCF sites based on Hi-C data in RAD21-AID-eGFP cells either untreated (left), treated for 1.5 h (middle) or 4 h (right) with auxin. Right: Differences in enrichment at CTCF peaks. Peaks were called on Hi-C data from untreated cells. **C.** Flow cytometry analysis of fixed cells stained with DAPI showing cell-cycle stage distributions of RAD21-AID-eGFP mESC cultured with serum, LIF and 2i, either before (green) or after 1.5 h (blue) and 6 h (red) auxin treatment. **D.** Integration site numbers in two clones of RAD21-AID-eGFP lines with and without 3xCTCF sites. **E.** Distribution of integration sites from lines shown in panel D that belong to A and B compartments called on distance-normalized Hi-C map (same as panel B). **F.** Integration sites distances from the

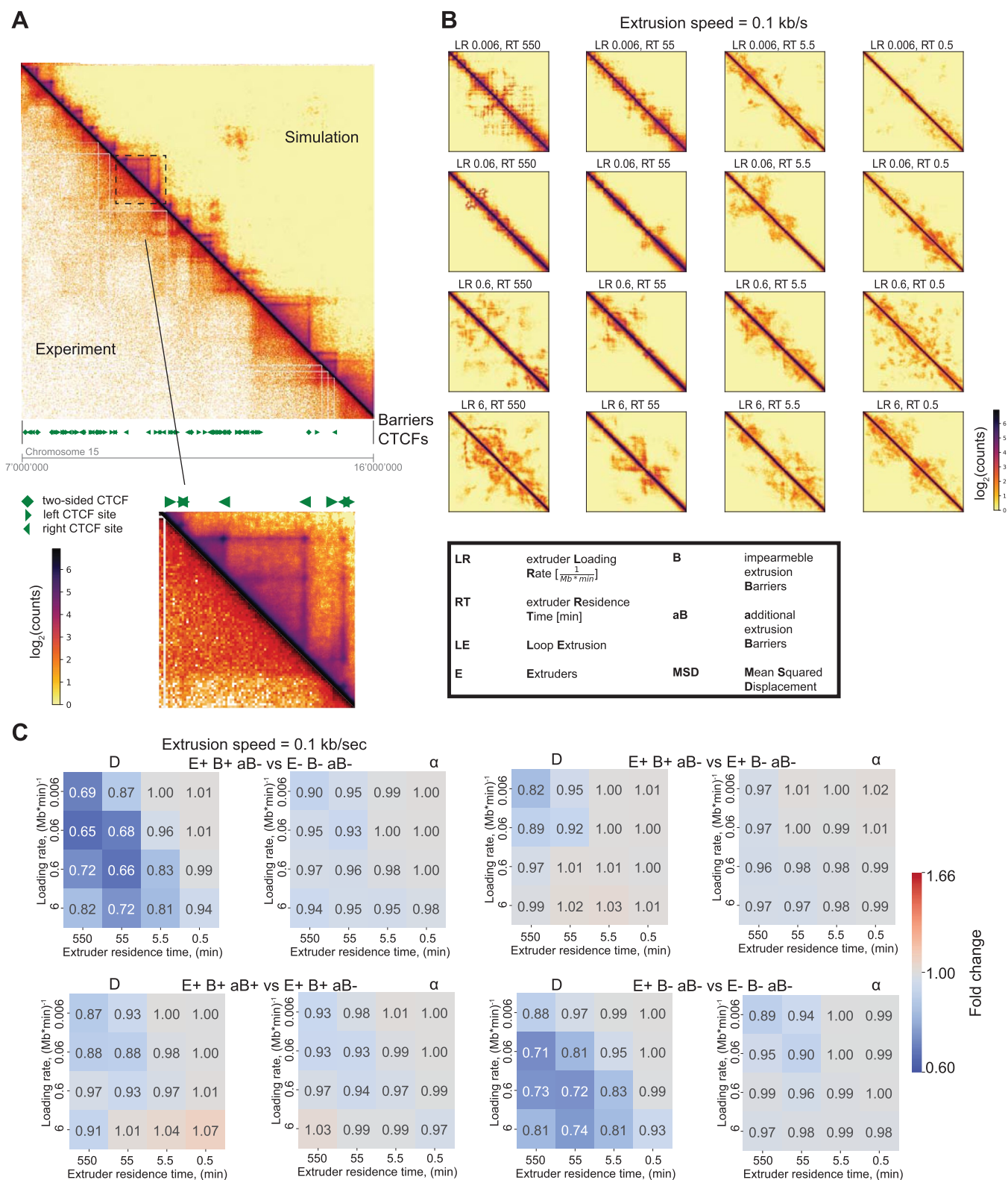
closest endogenous CTCF site. Boxplot: lower and upper quartiles (Q1 and Q3, respectively); whiskers: 1.5x interquartile region (IQR) below Q1 and above Q3. $n = 15$ and 19 insertions for -3xCTCF-TetO clones 1 and 2, respectively, $n = 14$ and 19 insertions for +3xCTCF-TetO clones 1 and 2, respectively. **G.** Example of genotyping PCR upon removal of 3xCTCF sites in a RAD21-AID-eGFP +3xCTCF-TetO clonal line. PCR1 amplifies the entire 3xCTCF cassette and product size changes from 470 bp to 147 bp if the cassettes are successfully removed. PCR2 amplifies half of the 3xCTCF cassette and no product is expected if 3xCTCF cassettes were removed from all insertion sites; otherwise a PCR band of 303 bp is expected. **H.** Representative 4C-seq profiles from insertions on chromosomes 6 and 9 using TetO as a viewpoint showing that 3xCTCF-TetOs lead to the formation of ectopic contacts (dashed red lines) with nearby endogenous CTCF sites in the presence of RAD21. Contacts are lost upon deletion of 3xCTCF cassette (-3xCTCF-TetO) and upon degradation of RAD21 (-RAD21).



Extended Data Fig. 2 | See next page for caption.

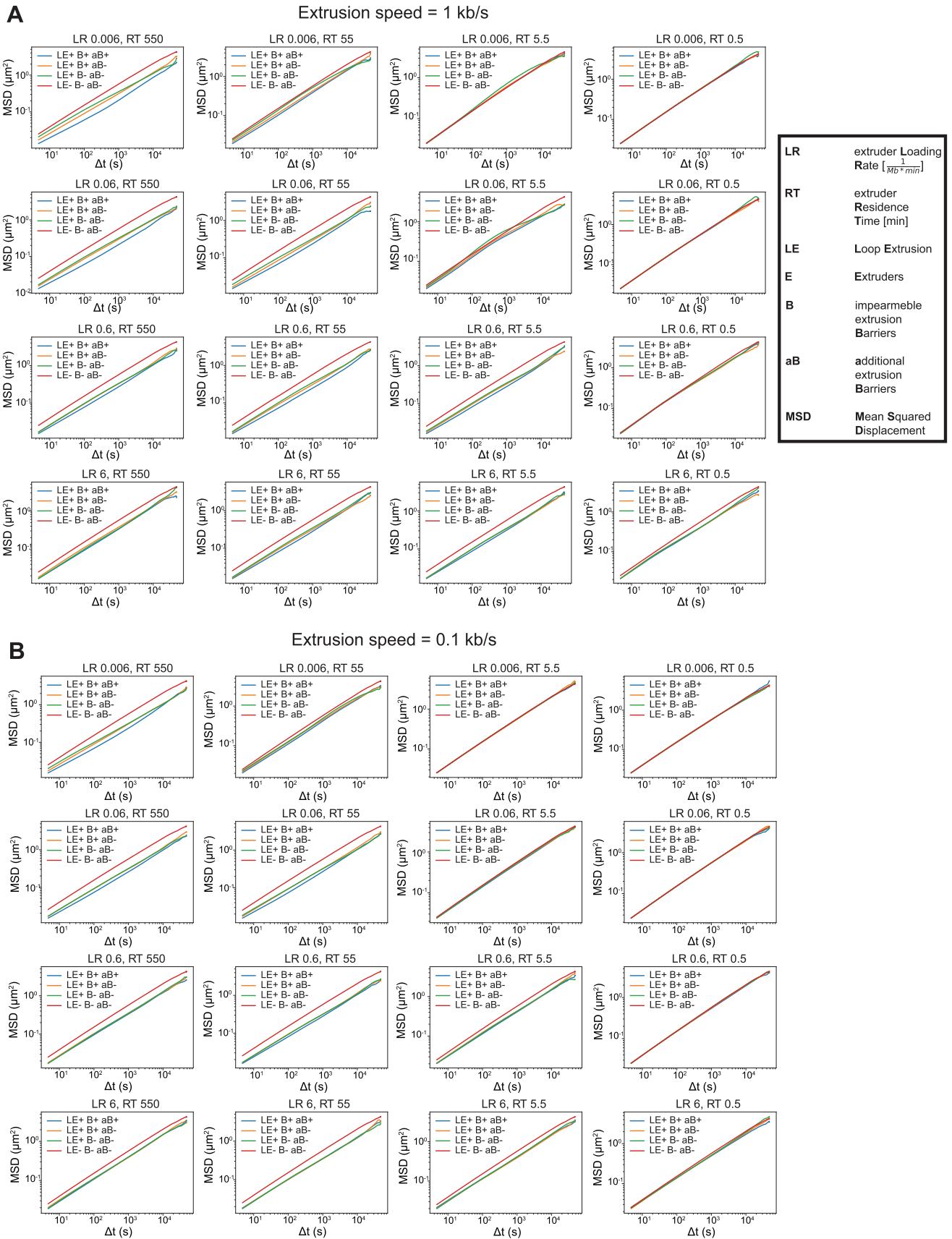
Extended Data Fig. 2 | Chromosome dynamics is modulated by degradation of factors involved in loop extrusion. **A.** Mean Square Displacement (MSD) of trajectories from TetO insertions within the same cell (MSD, mean \pm s.e.m., $n = 45$ tracks) before (cyan) and after applying cell motion (light blue, $n = 45$ tracks) and localisation error correction (dark blue, $n = 45$ tracks). **B.** Scaling exponents (α) and generalized diffusion coefficients (D) across all conditions and cell lines were fitted by pooling all three biological replicates. Shown are the numbers for the best fit \pm error of the fit. **C.** MSD (mean \pm s.e.m.) plots for a single clonal cell line (biological replicate) when looking at removal of 3xCTCF sites (top row) next to the array or degrading all CTCF (bottom row). **D.** MSD (mean \pm s.e.m.) in the cell lines ($n = 3$ replicates per clonal cell line, three cell lines) where the 3xCTCF cassette was excised. Shown are the MSDs for cells either depleted of RAD21 for 90 min (red, 266 cells, 9,020 trajectories analyzed) or not (blue, 271 cells, 11,082

trajectories analyzed). Global depletion of RAD21 increases mobility. p -values in panel E. **E.** Distributions of α and D fitted based on single trajectory MSD and significance test for differences in generalized diffusion coefficients (D) and scaling exponents (α). The p -value is calculated using Student t-test (two-sided) (see Methods). **F.** Same as in C for a single clonal cell line (biological replicate) with integrations with 3xCTCF-TetO (top row) or without 3xCTCF-TetO (bottom row) when degrading RAD21. Global depletion of RAD21 increases mobility. **G.** Same as in D in the cell lines that contain integrations of 3xCTCF-TetO and the TIR1 protein, but do not contain any AID-tag for targeted degradation. MSDs for cells either treated with auxin for 90 min (red, 97 cells, 2,155 trajectories analyzed) or not (blue, 111 cells, 3,711 trajectories analyzed). No significant changes were detected. p -values in panel E.

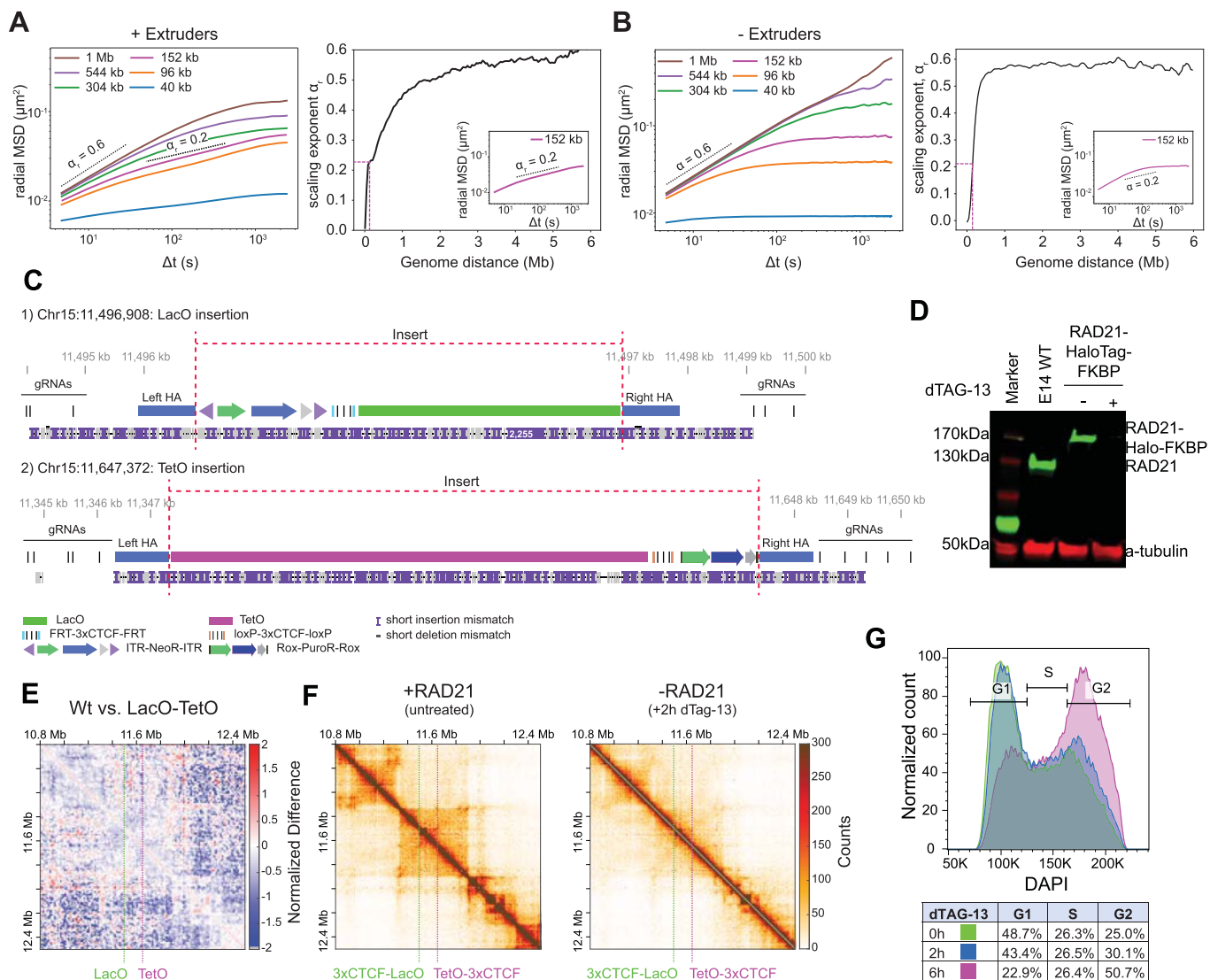


Extended Data Fig. 3 | Simulations of chromosome dynamics and effects of loop extrusion. A. Visual comparison of experimental Hi-C contact map with contact maps of simulations at extrusion speed 1 kb/s, extruder loading rate $0.06 \text{ (Mb} \times \text{min)}^{-1}$ and residence time 5.5 min. **B.** Contact maps for the polymer simulations at extrusion speed 0.1 kb/s and barriers from the range 7–16 Mb of

chromosome 15. Acronyms used in this figure are indicated in the black box on the right. **C.** Pairwise comparison for conditions indicated in the title of each pair of heatmaps. Pair of heatmaps contains ratios of generalized diffusion coefficients (D) and scaling exponent (α), and represents fold change between the conditions.



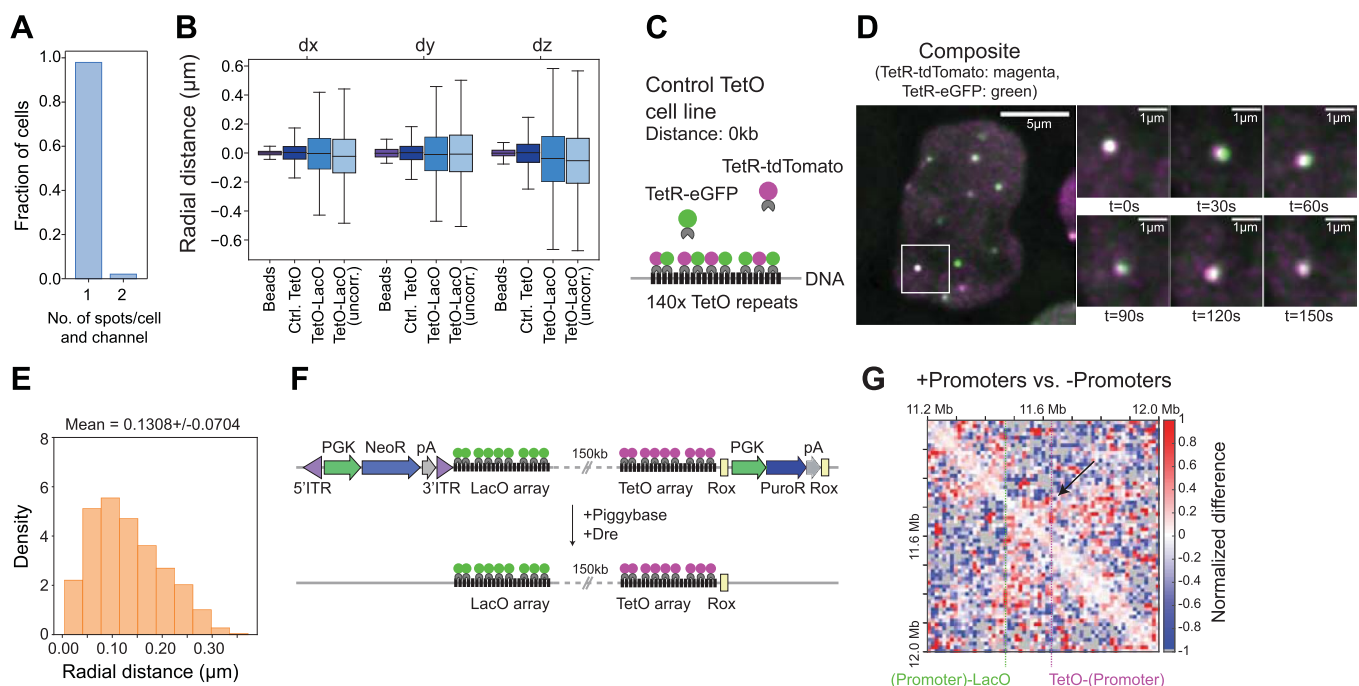
Extended Data Fig. 4 | MSDs of systems for two extruder speeds. A. MSDs for all 16 conditions for each set of loop extrusion parameters and extrusion speed of 1 kb/s. **B.** Same as A but for the extrusion speed of 0.1 kb/s.



Extended Data Fig. 5 | Characterization of TetO and LacO array integrations.

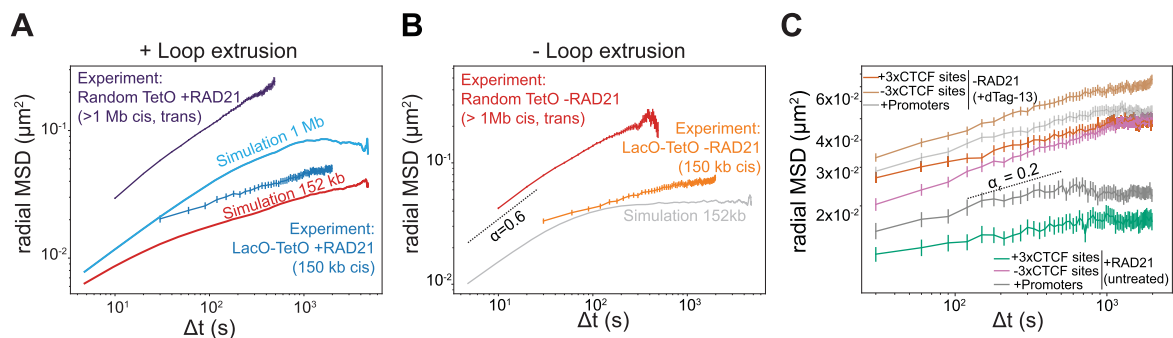
A. Left panel: radial MSD of distances between multiple pairs of monomers separated by distances equivalent to 40 kb - 1 Mb for a polymer with loop extrusion but no barriers. Dashed scaling exponents $\alpha = 0.2$ and $\alpha = 0.6$ serve as an eye guide. Right panel: Slopes of radial MSD curves for two loci separated by varying linear distances, estimated from linear fitting between 5 and 60 seconds. Inset: detail of radial MSD and fit for monomers separated by 152 kb. **B.** Left panel: radial MSD of multiple pairs of monomers separated by various distances (40 kb-1 Mb). Simulations were performed for the polymer without extruders and barriers. Values were averaged with a sliding window without considering the first and last 200 monomers (1.6 Mb). Dashed scaling exponent $\alpha = 0.6$ serves as an eye guide. Right panel: Distance dependency of the scaling exponent (α) on the genomic distance between loci. **C.** Integrated Genomic Viewer (IGV) snapshot showing an example of a Nanopore sequencing read mapped to a modified

mouse genome including the respective insertions. Reads that spanned from a guide RNA (gRNA) binding site upstream of the left homology arm (left HA) to a gRNA binding site downstream the right homology arm (right HA) confirmed single insertion of the transgene. **D.** Western Blots showing the targeted degradation of RAD21 after 2 h of treatment with 500 nM dTAG-13. Loading control: anti-tubulin, n = 2 replicates. **E.** Differential map at 6.4 kb resolution for the structural differences between a E14 wild-type (WT) and the E14 cell line containing LacO and TetO insertions (see Methods). Dashed lines indicate the insertion sites. No structural changes are detected upon integration of the operator arrays. **F.** Capture-C maps at 6.4 kb resolution in the region on chr15 (10.8 Mb-12.5 Mb) in the untreated cells (left) and in cells treated with 500 nM dTag-13 (left) showing that RAD21 degradation leads to loss of chromosome structure. **G.** Flow cytometry analysis of fixed cells stained with DAPI to show cell cycle stage distribution of E14 RAD21-HaloTag-FKBP cells.



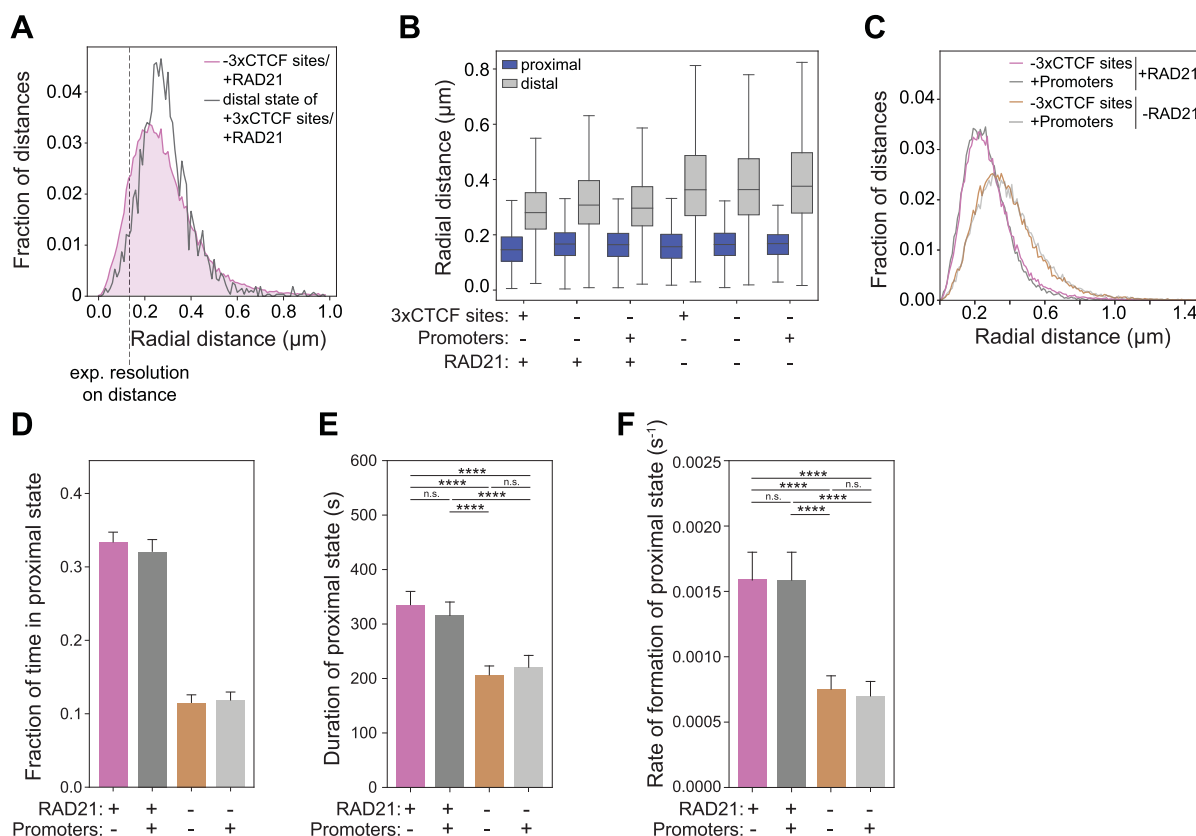
Extended Data Fig. 6 | Correction of chromatic aberrations and characterization of mESC lines with promoters flanking TetO and LacO arrays. **A.** Bar plot showing the number of detected spots per cell per channel for 1,400 manually annotated images subsampled from the images series. In 3% of the images 2 spots per cell are detected indicating the presence of sister-chromatids. **B.** Distribution of pairwise distances in each dimension for the control TetO cell line ($n = 69,453$ timepoints), as well as for chromatic-aberration corrected and uncorrected images from TetO-LacO cell lines (in the presence of cohesin and 3xCTCF sites, $n = 848,955$ timepoints). Boxplot: boxes denote lower and upper quartiles (Q1 and Q3, respectively); whiskers denote 1.5x the interquartile region (IQR) below Q1 and above Q3. **C.** Schematic representation of the 'Control TetO' cell line that contains multiple TetO array integrations as well as stable integrations of TetR-eGFP and TetR-tdTomato. This allows labeling of each TetO array with two separate fluorophores. **D.** Representative images of the 'Control

TetO' cell line. The time series shows a zoomed version of the region indicated by the white square. **E.** Radial distance distribution of the 'Control TetO' cell line as defined in panel C and D showing that the resolution on the 3D distance is ~130 nm. **F.** Schematic representation of cell line containing 3-phosphoglycerate kinase (PGK) promoters driving the expression of resistance gene directly adjacent to the operator arrays. The expression cassettes can be excised using Dre recombination or piggyBac transposition to yield the cell line with operator arrays only (PGK = PGK promoter, NeoR = Neomycin resistance gene, PuroR = Puromycin resistance gene, pA = polyadenylation signal, ITR = inverted terminal repeats for piggyBac recognition, Rox = Rox sites for Dre recombination). **G.** Differential map at 6.4 kb resolution for the structural differences between the E14 cell line containing LacO and TetO insertions with the adjacent promoters vs. the E14 cell line containing the operator arrays only (see Methods). Dashed lines indicate the insertion points.



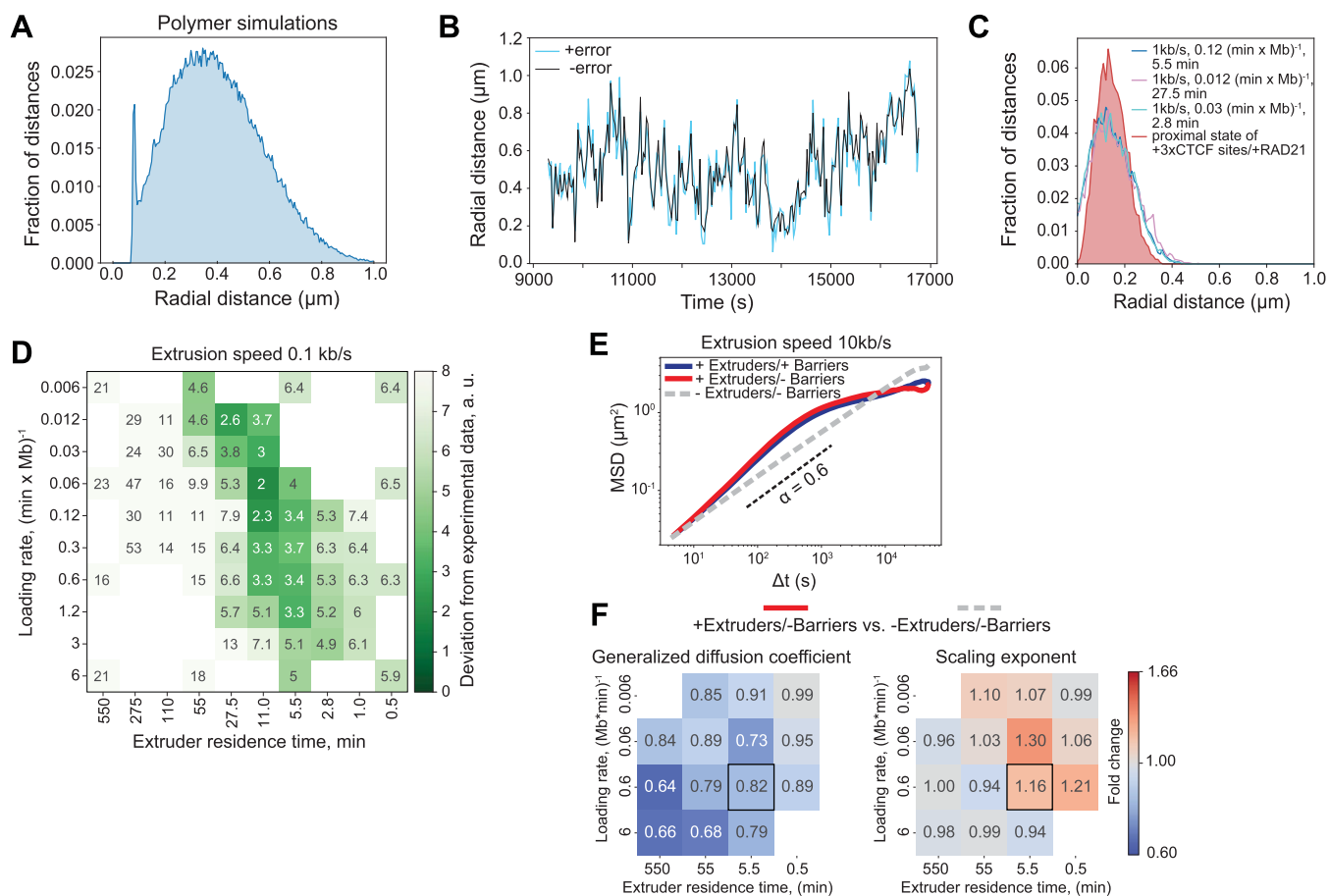
Extended Data Fig. 7 | Polymer simulations of two genomic locations within the same TAD. **A.** Radial MSD of TetO random integrations (mean \pm s.e.m., purple, see Fig. 1E, $n = 271$ cells examined over 3 pooled biological replicates) and of targeted LacO and TetO insertions on Chr15 (mean \pm s.e.m., dark blue, $n = 214$ cells examined over 4 replicates) are compared to model predictions for pairs of loci containing extrusion barriers at a distance of 1 Mb (light blue) and 152 kb (red). Note that random TetO insertions often occur on different chromosomes and thus have larger absolute radial MSD than 1 Mb simulations (but similar scaling). **B.** Radial MSD for cell lines containing multiple random integrations of TetO as shown in Extended Data Fig. 2D (mean \pm s.e.m., red, 266 cells examined over 3 pooled replicates) or the targeted integrations of LacO and TetO on chr15 (mean \pm s.e.m., orange, $n = 277$ cells examined over 6 replicates) in the absence of

RAD21 compared to the predicted radial MSD of two loci at a distance of 150 kb in the absence of extruders (gray) as predicted from polymer simulations. **C.** Radial MSD of TetO-LacO distances in mESC lines with or without convergent 3xCTCF sites (or promoters, respectively), either before or after treatment with 500 nM dTag-13 for 2 hours to induce degradation of RAD21 ($dt = 30$ s). radial MSDs are plotted as mean \pm s.e.m. over conditions: +CTCF sites/+RAD21: $n = 152$ cells examined over 4 replicates, -CTCF sites/+RAD21: $n = 214$ cells examined over 4 replicates, +CTCF sites/-RAD21: $n = 248$ cells examined over 7 replicates, -CTCF sites/-RAD21: $n = 277$ cells examined over 6 replicates, +Promoters/+RAD21: $n = 155$ cells examined over 3 replicates, +Promoters/-RAD21: $n = 170$ cells examined over 3 replicates.



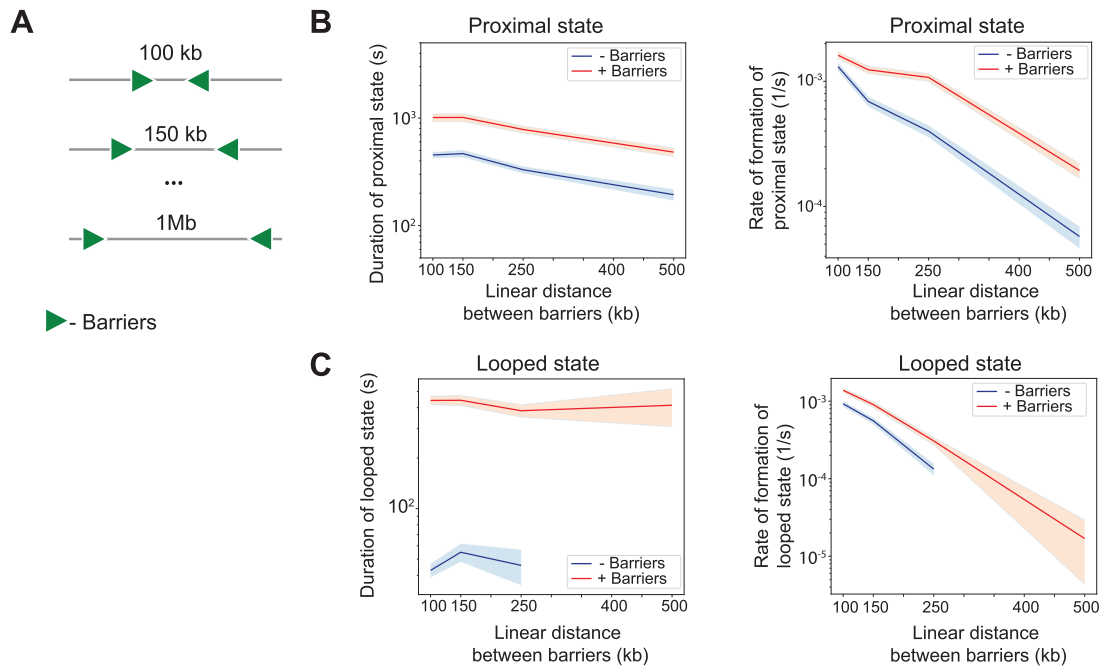
Extended Data Fig. 8 | Live-cell imaging of two genomic locations within the same TAD. A. Radial distance distribution for the condition -3xCTCF sites/+RAD21 (magenta) overlaid with the distal state called by HMM on the +3xCTCF sites/+RAD21 (gray) showing that the distal state identified by HMM largely overlaps with the distance distribution of the two loci in the absence of the CTCF sites. **B.** Boxplot for the radial distances for the proximal and distal state called by HMM on all six conditions. The horizontal line indicates the median. Box plots are as in Extended Data Fig. 1F. Boxplot: boxes denote lower and upper quartiles (Q1 and Q3, respectively); whiskers denote 1.5x the interquartile region (IQR) below Q1 and above Q3. **C.** Distribution of TetO-LacO radial distances in the four experimental conditions. -CTCF sites/+RAD21: n = 214 cells examined over 4 replicates, -CTCF sites/-RAD21: n = 277 cells examined over 6 replicates, +Promoters/+RAD21: n = 155 cells examined over 3 replicates, +Promoters/-RAD21: n = 170 cells examined over 3 replicates). **D.** Fraction of time

spent in the proximal state called by HMM in the four experimental conditions comparing +Promoters vs. -Promoters +/-RAD21 (no. of cells is as indicated in panel C). Shown average across experimental conditions and error bars represent bootstrapped (n = 10,000) standard deviations. **E.** Average duration of proximal states (mean \pm 95% confidence interval, n = 680 cells (-promoter +RAD21); n = 466 cells (+promoter +RAD21); n = 268 cells (-promoter -RAD21); n = 253 cells (+promoter -RAD21)) for the conditions +Promoters vs. -Promoters, +/-RAD21. p-values (two-sided Kolmogorov-Smirnov): * - p < 0.05, ** - p < 0.01, *** - p < 0.001, **** - p < 0.0001. p-values can be found in Suppl. Table S2. **F.** Average rates of contact formation - time elapsed between the end of a proximal state and the beginning of the next (mean \pm 95% confidence interval, n = 726 (-promoter +RAD21); n = 495 (+promoter +RAD21); n = 323 (-promoter -RAD21); n = 296 (+promoter -RAD21)) for the conditions +Promoters vs. -Promoters, +/-RAD21. p-values legend is as in panel E.



Extended Data Fig. 9 | HMM analysis of simulations compared to experimental data. **A.** Bimodal distribution of pairwise distances from simulations corresponding to the set of parameters with a loading rate of $0.06 \text{ (min} \times \text{Mb)}^{-1}$, extruder residence time of 5.5 min, extruder speed of 1 kb/s, and in the absence of barriers. Data were sampled every 1 s and merged from 10 simulation runs. **B.** Representative radial distance trajectory of a simulated system with and without an additional error on the distance that is in the range of the experimental error. **C.** Radial distance distribution for the proximal state of the +3xCTCF sites/+RAD21 condition overlaid with the distributions of the

proximal states from the three best matching parameters sets when comparing only the average radial distances. **D.** Heatmap showing the agreement of all simulated systems (for extrusion speed 0.1 kb/s) with the experimental data. The score is as described in Fig. 6A (see Methods). **E.** MSDs for three conditions for extruder residence time of 5.5 min, loading rate of $0.6 \text{ (Mb} \times \text{min)}^{-1}$ and extrusion speed of 10 kb/s. Pairwise comparison for conditions indicated in the title of each pair of heatmaps. **F.** Heatmap showing the fold change of generalized diffusion coefficients (**D**) and scaling exponent (α), and represents fold change between the conditions.



Extended Data Fig. 10 | Polymer simulations of landscapes with two barriers at different distances. **A.** Scheme of simulated polymers with varying distances between (optional) convergent loop extrusion barriers, corresponding to 100, 150, 250, 500, and 1000 kb. **B.** Duration (left) and rate of formation (right) of the HMM proximal state detected on simulated pairwise distances (after addition of experimental error) between monomers in the presence or absence of extrusion barriers, as a function of the intervening linear genomic distance. Lines are means, shaded areas are s.e.m. Note that the average duration of the

HMM proximal state slightly decreases although the average duration of the underlying cohesin-mediated CTCF-CTCF interaction doesn't (see panel C). This is due to non-CTCF mediated interactions, which also contribute to the proximal state, and decrease with increasing genomic distance. **C.** Average duration (left) and rate of formation (right) of the looped state (that is cohesin-mediated CTCF-CTCF interaction) extracted from polymer simulations. Lines are means, shaded areas are s.e.m.

Nonlinear control of transcription through enhancer–promoter interactions

<https://doi.org/10.1038/s41586-022-04570-y>

Received: 22 April 2021

Accepted: 21 February 2022

Published online: 13 April 2022

Open access

 Check for updates

Jessica Zuin^{1,5}, Gregory Roth^{1,5}, Yinxiu Zhan¹, Julie Cramard¹, Josef Redolfi¹, Ewa Piskadlo¹, Pia Mach^{1,2}, Mariya Kryzhanovska¹, Gergely Tihanyi^{1,2}, Hubertus Kohler¹, Mathias Eder³, Christ Leemans³, Bas van Steensel³, Peter Meister⁴, Sebastien Smallwood¹ & Luca Giorgetti¹✉

Chromosome structure in mammals is thought to regulate transcription by modulating three-dimensional interactions between enhancers and promoters, notably through CTCF-mediated loops and topologically associating domains (TADs)^{1–4}. However, how chromosome interactions are actually translated into transcriptional outputs remains unclear. Here, to address this question, we use an assay to position an enhancer at large numbers of densely spaced chromosomal locations relative to a fixed promoter, and measure promoter output and interactions within a genomic region with minimal regulatory and structural complexity. A quantitative analysis of hundreds of cell lines reveals that the transcriptional effect of an enhancer depends on its contact probabilities with the promoter through a nonlinear relationship. Mathematical modelling suggests that nonlinearity might arise from transient enhancer–promoter interactions being translated into slower promoter bursting dynamics in individual cells, therefore uncoupling the temporal dynamics of interactions from those of transcription. This uncovers a potential mechanism of how distal enhancers act from large genomic distances, and of how topologically associating domain boundaries block distal enhancers. Finally, we show that enhancer strength also determines absolute transcription levels as well as the sensitivity of a promoter to CTCF-mediated transcriptional insulation. Our measurements establish general principles for the context-dependent role of chromosome structure in long-range transcriptional regulation.

Transcriptional control in mammals critically depends on enhancers, which control tissue specificity and developmental timing of many genes⁵. Enhancers are often located hundreds of kilobases away from target promoters and are thought to control gene expression by interacting with the promoters in the three-dimensional space of the nucleus. Chromosome conformation capture (3C) methods⁶ revealed that enhancer–promoter interactions predominantly occur within sub-megabase domains known as topologically associating domains (TADs). These mainly arise from nested looping interactions between sites that are bound by the DNA-binding protein CTCF that act as barriers for the loop extrusion activity of cohesin⁷.

TAD boundaries and CTCF loops are thought to favour enhancer–promoter communication within specific genomic regions and disfavor it with respect to surrounding sequences^{1,3,4,8}. However, this view has recently been challenged by reports that disruption of TAD boundaries^{9,10} or depletion of CTCF and cohesin^{11,12} do not lead to systematic changes in gene expression, and that some regulatory sequences can act across TAD boundaries¹³. The manipulation of single CTCF sites has also been reported to result in variable effects on gene expression^{2,4,10,14–18}. The very notion that physical proximity is required for transcriptional regulation has been questioned by the observed lack

of correlation between transcription and proximity in single cells^{19,20}. Thus, it is highly debated whether there are indeed general principles that determine how physical interactions enable or prevent enhancer action²¹. Enhancer–promoter genomic distance might also contribute to transcriptional regulation^{22,23}, but it is unclear whether an enhancer acts uniformly within a TAD^{24,25}, or whether its effect depends on the genomic distance from a promoter^{23,26}.

Enhancer action depends on genomic distance

Addressing these questions requires a quantitative understanding of the relationship between transcription and enhancer–promoter interactions in conditions in which confounding effects by additional regulatory and structural interactions are minimized. Here we provide such a description using an experimental assay in which an enhancer is mobilized from an initial location and reinserted at large numbers of genomic positions with respect to a promoter. This enables the measurement of transcription levels as a function of the enhancer location and, therefore, of enhancer–promoter contact frequencies (Fig. 1a). Specifically, we generated mouse embryonic stem (mES) cells carrying a transgene in which a promoter drives the expression of enhanced green fluorescent protein (eGFP).

¹Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. ²University of Basel, Basel, Switzerland. ³Division of Gene Regulation and Oncode Institute, Netherlands Cancer Institute, Amsterdam, The Netherlands. ⁴University of Bern, Bern, Switzerland. ⁵These authors contributed equally: Jessica Zuin, Gregory Roth. ✉e-mail: luca.giorgetti@fmi.ch

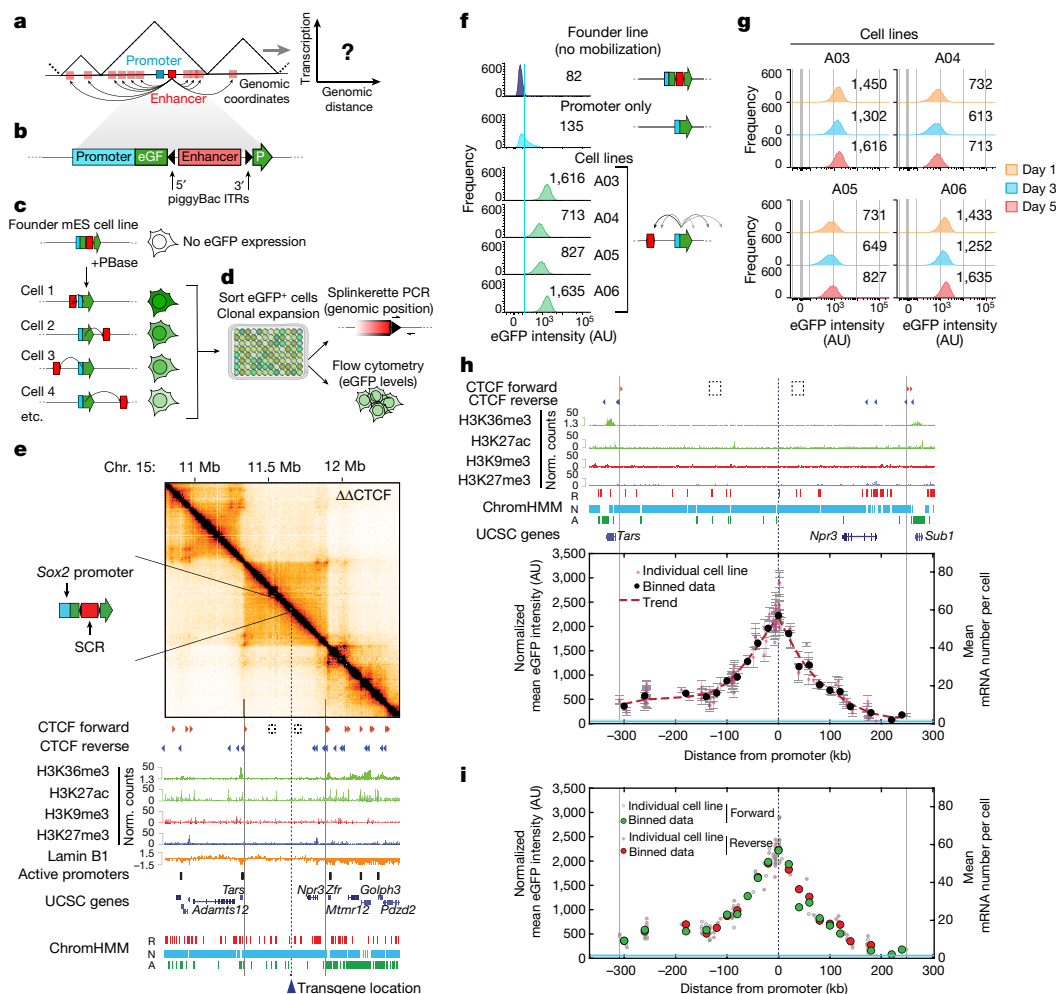


Fig. 1 | Enhancer action depends on the genomic distance from the promoter and is constrained by TAD boundaries. **a**, Mobilization of an enhancer around its target promoter to measure transcription as a function of their genomic distance. **b**, Schematic of the transgene: a promoter drives transcription of an *eGFP* gene split by a piggyBac-enhancer cassette. ITR, inverted terminal repeats. **c**, After expression of PBBase, the piggyBac-enhancer cassette is excised and randomly reinserted, occasionally leading to *eGFP* expression. **d**, Sorting of single *eGFP*⁺ cells results in cell lines in which the enhancer drives transcription from a single position. Splinkerette PCR and flow cytometry analysis are used to determine the enhancer position and promoter expression levels. **e**, Capture-C (6.4 kb resolution) analysis and genomic datasets in mES cells across 2.6 Mb centred around the selected TAD with both of the internal CTCF motifs deleted (dashed squares; $\Delta\Delta$ CTCF, double CTCF site deletion). The dashed line indicates the position of the future insertion of the transgene carrying the *Sox2* promoter and SCR. A, active; N, neutral; R,

repressive; Chr, chromosome. **f**, Representative flow cytometry profiles from founder mES cells, a promoter-only control cell line and *eGFP*⁺ cell lines with mobilized SCR. The light blue line indicates the mean *eGFP* levels in the promoter-only line. The numbers show the median *eGFP* intensities. AU, arbitrary units. **g**, *eGFP* levels in individual *eGFP*⁺ cell lines over cell passages. The numbers show the median *eGFP* values. **h**, Normalized mean *eGFP* intensities in individual *eGFP*⁺ cell lines as a function of SCR genomic position. The red dots are data from 135 individual cell lines; data are mean \pm s.d. $n = 3$ measurements on different days. The black dots show the average values within equally spaced 20 kb bins. The dashed red line shows the spline interpolation of average values. Mean mRNA numbers were inferred using smRNA-FISH calibration (Extended Data Fig. 1h). The light blue area shows the interval between the mean \pm s.d. of *eGFP* levels in three promoter-only cell lines. **i**, Data as in **h**, colour-coded according to SCR genomic orientation.

The *eGFP* transcript is split in two by a piggyBac transposon containing the cognate enhancer of the promoter (Fig. 1b). After expression of the PBBase transposase, the transposon is excised and reintegrated randomly into the genome, but preferentially in the vicinity of the initial site²⁷. Excision leads to reconstitution of functional *eGFP* of which the expression is used to isolate clonal cell lines by sorting single *eGFP*⁺ cells (Fig. 1c, d). This enables the rapid generation of hundreds of cell lines, each with the enhancer in a distinct genomic position. Enhancer position and *eGFP* expression are then determined in every cell line (Fig. 1d).

To minimize confounding effects, we integrated the transgene within a 560 kb TAD on chromosome 15 carrying minimal regulatory and structural complexity. This TAD does not contain expressed genes or active enhancers, is mostly composed of ‘neutral’ chromatin²⁸ except for a repressive -80 kb region at its 3’ side (Extended Data Fig. 1a), and

displays minimal structure mediated by two internal forward CTCF sites (Extended Data Fig. 1a, b). To further decrease the structural complexity, we deleted the two internal CTCF sites. This led to the loss of the associated loops (Extended Data Fig. 1c) and resulted in a simple homogeneous internal structure, as revealed by capture-C with tiled oligonucleotides spanning 2.9 Mb around the transgene (Fig. 1e and Extended Data Fig. 1c).

We first heterozygously inserted a single copy (Extended Data Fig. 1e) of a version of the transgene carrying the mouse *Sox2* promoter and the essential 4.8 kb region of its distal enhancer known as *Sox2* control region (SCR)^{29,30} (Extended Data Fig. 1d and Methods), from which we deleted its single CTCF site, which is not essential for transcriptional regulation at the endogenous locus¹⁷. Transgene insertion did not lead to substantial structural rearrangements within the TAD besides new moderate interactions with the CTCF sites at the 3’ and 5’ end of the

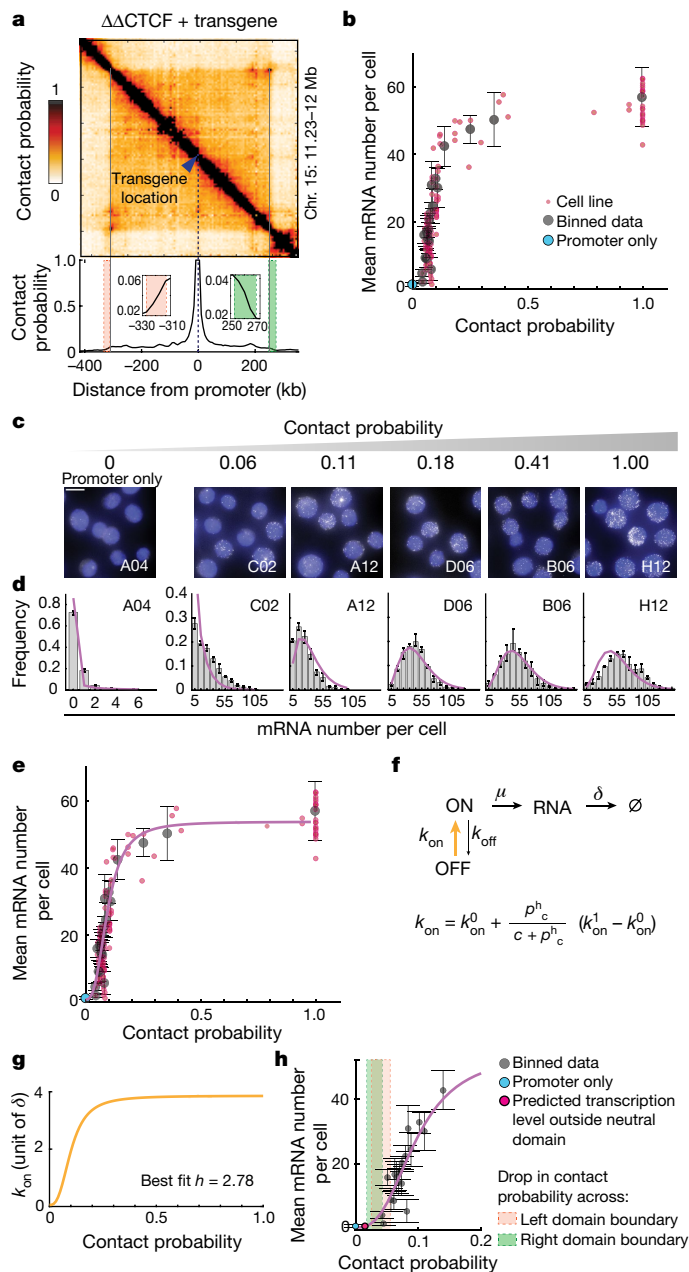


Fig. 2 | The promoter on rate is a sigmoidal function of enhancer–promoter contact probabilities. **a**, Capture-C (6.4 kb resolution) analysis of the founder cell line used for the experiments in Fig. 1 after converting read counts into contact probabilities (top) (Methods). Bottom, cross-section showing contact probabilities from the ectopic *Sox2* transgene. Insets: magnification of contact probability across the TAD boundaries. **b**, Mean *eGFP* mRNA numbers per cell plotted against contact probabilities between the ectopic *Sox2* promoter and SCR insertions. The red dots show individual cell lines. The black dots show the average values within equally spaced 20 kb bins \pm s.d. The number of cell lines per bin varies from 1 to 28. **c**, Representative smRNA-FISH images from cell lines in which *eGFP* transcription is driven by the *Sox2* promoter alone (left) or by the SCR located at different distances and contact probabilities (right). Scale bar, 10 μ m. **d**, Distributions of mRNA numbers per cell measured in the cell lines shown in **c**. The error bars show the minimum and maximum frequency. $n = 3$ technical replicates. The line shows the best fit of the phenomenological two-state model to the experimental data shown in **b** and **d**. **e**, Best fit to experimental data of **b** and **d**. Best-fit parameters are shown in Extended Data Fig. 3b. **f**, Description of the phenomenological two-state model with a variable on rate. The Hill function describes the dependency of k_{on} on contact probability (p_c). k_{on}^0 and k_{on}^1 are the minimum and maximum on rates, respectively; c and h are the Hill function critical threshold and the sensitivity parameter, respectively. \emptyset symbolizes degraded RNA. **g**, The best-fitting Hill function for k_{on} (in units of mRNA lifetime δ), corresponding to a sigmoidal curve. **h**, Close-up of **e**, highlighting the predicted insulation outside the TAD boundaries (red and green shaded areas). Data are presented as in **b**.

tenfold dynamic range in gene expression, from around 5 to 60 mRNAs per cell on average on the basis of smRNA-FISH calibration (Extended Data Fig. 1h). Insertions downstream of the non-transcribed *Npr3* gene generated lower transcription levels (Fig. 1h), possibly because this is a predominantly repressive region. Mild positive and negative deviations from the average decay in transcription levels indeed correlated with local enrichment in active and repressive chromatin states, respectively (Extended Data Fig. 1k). Consistent with the classical notion derived from reporter assays that enhancer activity is independent of genomic orientation³¹, enhancers inserted in forward or reverse orientations generated equivalent transcription levels (Fig. 1i). Interestingly, cell-to-cell heterogeneity in eGFP levels (assessed using coefficients of variation (CVs)) showed an opposite trend to mean expression levels and increased with increasing enhancer–promoter genomic distance (Extended Data Fig. 1l; examples of eGFP intensity distributions are provided in Extended Data Fig. 1m). Importantly, these results did not depend on the specific fluorescence gate used to define eGFP⁺ cells (Extended Data Fig. 1n, o). Together, these data show that the range of activity of the enhancer extends to the entire TAD and is delimited by its boundaries. However, transcription levels and their cell-to-cell variability quantitatively depend on enhancer–promoter genomic distance.

Enhancer contacts modulate burst frequency

We next examined the relationship between transcription levels and contact probabilities. Although reads from the wild-type allele might underemphasize changes introduced by the heterozygous insertion of the transgene, contact patterns detected in capture-C did not change substantially in individual cell lines in which the SCR was mobilized compared to the founder line before piggyBac mobilization (Extended Data Fig. 2a). Thus, the ectopic enhancer and promoter do not create prominent specific interactions, which enabled us to use capture-C data from the founder line (Methods)³² to infer contact probabilities between promoter and enhancer locations (Fig. 2a). Contact probabilities steeply decayed with increasing genomic distance from the promoter, fell considerably while approaching TAD boundaries (from 1 to around 0.05) and further dropped by a factor of around 3 across boundaries (Fig. 2a). This is consistent with previous estimations³³ confirmed using cross-linking and ligation-free methods³⁴ and is representative of the contact probabilities

TAD (Extended Data Fig. 1f). Mobilization of the piggyBac-SCR cassette led to random genomic reinsertions with a preference for chromosome 15 itself (Extended Data Fig. 1g). Individual experiments resulted in several tens of cell lines of which the eGFP levels were unimodally distributed (Fig. 1f), generally higher than those detected in control lines in which transcription was driven by the *Sox2* promoter alone (Fig. 1f), and remained stable over cell passages (Fig. 1g). Mean eGFP levels in single cell lines were linearly correlated with average numbers of *eGFP* mRNAs measured using single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) (Extended Data Fig. 1h). We therefore used flow cytometry as a readout of transcriptional activity.

Mapping of piggyBac-SCR positions in more than 300 cell lines revealed that, although in around 15% of them the transposon had not been successfully mobilized, in 99% of those in which it had (262 out of 264), the enhancer reinserted within the initial TAD (Fig. 1h and Extended Data Fig. 1i). In the two cell lines in which the enhancer transposed outside the TAD, eGFP levels were comparable to basal transcription driven by promoter-only control cell lines (Extended Data Fig. 1j). Notably, within the TAD, expression levels decreased with increasing enhancer–promoter genomic distance (Fig. 1h). Genomic distance accounted for a

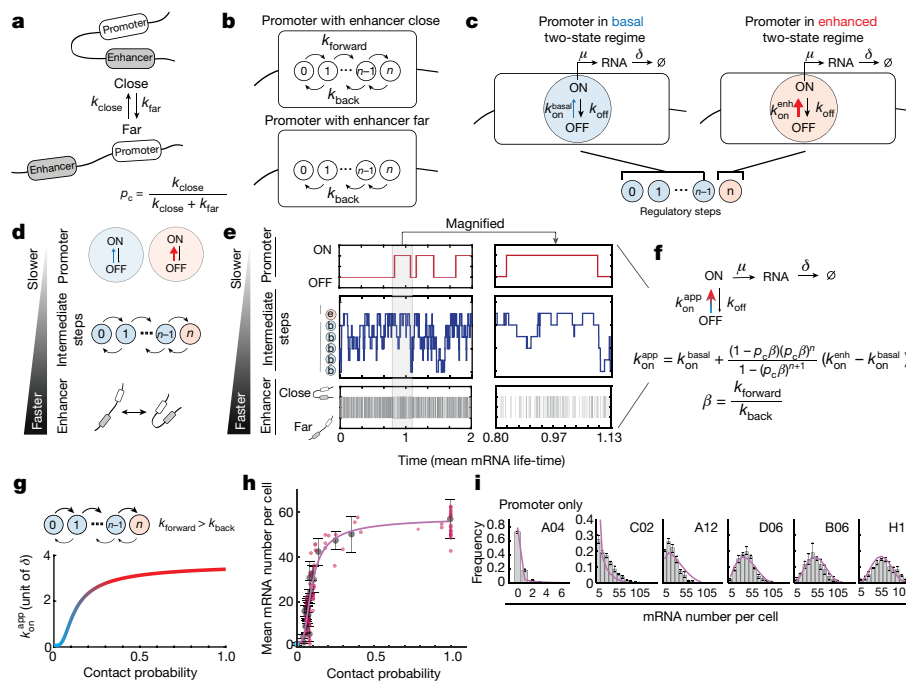


Fig. 3 | A mechanistic model of enhancer–promoter communication.

a, Stochastic promoter–enhancer interactions occur and disassemble with rates k_{close} and k_{far} . **b**, In the close state, the enhancer can trigger n reversible regulatory steps with forward and reverse rates k_{forward} and k_{back} . In the far state, regulatory steps can revert only at rate $k_{\text{back}}^{\text{basal}}$. **c**, The promoter operates in a basal two-state regime with a small on rate ($k_{\text{on}}^{\text{basal}}$) unless all n regulatory steps have been completed, in which case it transiently enters an enhanced two-state regime with a higher on rate ($k_{\text{on}}^{\text{enh}}$). **d**, Schematic of the parameter constraints under which the mechanistic model reduces to an apparent two-state model: $k_{\text{close,far}} \gg k_{\text{forward,back}} \gg k_{\text{on}}^{\text{basal,enh}}, k_{\text{off}}, \mu$. **e**, Representative single-cell dynamics of enhancer–promoter interactions, promoter regulatory steps and promoter

states predicted by the mechanistic model with $n = 5$ and rates satisfying the constraint on timescales described in **d** (time unit, $1/\delta$). **f**, Reduction of the mechanistic model to an apparent two-state model. The equation describes how the apparent on rate $k_{\text{on}}^{\text{app}}$ depends on contact probability (p_c) and other parameters of the mechanistic model. **g**, Dependency of $k_{\text{on}}^{\text{app}}$ on contact probability, illustrated for the best fitting parameters shown in **h** and **i**. **h**, Best fit of the apparent two-state model to the experimental transcriptional response shown in Fig. 2b. **i**, Best fit of the apparent two-state model to the experimental mRNA distributions shown in Fig. 2c. Best-fit parameters are shown in Extended Data Fig. 4c.

experienced by promoters in mES cells (Extended Data Fig. 2b, c). However, such a trend is at odds with our observation that transcription levels rather mildly decreased inside the TAD and dropped to promoter-only levels outside its boundaries (Fig. 1h and Extended Data Fig. 2d). Interestingly, plotting the mean *eGFP* mRNA numbers as a function of contact probabilities revealed a highly nonlinear relationship (Fig. 2b).

We sought to understand whether such a nonlinear relationship could be related to how enhancer–promoter interactions translate into transcription in individual cells. Transcription occurs in intermittent bursts³⁵ that give rise to variable mRNA numbers in single cells. smRNA-FISH analysis revealed substantial cell-to-cell variability in *eGFP* mRNA numbers in a panel of cell lines in which promoter–SCR contact probabilities ranged from zero (promoter-only control cell line) to one (Fig. 2c). Similar to *eGFP* protein distributions (Extended Data Fig. 2e), CVs of mRNA distributions increased with decreasing contact probabilities (Extended Data Fig. 2f). Bursty promoter behaviour can generally be described in terms of a two-state model of gene expression³⁶ in which the promoter stochastically switches with rates k_{on} and k_{off} between an OFF and an ON state in which transcription can initiate with rate μ . Consistent with this notion, mRNA number distributions (Fig. 2d) and mean transcription levels (Fig. 2e) in individual cell lines could be well approximated by a phenomenological two-state model in which the ‘on’ rate k_{on} (and therefore the burst frequency) nonlinearly depends on enhancer–promoter contact probability through a Hill function (Fig. 2f and Supplementary Information, model description). Interestingly, the best agreement with experimental data occurred with a Hill coefficient (h) of 2.8 (95% confidence interval = 2.4–3.2; Extended Data Fig. 3a, b). This corresponds to a sigmoidal transcriptional response in which the enhancer would be no

longer able to activate the promoter outside the approximately threefold drop in contact probabilities generated by TAD boundaries (Fig. 2g, h). Importantly the sigmoidal behaviour of k_{on} was not an artefact due to systematic errors in estimation of contact probabilities (Extended Data Fig. 3c), confounding effects of CTCF sites and repressive chromatin in the 3’ part of the TAD, or inclusion of promoter-only cell lines in the fit (Extended Data Fig. 3d). Alternative two-state models in which ‘off’ or initiation rates depend on contact probability rather than the on rate failed to reproduce the observed decrease in CV with contact probabilities (Supplementary Information, model description).

Mechanistic model of enhancer regulation

We next examined which mechanism could in principle generate such a phenomenological two-state model with sigmoidal modulation of k_{on} . Enhancer–promoter contacts are stochastic^{32,37,38} and probably dynamic³⁹ in single cells. Molecular processes that are thought to transmit regulatory information from enhancers to promoters (such as recruitment of transcription factors and coactivators, assembly of the Mediator complex⁴⁰), as well as those that are associated with promoter operation itself (such as pre-initiation complex assembly, RNA polymerase II pausing and release^{41,42}) are also stochastic and dynamic⁴³. We reasoned that the interplay between the timescales of these processes might generate nonlinear effects, as was recently hypothesized to explain promoter bursting⁴⁴. To investigate this concept in a quantitative manner, we developed a mechanistic model describing the simple hypothesis that, in single cells, the on rate of the promoter is transiently increased after stochastic interactions with an enhancer. We assumed that

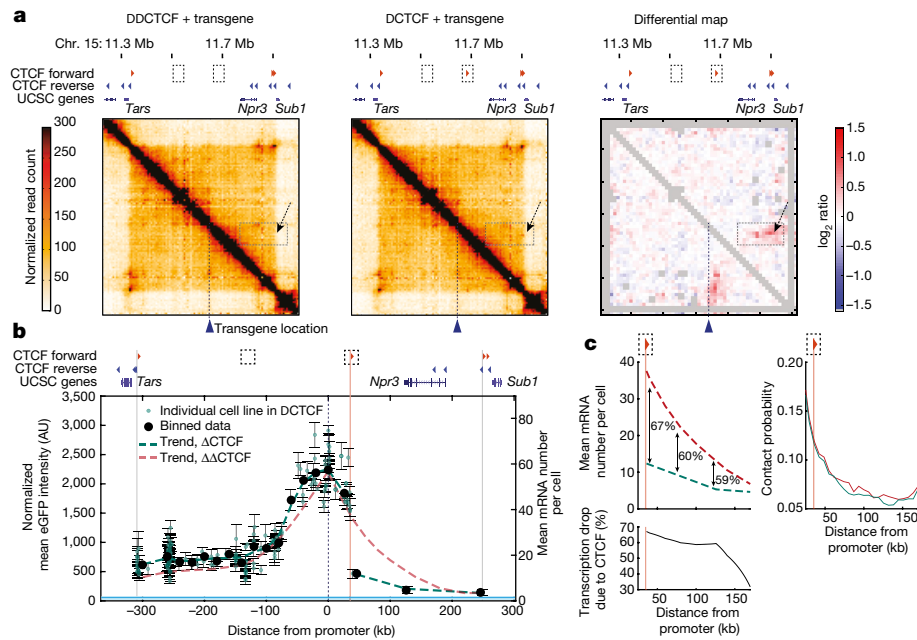


Fig. 4 | Insulation by a single CTCF site exceeds contact probability changes. **a**, Capture-C (6.4 kb resolution) analysis of founder mES cell lines in the absence ($\Delta\Delta$ CTCF) or presence (single CTCF-site deletion, Δ CTCF) of a forward CTCF motif 36 kb downstream of the transgene, and the corresponding differential map. The grey pixels show ‘noisy’ interactions that did not pass quality control filters (Methods). The dotted boxes and arrows indicate the position of the CTCF site and the structural changes it generates. **b**, The normalized mean \pm s.d. eGFP levels in 172 individual eGFP⁺ cell lines following SCR mobilization in Δ CTCF mES cells (green dots); $n = 3$ measurements performed on different days. The black dots show the mean \pm s.d. values within equally spaced 20 kb bins. The green dashed line

shows the spline interpolation of average eGFP values. The vertical pink line shows the position of the CTCF site at +36 kb. The red dashed line shows the trend of eGFP levels in the $\Delta\Delta$ CTCF background (compare with Fig. 1h). The blue line shows the promoter-only eGFP level as in Fig. 1h. **c**, Magnification of spline interpolants of GFP⁺ cell lines in the absence ($\Delta\Delta$ CTCF, red dashed line) or presence (Δ CTCF, green dashed line) of the CTCF binding site at +36 kb (vertical pink line) (left). The numbers represent the percentage fold changes between trendlines. Bottom, the percentage fold changes as a function of distance from the promoter. Right, contact probabilities from the location of the ectopic *Sox2* transgene in $\Delta\Delta$ CTCF (red line) and Δ CTCF (green line) mES cells.

enhancer–promoter interactions occur and disassemble with rates k_{close} and k_{far} , corresponding to a steady-state contact probability of $k_{\text{close}} / (k_{\text{close}} + k_{\text{far}})$ (Fig. 3a). When the enhancer is close to the promoter, it triggers one or more (n) reversible regulatory steps that transmit information to the promoter with forward and reverse rates k_{forward} and k_{back} (Fig. 3b). These steps are an abstract representation of any stochastic regulatory processes occurring at the enhancer–promoter interface. When the enhancer is far, no information is transmitted to the promoter and regulatory steps can only revert at rate k_{back} (Fig. 3b). The promoter operates in a basal two-state regime with a small on rate ($k_{\text{on}}^{\text{basal}}$) (Fig. 3c) unless all regulatory steps have been completed. In this case, the promoter transiently enters an ‘enhanced’ two-state regime with a higher on rate ($k_{\text{on}}^{\text{enh}}$), thus transiently increasing its transcriptional activity (Fig. 3c and Supplementary Information, model description). A transient increase in promoter activity therefore requires enhancer interactions that are either long enough (Extended Data Fig. 4a) or frequent enough (Extended Data Fig. 4b) to allow the completion of the n regulatory steps.

This mechanistic model does not generally reproduce the phenomenological two-state behaviour observed in Fig. 2e, f for the ectopic *Sox2* promoter. However, when the timescales of enhancer–promoter interactions are faster than those of intermediate regulatory steps, and both are faster than the promoter’s intrinsic bursting dynamics ($k_{\text{close, far}} \gg k_{\text{forward, back}} \gg k_{\text{on}}^{\text{basal, enh}}, k_{\text{off}}, \mu$) (Fig. 3d, e), the mechanistic model reduces to an apparent two-state model (Fig. 3f and Supplementary Information, model description). If forward transitions through $n > 1$ regulatory steps are favoured over backward reactions ($k_{\text{forward}} > k_{\text{back}}$), then the on rate of the apparent two-state model ($k_{\text{on}}^{\text{app}}$) depends sigmoidally on contact probabilities (Fig. 3g). This shows that, in principle, the promoter’s phenomenological two-state behaviour with sigmoidal modulation of k_{on} observed in Fig. 2e, f could arise from stochastic enhancer–promoter

interactions being transmitted into slower promoter ON/OFF dynamics through small numbers of intermediate regulatory processes. The resulting sigmoidal transcriptional response would enable an enhancer to act efficiently even when contact probabilities rapidly decay away from the promoter (Extended Data Fig. 2d), and contribute to block enhancer action when small drops in contact probabilities occur across TAD boundaries (Fig. 2h). The mechanistic model also predicts that enhancer–promoter contacts should not correlate with transcription bursts (Fig. 3e), as recently suggested by simultaneous imaging of *Sox2* transcription and genomic locations flanking the endogenous *Sox2* and *SCR*²⁰.

Finally, we verified that, when reduced to a two-state model, the mechanistic model could simultaneously fit the experimental transcriptional response to contact probabilities and smRNA-FISH distributions (Fig. 3h, i). Best agreement occurred with five intermediate regulatory steps (95% confidence interval = 3–7; Extended Data Fig. 4c, d and Supplementary Information, model description) and, consistent with previous observations²⁰, promoter ON/OFF transitions that occur in the timescale of several minutes (considering that the time unit in the model is mRNA lifetime, expected to be around 1.5 h)⁴⁵ (Extended Data Fig. 4c, d). Regulatory processes at the interface between enhancers and promoters have been estimated to occur in the order of tens of seconds^{41,43,46}, consistent with the condition that intermediate regulatory steps should be faster than bursting kinetics (Fig. 3f). The requirement that enhancer–promoter interactions should be even faster (Fig. 3f) therefore predicts that they should occur on a timescale of seconds or less.

Enhancer strength controls insulation levels

We next set out to examine whether CTCF binding affects the observed nonlinear relationship between transcription and contact probabilities.

Article

To this aim, we repeated the enhancer mobilization assay in mES cells in which only one of the two internal CTCF sites was homozygously deleted. The remaining forward CTCF site is located 36 kb downstream of the transgene and loops onto the reverse CTCF sites at the 3' end of the domain (Fig. 4a). SCR mobilization in this context resulted in 172 cell lines of which the transcription levels were indistinguishable from those generated in the 'empty' TAD, except across the CTCF site that severely, but not completely, insulated the ectopic *Sox2* promoter from the enhancer (Fig. 4b). Transcription levels across the CTCF site were about 60% lower than those generated in the absence of the CTCF site (Fig. 4c). Strikingly, this occurred in the absence of notable changes in the promoter's interaction probabilities with the region downstream of the CTCF site, at least in the current experimental set-up (capture-C data with 6.4 kb resolution) (Fig. 4c). This suggests that a single CTCF site might exert transcriptional insulation through additional mechanisms beyond simply driving physical insulation, possibly depending on site identity⁴⁷ and flanking sequences¹⁶.

The SCR is a strong enhancer that accounts for most of the transcriptional activity of endogenous *Sox2*^{29,30}. We reasoned that a weaker enhancer should lead to a different transcriptional response to contact probabilities with the promoter. There are two ways in which the parameters in the model shown in Fig. 3f might change when reducing enhancer strength. The ratio between transition rates through regulatory steps k_{forward} and k_{back} (β in Fig. 3h) might decrease, resulting in a slower transmission of regulatory information (Fig. 5a). This would generate a transcriptional response with maximal transcriptional levels that are similar to those generated by the SCR but different sensitivity to changes in contact probabilities (Fig. 5a). Alternatively (although not exclusively), the on rate in the enhanced promoter regime $k_{\text{on}}^{\text{enh}}$ could decrease (Fig. 5b). This would conserve the shape of the transcriptional response but decrease the maximal transcription level (Fig. 5b). To test these predictions, we performed the enhancer mobilization assay using a truncated version of the SCR (Extended Data Fig. 5a). This contained only one of the two -1.5 kb subregions that share similar transcription-factor-binding sites²⁹ and independently operate as weaker enhancers of the *Sox2* promoter in transient reporter assays²⁹ (Extended Data Fig. 5b). Mobilization of the truncated SCR in mES cells with a forward CTCF site downstream of the promoter (compare with Fig. 4a) led to 74 eGFP⁺ cell lines displaying approximately twofold lower transcription levels compared with those generated by the full-length SCR at comparable genomic distances (Fig. 5c). In contrast to the full-length SCR, the truncated enhancer was completely insulated from the promoter by the CTCF site (Fig. 5c). Thus, the level of functional insulation generated by the same CTCF site depends on the strength of the enhancer. In the region upstream of the CTCF sites, the transcriptional response generated by the truncated SCR (Fig. 5d) was in quantitative agreement with model predictions under the hypothesis that enhancer strength decreases the on rate rather than changing the intermediate regulatory steps (Fig. 5b), and could be predicted using the full-length SCR best-fit parameters with a two-fold decreased $k_{\text{on}}^{\text{enh}}$. This further strengthens our interpretation that enhancer strength modulates the ability of the promoter to turn on, possibly by regulating chromatin state, transcription factor binding or RNA polymerase II dynamics at the promoter^{35,44}.

In the nonlinear transcriptional response that we identified, high sensitivity in the low contact probability regime (that is, at long genomic distances) might contribute to secure insulation by TAD boundaries of even strong enhancers such as the SCR. Interestingly, in mES cells, the contact probabilities of most (~75%) active promoters with the nearest TAD boundary are comparable to those experienced by the ectopic *Sox2* promoter in our experiments (lower than 0.2) (Extended Data Fig. 5c). These promoters should therefore experience the same insulation mechanisms. The remaining promoters are closer (or adjacent) to a TAD boundary and therefore experience larger contact probabilities with the boundary, at which the transcriptional response is less sensitive (Extended Data Fig. 5d). However, interestingly, drops in contact

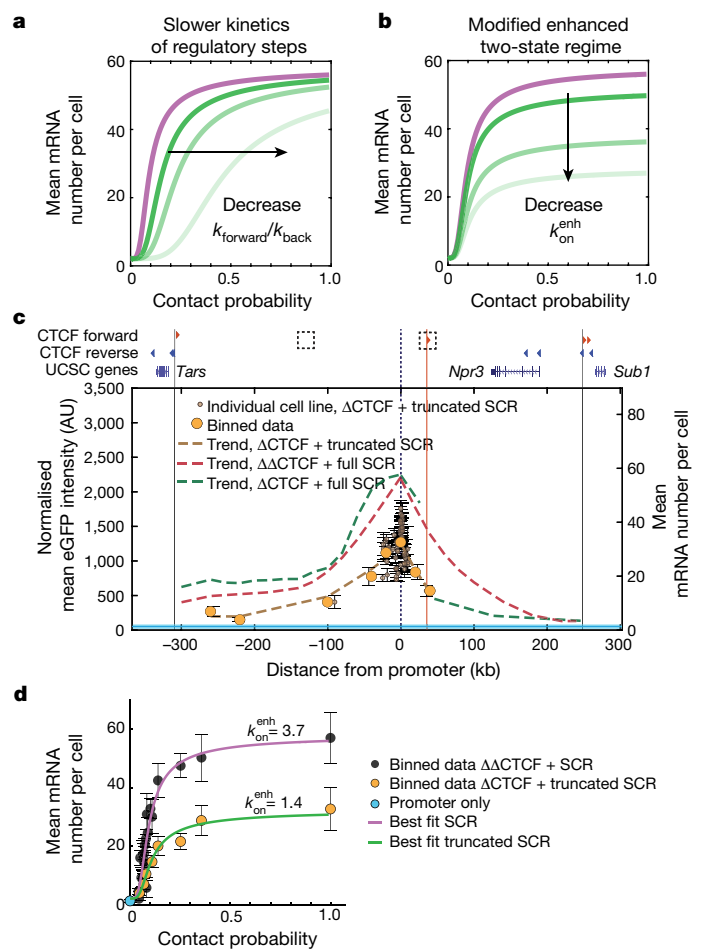


Fig. 5 | Enhancer strength modulates promoter on rates and determines insulation levels through a CTCF site. a, Model predictions under the hypothesis that decreasing enhancer strength results in a slower flow of regulatory information to the promoter. **b**, Model predictions as in **a**, under the alternative hypothesis that decreasing enhancer strength modifies the enhanced on rate ($k_{\text{on}}^{\text{enh}}$). **c**, Normalized eGFP levels in 74 individual GFP⁺ cell lines (brown dots; the error bars show the s.d. of $n = 3$ measurements performed on different days), binned data (orange dots) and data trend (brown dashed line) after mobilization of the truncated SCR in the $\Delta\Delta\text{CTCF}$ background. Trends of eGFP levels in individual GFP⁺ cell lines in which the SCR was mobilized either in the $\Delta\Delta\text{CTCF}$ background (red dashed line; Fig. 1h) or in the ΔCTCF background (green dashed line; Fig. 4b) are shown for comparison. Promoter-only eGFP levels (light blue) are shown as in Fig. 1h. **d**, The transcriptional response of the truncated SCR (green line) can be predicted from the best fit to the full-length SCR (purple line) with a modified enhanced on rate ($k_{\text{on}}^{\text{enh}}$). Data are mean \pm s.d. eGFP values were calculated within equally spaced 20 kb bins as in **c**; the number of cell lines per bin varies from 1 to 56.

probabilities across a boundary increase with decreasing genomic distance from the boundary itself (Extended Data Fig. 5d). This might contribute to the functional insulation of this class of promoters. Boundaries associated with clusters of CTCF sites might also benefit from the fact that insulation from CTCF sites can exceed the changes in contact probabilities that they generate (Fig. 4).

Discussion

Our study provides unbiased and systematic measurements of promoter output as a function of large numbers of enhancer positions with minimal confounding effects. The analysis of hundreds of cell lines enables us to move beyond locus-specific observations, and establishes a quantitative framework for understanding the role of chromosome

structure in long-range transcriptional regulation. Our data reveal that, within a TAD, absolute transcription levels generated by an enhancer depend on its genomic distance from the promoter and are determined by a nonlinear relationship with their contact probabilities. Minimal regulatory and structural complexities introduce deviations from this behaviour and might therefore confound its detection outside a highly controlled genomic environment, notably when studying regulatory sequences in their endogenous context²³. Mathematical modelling suggests that the observed nonlinear transcriptional response involves a modulation of the promoter's burst frequency, which could arise from transient enhancer–promoter interactions being translated into slower promoter bursting dynamics in individual cells. In addition to readily explaining the absence of correlation between transcription and physical proximity in single-cell experiments, this argues that the absence of such correlation should not be interpreted as the absence of causality. Although alternative explanations cannot be ruled out (such as cooperative effects through biomolecular condensates^{21,48}), our model provides a simple explanatory framework for both population-averaged and single-cell behaviour of enhancer-driven transcription, based on a minimal set of general and realistic hypotheses. Future live-cell imaging experiments with improved spatial and temporal resolution⁴⁹ will probably enable the testing of the model's prediction that enhancer–promoter interactions should occur on a timescale of seconds or less, therefore enabling the assessment of the model's premises. Finally, our study reveals that enhancer strength is not only a determinant of absolute transcription levels, but also of the level of insulation provided by CTCF. Our data therefore imply that transcriptional insulation is not an intrinsic absolute property of TAD boundaries or CTCF interactions but, rather, a graded variable depending on enhancer strength, boundary strength and distance from a promoter.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04570-y>.

- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Down, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
- Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, aad9024 (2016).
- Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
- 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).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
- Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
- Despang, A. et al. Functional dissection of the *Sox9-Kcnj2* locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
- Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
- Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
- Galupa, R. et al. A conserved noncoding locus regulates random monoallelic *Xist* expression across a topological boundary. *Mol. Cell* **77**, 352–367 (2020).
- Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
- Hanssen, L. L. P. et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.* **19**, 952–961 (2017).
- Huang, H. et al. CTCF mediates dosage and sequence-context-dependent transcriptional insulation through formation of local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).
- de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
- Rodríguez-Carballo, E. et al. Chromatin topology and the timing of enhancer function at the *HoxD* locus. *Proc. Natl Acad. Sci. USA* **117**, 31231–31241 (2020).
- Benabdallah, N. S. et al. Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484 (2019).
- Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
- Xiao, J. Y., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large changes in transcription. *eLife* **10**, e64320 (2021).
- Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
- Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
- Anderson, E., Devenney, P. S., Hill, R. E. & Lettice, L. A. Mapping the Shh long-range regulatory domain. *Development* **141**, 3934–3943 (2014).
- Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
- Symmons, O. et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
- Li, M. A. et al. The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. *Mol. Cell Biol.* **33**, 1317–1330 (2013).
- Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
- Zhou, H. Y. et al. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* **28**, 2699–2711 (2014).
- Li, Y. et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS ONE* **9**, e114485 (2014).
- Schaffner, W. Enhancers, enhancers—from their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).
- Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
- Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
- Redolfi, J. et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019).
- Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14**, 572–584 (2013).
- Peccoud, J. & Ycart, B. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).
- Amano, T. et al. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
- Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
- Chen, H. et al. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
- Vernimmen, D. & Bickmore, W. A. The hierarchy of transcriptional activation: from enhancer to promoter. *Trends Genet.* **31**, 696–708 (2015).
- Nguyen, V. Q. et al. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. *Mol. Cell* **81**, 3560–3575 (2021).
- Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422 (2017).
- Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
- Lammers, N. C., Kim, Y. J., Zhao, J. & Garcia, H. G. A matter of time: using dynamics and theory to uncover mechanisms of transcriptional bursting. *Curr. Opin. Cell Biol.* **67**, 147–157 (2020).
- Horvathova, I. et al. The dynamics of mRNA turnover revealed by single-molecule imaging in single cells. *Mol. Cell* **68**, 615–625 (2017).
- Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
- Anania, C. et al. In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.14.439779> (2021).
- Bialek, W., Gregor, T. & Tkacik, G. Action at a distance in transcriptional regulation. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1912.08579> (2019).
- Brandão, H. B., Gabriele, M. & Hansen, A. S. Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* **70**, 18–26 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Article

Methods

Culture of embryonic stem cells

All cell lines are based on E14 mES cells, provided by E. Heard's laboratory. Cells were cultured on gelatin-coated culture plates in Glasgow minimum essential medium (Sigma-Aldrich, G5154) supplemented with 15% fetal calf serum (Eurobio Abcys), 1% L-glutamine (Thermo Fisher Scientific, 25030024), 1% sodium pyruvate MEM (Thermo Fisher Scientific, 11360039), 1% MEM non-essential amino acids (Thermo Fisher Scientific, 11140035) 100 μM β -mercaptoethanol, 20 U ml^{-1} leukaemia inhibitory factor (Miltenyi Biotec, premium grade) in 8% CO_2 at 37 °C. Cells were tested for mycoplasma contamination once a month and no contamination was detected. After piggyBac-enhancer transposition, cells were cultured in standard E14 medium supplemented with 2i (1 μM MEK inhibitor PDO35901 (Axon, 1408) and 3 μM GSK3 inhibitor CHIR 99021 (Axon, 1386)).

Generation of enhancer–promoter piggyBac targeting vectors

Homology arms necessary for the knock-in, the *Sox2* promoter, the SCR and the truncated version of the SCR (Ei) were amplified from E14 mES cell genomic DNA by Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F549) using primers compatible with Gibson assembly cloning (NEB, E2611). The targeting vector was generated starting from the 3-SB-EF1-PBBAR-SB plasmid⁵⁰, gifted by Rob Mitra. To clone homology arms into the vector, BspEI and BclI restriction sites were introduced using Q5 Site-Directed Mutagenesis Kit (NEB, E0554). The left homology arm was cloned using Gibson assembly strategy by linearizing the vector with BspEI (NEB, R0540). The right homology arm was cloned using Gibson assembly strategy by linearizing the vector with BclI (NEB, R0160). The *Sox2* promoter was cloned by first removing the Efla promoter from the 3-SB-EF1-PBBAR-SB vector using NdeI (NEB, R0111) and Sall (NEB, R0138) and subsequently using Gibson assembly strategy. The SCR and its truncated version (truncated SCR or Ei) were cloned between the piggyBac transposon-specific inverted terminal repeat sequences (ITR) by linearizing the vector with BamHI (NEB, R3136) and NheI (NEB, R3131). A transcriptional pause sequence from the human alpha2 globin gene and an SV40 poly(A) sequence were inserted at both 5' and 3' ends of the enhancers using Gibson assembly strategy. A selection cassette carrying the puromycin resistance gene driven by the PGK promoter and flanked by FRT sites was cloned in front of the *Sox2* promoter by linearizing the piggyBac vector with the AsiSI (NEB, R0630) restriction enzyme. A list of the primers used for cloning is provided in Supplementary Table 1.

Generation of founder mES cell lines carrying the piggyBac transgene

The gRNA sequence for the knock-in of the piggyBac transgene on chromosome 15 was designed using the online tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) and purchased from Microsynth AG. gRNA sequence was cloned into the PX459 plasmid (Addgene) using the BsaI restriction site. E14 mES cell founder lines carrying the piggyBac transgene were generated using nucleofection with the Amaxa 4D-Nucleofector X-Unit and the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024KT). Cells (2×10^6) were collected with accutase (Sigma-Aldrich, A6964) and resuspended in 100 μl transfection solution (82 μl primary solution, 18 μl supplement, 1 μg piggyBac targeting vector carrying the SCR, truncated SCR or promoter alone, and 1 μg of PX459 ch15_gRNA/Cas9) and transferred into a single Nucleocuvette (Lonza). Nucleofection was performed using the protocol CG110. Transfected cells were directly seeded in prewarmed 37 °C culture in E14 standard medium. Then, 24 h after transfection, 1 $\mu\text{g ml}^{-1}$ of puromycin (InvivoGen, ant-pr-1) was added to the medium for 3 days to select cells transfected with PX459 gRNA/Cas9 vector. Cells were then cultured in standard E14 medium for an additional 4 days. To select cells with insertion of the piggyBac targeting

vector, a second pulse of puromycin was carried out by culturing cells in standard medium supplemented with 1 $\mu\text{g ml}^{-1}$ of puromycin. After 3 days of selection, single cells were isolated by fluorescence-activated cell sorting (FACS) on 96-well plates. Sorted cells were kept for 2 days in standard E14 medium supplemented with 100 $\mu\text{g ml}^{-1}$ primocin (InvivoGen, ant-pm-1) and 10 μM ROCK inhibitor (STEMCELL Technologies, Y-27632). Cells were then cultured in standard E14 medium with 1 $\mu\text{g ml}^{-1}$ of puromycin. Genomic DNA was extracted by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl, proteinase K and RNase) and subsequent isopropanol precipitation. Individual cell lines were analysed by genotyping PCR to determine heterozygous insertion of the piggyBac donor vector. Cell lines showing the corrected genotyping pattern were selected and expanded. A list of the primers used for genotyping is provided in Supplementary Table 1.

Puromycin resistance cassette removal

Cells (1×10^6) were transfected with 2 μg of a pCAG-FlpO-P2A-HygroR plasmid encoding for the flippase (Flp) recombinase using Lipofectamine 3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. Transfected cells were cultured in standard E14 medium for 7 days. Single cells were then isolated using FACS on 96-well plates. Genomic DNA was extracted by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl, proteinase K and RNase) and subsequent isopropanol precipitation. Individual cell lines were analysed by genotyping PCR to verify the deletion of the puromycin resistance cassette. A list of the primers used for genotyping is provided in Supplementary Table 1. Cell lines showing the correct genotyping pattern were selected and expanded. Selected cell lines were processed for targeted Nanopore sequencing with Cas9-guided adapter ligation (nCATS)⁵¹ and only the ones showing unique integration of the piggyBac donor vector were used as founder lines for the enhancer mobilization experiments.

Mobilization of the piggyBac-enhancer cassette

A mouse codon-optimized version of the piggyBac transposase (PBase) was cloned in frame with the red fluorescent protein tagRFpT (Evrogen) into a pBroad3 vector (pBroad3_hyPBase_IRES_tagRFpT) using Gibson assembly cloning (NEB, E2611). Cells (2×10^5) were transfected with 0.5 μg of pBroad3_hyPBase_IRES_tagRFpT using Lipofectamine 3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. To increase the probability of enhancer transposition, typically 12 independent PBase transfections were performed at the same time in 24-well plates. Transfection efficiency as well as expression levels of hyPBase_IRES_tagRFpT transposase within the cell population were monitored by flow cytometry analysis. Then, 7 days after transfection with PBase, individual eGFP⁺ cell lines were isolated using FACS in 96-well plates. Sorted cells were kept for 2 days in standard E14 medium supplemented with 100 $\mu\text{g ml}^{-1}$ primocin (InvivoGen, ant-pm-1) and 10 μM ROCK inhibitor (STEMCELL Technologies, Y-27632). Cells were cultured in E14 standard medium for additional 7 days and triplicated for genomic DNA extraction, flow cytometry analysis and freezing.

Sample preparation for mapping piggyBac-enhancer insertion sites in individual cell lines

Mapping of enhancer insertion sites in individual cell lines was performed using splinkerette PCR. The protocol was performed as described previously⁵² with a small number of modifications. Genomic DNA from individual eGFP⁺ cell lines was extracted from 96-well plates using the Quick-DNA Universal 96 Kit (Zymo Research, D4071) according to the manufacturer's instructions. Purified genomic DNA was digested by 0.5 μl of Bsp143I restriction enzyme (Thermo Fisher Scientific, FD0784) for 15 min at 37 °C followed by a heat-inactivation step at 65 °C for 20 min. Long (HMSpAa) and short (HMSpBb) splinkerette adapters were first resuspended with 5 \times NEBuffer 2 (NEB, B7002) to reach a concentration of 50 μM . Then, 50 μl of HMSpA adapter was

mixed with 50 μl of HMSpBb adapter (Aa+Bb) to reach a concentration of 25 μM . The adapter mix was denatured and annealed by heating it to 95 $^{\circ}\text{C}$ for 5 min and then cooling to room temperature. Then, 25 pmol of annealed splinkerette adapters was ligated to the digested genomic DNA using 5 U of T4 DNA ligase (Thermo Fisher Scientific, EL0011) and incubating the samples for 1 h at 22 $^{\circ}\text{C}$ followed by a heat-inactivation step at 65 $^{\circ}\text{C}$ for 10 min. For splinkerette amplifications, PCR 1 was performed combining 2 μl of the splinkerette sample, 1 U of Platinum Taq polymerase (Thermo Fisher Scientific, 10966034), 0.1 μM of HMSp1 and 0.1 μM of PB5-1 (or PB3-1) primer, and splinkerette PCR 2 was performed using 2 μl of PCR 1, 1 U of Platinum Taq polymerase (Thermo Fisher Scientific, 10966034), 0.1 μM of HMSp2 and 0.1 μM of PB5-5 (or PB3-2) primer. The quality of PCR amplification was checked by agarose gel electrophoresis. Samples were sent for Sanger Sequencing (Microsynth AG) using the PB5-2 (or PB3-2) primer. A list of the primers used for splinkerette PCRs and sequencing is provided in Supplementary Table 1. Mapping of enhancer insertion sites in individual cell lines was performed as described in the 'Mapping of piggyBac-enhancer insertion sites in individual cell lines' section.

Flow cytometry eGFP fluorescence intensity measurements and analysis

eGFP⁺ cell lines were cultured in serum + 2i medium for 2 weeks before flow cytometry measurements. eGFP levels of individual cell lines were measured on the BD LSR II SORP flow cytometer using BD High Throughput Sampler (HTS), which enabled sample acquisition in 96-well plate format. Measurements were repeated three times for each clone. Mean eGFP fluorescence intensities were calculated for each clone using FlowJo and all three replicates were averaged.

Normalization of mean eGFP fluorescence intensities

Mean eGFP fluorescence levels of each cell line measured in flow cytometry were first corrected by subtracting the mean eGFP fluorescence intensities measured in wild-type E14 mES cells cultured in the same 96-well plate. The resulting mean intensities were then normalized by dividing them by the average mean intensities of all cell lines where the SCR was located within a 40 kb window centred at the promoter location, and multiplied by a common factor.

Sample preparation for high-throughput sequencing of piggyBac-enhancer insertion sites

Cells (5×10^5) were transfected with 2 μg of PBase using Lipofectamine 3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions. Transfection efficiency as well as expression levels of PBase within the cell population were monitored by flow cytometry analysis. Then, 5 days after transfection with PBase, genomic DNA was purified using the DNeasy Blood & Tissue Kit (Qiagen, 69504). To reduce the contribution from cells in which excision of piggyBac-enhancer did not occur, we depleted eGFP sequences using an in vitro Cas9 digestion strategy. gRNA sequences for eGFP depletion were designed using the online tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) (Supplementary Table 1). Custom-designed Alt-R CRISPR-Cas9 crRNAs containing the gRNA sequences targeting eGFP (gRNA_1_3PRIME and gRNA_2_3PRIME), Alt-R CRISPR-Cas9 tracrRNA (IDT, 1072532) and Alt-R *Streptococcus pyogenes* Cas9 enzyme (IDT, 1081060) were purchased from IDT. In vitro cleavage of the eGFP fragment by Cas9 was performed according to the IDT protocol 'In vitro cleavage of target DNA with ribonucleoprotein complex'. In brief, 100 μM of Alt-R CRISPR-Cas9 crRNA and 100 μM of Alt-R CRISPR-Cas9 tracrRNA were assembled by heating the duplex at 95 $^{\circ}\text{C}$ for 5 min and allowing to cool to room temperature (15–25 $^{\circ}\text{C}$). To assemble the RNP complex, 10 μM of Alt-R guide RNA (crRNA:tracrRNA) and 10 μM of Alt-R SpCas9 enzyme were incubated at room temperature for 45 min. To perform in vitro digestion of eGFP, 300 ng of genomic DNA extracted from the pool cells transfected with the PBase was incubated for 2 h with

1 μM Cas9/RNP. After the digestion, 40 μg of proteinase K was added and the digested sample was further incubated at 56 $^{\circ}\text{C}$ for 10 min to release the DNA substrate from the Cas9 endonuclease. After purification using AMPURE beads XP (Beckman Coulter, A63881), genomic DNA was digested by 0.5 μl of BspI43I restriction enzyme (Thermo Fisher Scientific, FD0784) for 15 min at 37 $^{\circ}\text{C}$ followed by a heat-inactivation step at 65 $^{\circ}\text{C}$ for 20 min. Annealed splinkerette adapters (Aa+Bb; 125 pmol) were then ligated to the digested genomic DNA using 30 U of T4 DNA ligase HC (Thermo Fisher Scientific, EL0013), and the samples were incubated for 1 h at 22 $^{\circ}\text{C}$ followed by a heat-inactivation step at 65 $^{\circ}\text{C}$ for 10 min. For splinkerette amplifications, 96 independent PCR 1 reactions were performed combining 100 ng of the splinkerette sample, 1 U of Platinum Taq polymerase (Thermo Fisher Scientific, 10966034), 0.1 μM of HMSp1 and 0.1 μM of PB3-1 primer, and splinkerette PCR 2 was performed using 4 μl of PCR 1 product, 1 U of Platinum Taq polymerase (Thermo Fisher Scientific, 10966034), 0.1 μM of HMSp2 and 0.1 μM of PB3-2 primer. A list of the primers used for splinkerette PCRs is provided in Supplementary Table 1. Splinkerette amplicon products were processed using the NEB Ultra II kit according to the manufacturer's protocol, using 50 ng of input material. Mapping of genome-wide insertions was performed as described in the 'Mapping of piggyBac-enhancer insertion sites in population-based splinkerette PCR' section.

Sample preparation for tagmentation-based mapping of PiggyBac insertions

PiggyBac integrations in pools of cells were mapped using a Tn5-transposon-based ITR mapping technique based on ref.⁵³ with minor alterations. Cells (2×10^5) were transfected with 0.5 μg of PBase using Lipofectamine 3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's instructions in 24-well plates. Eight independent transfections were performed in parallel. Transfection efficiency as well as expression levels of PBase within the cell population were monitored by flow cytometry analysis. Then, 7 days after transfection with PBase, 6 cell pools of 10,000 cells from low GFP values (gates low 1 and low 2) and 6 cell pools of 337 cells of high GFP values (gate high) were sorted in a 24-well plate. Sorted cells were kept for 2 days in standard E14 medium supplemented with 100 $\mu\text{g ml}^{-1}$ primocin (InvivoGen, ant-pm-1) and 10 μM ROCK inhibitor (StemCell Technologies, Y-27632). Cells were cultured in E14 standard medium for either 1 passage (pools from gates low 1 and low 2) or 2 passages (pools from gate high) and genomic DNA from individual pools was extracted using the Quick-DNA Miniprep Plus Kit (Zymo Research, D4069) according to the manufacturer's instructions. The Tn5 transposon was produced as described in ref.⁵⁴. The tagmentation reaction was performed as follows. The primers TAC0101 & TAC0102 (45 μl of 100 μM) each were mixed with 10 μl 10 \times Tris-EDTA (pH 8) and annealed by heating to 95 $^{\circ}\text{C}$ followed by a slow ramp down (0.1 $^{\circ}\text{C s}^{-1}$) until 4 $^{\circ}\text{C}$. The transposome is obtained by combining the adapters (1 μl of 1:2 diluted adapters) and the Tn5 transposon (1.5 μl of 2.7 mg ml^{-1} stock) in 18.7 μl Tn5 dilution buffer (20 mM HEPES, 500 mM NaCl, 25% glycerol) and incubating the mix for 1 h at 37 $^{\circ}\text{C}$. The tagmentation was performed by mixing 100 ng of genomic DNA with 1 μl of assembled transposome, 4 μl 5 \times TAPS-PEG buffer (50 mM TAPS-NAOH, 25 mM MgCl_2 , 8% (v/v) PEG8000) in a final volume of 20 μl . The reaction was incubated at 55 $^{\circ}\text{C}$ for 10 min and quenched with 0.2% SDS afterwards. For the best mapping results, both sides of the PiggyBac transposon were processed to obtain 5' ITR- and 3' ITR-specific libraries. First, we enriched our target region by linear amplification PCR with 3' ITR-specific (TAC0006) and 5' ITR-specific (TAC0099) primers. The PCR mix was 3 μl of tagmented DNA, 1 μl of 1 μM enrichment primer, 2 μl dNTPs (10 mM), 4 μl 5 \times Phusion HF Buffer (NEB), 0.25 μl Phusion HS Flex polymerase (2 U μl^{-1} , NEB), in a final volume of 20 μl and amplified as follows: 30 s at 98 $^{\circ}\text{C}$; 45 cycles of 10 s at 98 $^{\circ}\text{C}$, 20 s at 62 $^{\circ}\text{C}$ and 30 s at 72 $^{\circ}\text{C}$; then 20 s at 72 $^{\circ}\text{C}$. PCR 1 of the library preparation was performed using TAC0161 (3' ITR) and TAC0110 (5' ITR) in combination with N5xx (Illumina, Nextera Index Kit).

Article

The PCR mix was 5 μ l of enrichment PCR, 1 μ l of 10 μ M primers, 2 μ l dNTPs (10 mM), 4 μ l 5 \times Phusion HF Buffer and 0.25 μ l Phusion HS Flex polymerase (NEB), in a final volume of 25 μ l and amplified as follows: 30 s at 98 $^{\circ}$ C; 3 cycles of 10 s at 98 $^{\circ}$ C, 20 s at 62 $^{\circ}$ C and 30 s at 72 $^{\circ}$ C; and 8 cycles of 10 s at 98 $^{\circ}$ C, 50 s at 72 $^{\circ}$ C. In PCR2 the N7xx (Illumina, Nextera Index Kit) adapters were added to the PiggyBac specific locations as follows. PCR was performed with TACO103 (both ITRs) and N7xx. The PCR mix was 2 μ l of PCR1, 1 μ l of 10 μ M primers, 2 μ l dNTPs (10 mM), 4 μ l 5 \times Phusion HF Buffer and 0.25 μ l Phusion polymerase (Thermo Fisher Scientific), in a final volume of 22 μ l and amplified as follows: 30 s at 98 $^{\circ}$ C; 10 cycles of 10 s at 98 $^{\circ}$ C, 20 s at 63 $^{\circ}$ C and 30 s at 72 $^{\circ}$ C. Then, 5 μ l of library was checked on a 1% agarose gel and different samples were pooled according to smear intensity. Finally, the library was purified by bead purification using CleanPCR (CleanNA) beads at a ratio 1:0.8 sample:beads. The final library was sequenced using the Illumina MiSeq (150 bp, paired-end) system. Mapping of genome-wide insertions was performed as described in the 'Mapping of piggyBac-enhancer insertion sites by tagmentation' section.

Deletion of genomic regions containing CTCF-binding sites

gRNA sequences for depletion of the genomic regions containing the CTCF-binding sites were designed using the online tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) and purchased from Microsynth AG (Supplementary Table 1). gRNA sequences were cloned into the PX459 plasmid (Addgene) using the BsaI restriction site. To remove the first forward CTCF-binding site (chromosome 15: 11520474–11520491), 3×10^5 cells were transfected with 0.5 μ g of PX459 CTCF_KO_gRNA3/Cas9 and 1 μ g of PX459 CTCF_KO_gRNA10/Cas9 plasmids using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) according to the manufacturer's instructions. To remove the second forward CTCF-binding sites (chromosome 15: 11683162–11683179), 1×10^6 cells were transfected with 1 μ g of PX459 gRNA2_CTCF_KO/Cas9 and 1 μ g of PX459 gRNA6_CTCF_KO/Cas9 plasmids using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) according to the manufacturer's instructions. Then, 24 h after transfection, 1 μ g ml⁻¹ of puromycin was added to the medium for 3 days. Cells were then cultured in standard E14 medium for an additional 4 days. To select cell lines with homozygous deletion, single cells were isolated by FACS on 96-well plate. Sorted cells were kept for 2 days in E14 standard medium supplemented with 100 μ g ml⁻¹ primocin (InvivoGen, ant-pm-1) and 10 μ M ROCK inhibitor (STEMCELL Technologies, Y-27632). Cells were then cultured in standard E14 medium. Genomic DNA was extracted by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl, proteinase K and RNase) and subsequent isopropanol precipitation. Individual cell lines were analysed by genotyping PCR to determine homozygous deletion of the genomic regions containing the CTCF-binding sites. Cell lines showing the corrected genotyping pattern were selected and expanded. A list of the primers used for genotyping is provided in Supplementary Table 1.

smRNA-FISH

Cells were collected with accutase (Sigma-Aldrich, A6964) and adsorbed on poly-L-lysine (Sigma-Aldrich, P8920) precoated coverslips. Cells were then fixed with 3% PFA (EMS, 15710) in PBS for 10 min at room temperature, washed with PBS and kept in 70% ethanol at -20 $^{\circ}$ C. After at least 24 h incubation in 70% ethanol, the coverslips were incubated for 10 min with freshly prepared wash buffer composed of 10% formamide (Millipore Sigma, S4117) in 2 \times SSC (Sigma-Aldrich, S6639). The coverslips were hybridized overnight (around 16 h) at 37 $^{\circ}$ C in freshly prepared hybridization buffer composed of 10% formamide, 10% dextran sulfate (Sigma-Aldrich, D6001) in 2 \times SSC and containing 125 nM of RNA-FISH probe sets against Sox2 labelled with Quasar 670 (Stellaris) and against eGFP labelled with Quasar 570 (Stellaris). After hybridization, the coverslips were

washed twice with wash buffer prewarmed to 37 $^{\circ}$ C for 30 min at 37 $^{\circ}$ C with shaking, followed by 5 min incubation with 500 ng ml⁻¹ DAPI solution (Sigma-Aldrich, D9564) in PBS (Sigma-Aldrich, D8537). The coverslips were then washed twice in PBS and mounted on slides with Prolong Gold medium (Invitrogen, P36934) and cured at room temperature for 24 h. The coverslips were then sealed and imaged within 24 h.

RNA-FISH image acquisition

Images were acquired on a Zeiss Axion Observer Z1 microscope equipped with 100 mW 561 nm and 100 mW 642 nm HR diode solid-state lasers, an Andor iXion 885 EMCCD camera, and an α Plan-Fluar \times 100/1.45 NA oil-immersion objective. Quasar 570 signal was collected with the DsRed ET filter set (AHF Analysentechnik, F46-005), Quasar 670 with Cy5 HC mFISH filter set (AHF Analysentechnik, F36-760) and DAPI with the Sp. Aqua HC-mFISH filter set (AHF Analysentechnik, F36-710). The typical exposure time for RNA-FISH probes was set to around 300–500 ms with 15–20 EM gain and 100% laser intensity. DAPI signal was typically imaged with an exposure time of 20 ms with EM gain 3 and 50% laser intensity. The pixel size of the images was 0.080 \times 0.080 μ m with a z-step of 0.25 μ m for around 55–70 z-planes.

Image processing and quantification of mRNA numbers

Raw images were processed in KNIME, python and Fiji to extract the numbers of RNAs per cell. The KNIME workflow described below is based on a previously published workflow⁵⁵. z-stacks were first projected to a maximal projection for each fluorescence channel. Individual cells were then segmented using the DAPI channel using Gaussian convolution ($\sigma = 3$), followed by filtering using global threshold with Otsu filter, watershed and connected component analysis for nuclei segmentation. Cytoplasmic areas were then estimated with seeded watershed. Cells with nuclei partially outside the frame of view were automatically excluded. Cells containing obvious artifacts, wrongly segmented or not fully captured in xyz dimensions were manually excluded from the final analysis. Spot detection is based on the Laplacian of Gaussian method implemented in TrackMate⁵⁶. For the channels containing RNA-FISH probes signal, RNAs spots were detected after background subtraction (rolling ball radius 20–25 pixels) by selecting spot size 0.2 μ m and threshold for spot detection based on visual inspection of multiple representative images. Spot detection is based on the Laplacian of Gaussian method from TrackMate. Subpixel localization of RNA spots was detected for RNA channels and a list of spots per cell for each experimental condition and replicate was generated. Spots in each channel were then aggregated by cell in python to extract the number of RNAs per cell.

Enhancer reporter assays

To generate vectors for the enhancer reporter assay, the Sox2 promoter, SCR and the truncated versions of the SCR (Ei and Eii) were amplified from E14 mES cell genomic DNA with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F549) using primers compatible with Gibson assembly strategy. The Sox2 promoter was cloned into the 3-SB-EF1-PBBAR-SB vector as described above. The SCR and the truncated versions Ei and Eii were cloned in front of the Sox2 promoter by linearizing the vector with AgeI (NEB, R3552) and subsequently using Gibson assembly cloning. A transcriptional pause sequence from the human α 2-globin gene and an SV40 poly(A) sequence was inserted at both the 5' and 3' ends of the enhancers. To test enhancers activity, 3×10^5 cells were co-transfected with 0.5 μ g of the different versions piggyBac vectors and 0.5 μ g of pBroad3_hyPBbase_IRES_tagRFPT using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) according to the manufacturer's instructions. As a control, only 0.5 μ g of the piggyBac vector carrying the Sox2 promoter was transfected. 24 h after transfection, cells were collected and analysed by flow cytometry.

Capture-C sample preparation

Cells (20×10^6) were cross-linked with 1% formaldehyde (EMS, 15710) for 10 min at room temperature and quenched with glycine (final concentration, 0.125 M). Cells were lysed in 1 M Tris-HCl pH 8.0, 5 M NaCl and 10% NP40 and complete protease inhibitor (Sigma-Aldrich, I1836170001) and enzymatically digested using 1,000 U of MboI (NEB, R0147). Digested chromatin was then ligated at 16 °C with 10,000 U of T4 DNA ligase (NEB, M0202) in ligase buffer supplemented with 10% Triton X-100 (Sigma-Aldrich, T8787) and 240 µg of BSA (NEB, B9000). Ligated samples were de-cross-linked with 400 µg proteinase K (Macherey Nagel, 740506) at 65 °C and phenol-chloroform purified. 3C library preparation and target enrichment using a custom-designed collection of 6,979 biotinylated RNA 'baits' targeting single MboI restriction fragments chromosome 15:10283500–13195800 (mm9) (Supplementary Table 2; Agilent Technologies; designed as in ref.⁵⁷) were performed according to the SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol. The only exceptions were the use of 9 µg of 3C input material (instead of 3 µg) and shearing of DNA using Covaris sonication with the following settings: duty factor: 10%; peak incident power: 175; cycles per burst: 200; treatment time: 480 s; bath temperature: 4 °C to 8 °C).

Targeted nCATS analysis

gRNA sequences targeting specific genomic regions of chromosome 15 external to the homology arms of the transgene were designed using the online tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) (Supplementary Table 1). Custom-designed Alt-R CRISPR-Cas9 crRNAs (5 crRNAs targeting the region upstream and 5 crRNAs targeting the region downstream the integrated transgene), Alt-R CRISPR-Cas9 tracrRNA (IDT, 1072532) and Alt-R SpCas9 enzyme (IDT, 1081060) were purchased from IDT. Sample preparation and Cas9 enrichment were performed according to a previously described protocol⁵¹ with a few modifications. Genomic DNA from mES cell founder lines was extracted using the Genra Puregene Cell Kit (Qiagen, 158745) according to the manufacturer's instructions. The quality of the high molecular mass DNA was checked using the TapeStation (Agilent) system. Typically, 5 µg of high molecular mass DNA was processed for incubation using shrimp alkaline phosphatase (rSAP; NEB, M0371) for 30 min at 37 °C followed by 5 min at 65 °C to dephosphorylate DNA-free ends. For Cas9 enrichment of the target region, all ten Alt-R CRISPR-Cas9 crRNAs were first pooled at an equimolar amount (100 µM) and subsequently incubated with 100 µM of Alt-R CRISPR-Cas9 tracrRNA at 95 °C for 5 min to assemble the Alt-R guide RNA duplex (crRNA:tracrRNA). To assemble the RNP complex, 4 pmol of Alt-R SpCas9 enzyme was incubated with 8 pmol Alt-R guide RNA (crRNA:tracrRNA) at room temperature for 20 min. In vitro digestion and A-tailing of the DNA were performed by adding 10 µl of the RNP complex, 10 mM of dATP (NEB, N0440) and 5 U of Taq Polymerase (NEB, M0267) and incubating the samples for 30 min at 37 °C followed by 5 min at 72 °C. Adapter ligation for Nanopore sequencing was performed using the Ligation Sequencing Kit (Nanopore, SQK-CAS109) according to the manufacturer's instructions. After purification with AMPure PB beads (Witec, 100-265-900), the samples were loaded into the MniON system, selecting the SQK-CAS109 protocol.

Nanopore sequencing analysis

To map Nanopore sequencing reads, we first built a custom genome consisting of the transgene sequence flanked by -10 kb mouse genomic sequence upstream and downstream of the target integration site. The custom genome can be found at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021/blob/main/Nanopore/cassette/cassette.fa). Reads were mapped to the custom genome using minimap2 (v.2.17-r941)

with the '-x map-ont' parameter. Nanopore sequencing analysis has been implemented using Snakemake workflow (v.3.13.3). Reads were visualized using IGV (v.2.9.4). The full workflow can be found at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021).

RNA-sequencing sample preparation and analysis

Mouse embryonic stem cells were collected with accutase (5 min, 37 °C) and counted. Cells (3×10^5) were lysed with 300 µl TRIzol reagent. RNA was extracted using the Direct-Zol RNA extraction kit from Zymo. Library preparation was performed after Illumina TruSeq Stranded mRNA-seq according to the manufacturer protocol. Reads were mapped to the *Mus musculus* genome (build mm9) using STAR⁵⁸, using the following options: --outSJfilterReads Unique --outFilterType BySJout --outFilterMultimapNmax 10 --alignSJoverhangMin 6 --alignSJDBoverhangMin 2 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --outSAMstrandField intronMotif --outFilterIntronMotifs RemoveNoncanonicalUnannotated --outSAMtype BAM SortedByCoordinate --seedSearchStartLmax 50 --twopassMode basic. Gene expression was quantified using qCount from QuasR package⁵⁹ using the 'TxDb.Mmusculus.UCSC.mm9.knownGene' database for gene annotation (Bioconductor package: Carlson M and Maintainer BP. TxDb.Mmusculus.UCSC.mm9.knownGene: Annotation package for TxDb object(s); R package v.3.2.2). Active promoters were defined as genes with $\log_2[\text{RPKM} + 0.1]$ higher than 1.5.

Capture-C analysis

Capture-C data were analysed using HiC-Pro⁶⁰ (v.2.11.4); the parameters can be found at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021). In brief, read pairs were mapped to the mouse genome (build mm9). Chimeric reads were recovered after recognition of the ligation site. Only unique valid pairs mapping to the target regions were used to build contact maps. Iterative correction⁶¹ was then applied to the binned data. The target regions can be found at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021). For SCR_ΔCTCF, SCR_ΔCTCF and the derived clonal lines, data from replicate one were used to make the quantification and plots throughout the manuscript.

Differential capture-C maps

To evaluate the structural perturbation induced by the insertion of the transgene and the mobilization of the enhancer (ectopic sequences), we accounted for differences in genomic distances due to the presence of the ectopic sequence. In the founder cell line (for example, SCR_ΔCTCF), insertion of the transgene modifies the genomic distance between loci upstream and downstream the insertion site. To account for these differences, we generated distance-normalized capture-C maps in which each entry corresponds to the interaction normalized to the corrected genomic distance between the interacting bins. Outliers (defined using the interquartile rule) or bins with no reported interactions from capture-C were treated as noise and filtered out. Singletons, defined as the top 0.1 percentile of Z-score, were also filtered out. The Z-score is defined as $(\text{obs} - \text{exp})/\text{stdev}$, where obs is the capture-C signal for a given interaction and exp and stdev are the genome-wide average and standard deviation, respectively, of capture-C signals at the genomic distance separating the two loci. We next calculated the ratios between distance normalized and noise-filtered capture-C maps. A bilinear smoothing with a window of 2 bins was applied to the ratio maps to evaluate the structural perturbation induced by the insertion of the ectopic sequence.

Chromatin state calling with ChromHMM

Chromatin states were called using ChromHMM²⁸ with four states. The list of histone modification datasets used is provided in Supplementary Table 3. States with enrichment in H3K9me3 and H3K27me3 were merged, therefore resulting in three chromatin states: active (enriched

Article

in H3K27ac, H3K36me3, H3K4me1 and H3K9ac), repressive (enriched in H3K9me3 and H3K27me3) and neutral (no enrichment).

Mapping of piggyBac-enhancer insertion sites in population-based splinkerette PCR

To identify true-positive enhancer re-insertion sites, we first filtered out reads containing eGFP fragments. We then retained only read pairs for which one side mapped to the ITR sequence and the other side mapped to the splinkerette adapter sequence. We mapped separately the ITR/splinkerette sides of the read pair to the mouse genome (build mm9) using BWA mem⁶² with the default parameters. Only integration sites that had more than 20 reads from both ITR and splinkerette sides were retained.

Mapping of piggyBac-enhancer insertion sites in individual cell lines

To map the enhancer position in individual cell lines, Sanger sequencing (Microsynth) without the adapter sequences were filtered out. The first 24 bp of each read after the adapter was then mapped to the mouse genome (mm9) using vmatchPattern (Biostrings v.2.58.0). The script used to map Sanger sequencing can be found at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021).

Mapping of piggyBac-enhancer insertion sites by tagmentation

Before aligning paired-end sequencing reads, reads were filtered using an adaptation of cutadapt⁶³, processing each read pair in multiple steps. Sequence patterns originating from Tn5 and each ITR were removed. The paired-end reads coming from both ITRs were treated the same. First, the presence of the unique part of the 5' ITR and 3' ITR sequence was detected at the start of the second read of the pair and, if present, this sequence was trimmed. Next, the sequence up to and including the TTAA site that was found on both the 5' ITR and 3' ITR was trimmed off. This sequence only partly contained the respective primers used for each ITR, and was used to filter reads that contained the sequence expected for a correct PCR product starting at the transposon. The sequence up to, but not including, the TTAA was removed. Next, all of the other sequence patterns coming from either Tn5 or the ITR were removed from the 5' end of the first read in the pair and the 3' end of both reads.

After filtering and trimming the reads, the reads were aligned to a reference genome with an in silico insertion of the split-GFP construct, but with a single TTAA motif instead of the PiggyBac transposon. This was done by aligning the homology arms found in the plasmid against mm10 reference genome. The complete sequence on the reference matching both arms was replaced by the plasmid sequence inserted.

Alignment was performed using Bowtie2 with the fragment length set to a minimum of 0 bp and maximum of 2,000 bp and the very-sensitive option was used. After reads were aligned to the genome, sambamba⁶⁴ was used to remove duplicates and samtools⁶⁵ was used to filter out read pairs that were not properly paired. We then designated, for each read pair, the position of the first 4 nucleotides of the second read as a putative insertion site. To calculate the fraction of reads originating from the non-mobilized position, the number of read pairs that overlapped the non-mobilized position (the TTAA replacing the PiggyBac of the in silico insert) was divided over the total number of reads originating from putative insertion sites supported by at least one read pair with a mapping quality higher than 2. Confident insertions were identified as those with at least one read for both 5' and 3' ITR.

Calibration of the mean number of mRNAs per cell with smRNA-FISH

A linear model was used to predict the average number of eGFP mRNAs on the basis of the mean eGFP intensity. The model was fitted on 7 data points corresponding to the average number of eGFP mRNAs obtained using single-molecule RNA fluorescence in situ and the mean eGFP

intensity obtained by flow cytometry (Extended Data Fig. 1h; $R^2 = 0.9749$, $P < 0.0001$, t -test).

Mathematical model and parameter fitting

The phenomenological two-state model (Fig. 2) and the apparent two-state model deduced from the mechanistic enhancer–promoter model (Fig. 3) were both fitted simultaneously to the mean eGFP levels measured in individual cell lines and to the distributions of RNA numbers measured by smRNA-FISH in six cell lines where the SCR was located at different distances from the promoter. The mean number of mRNAs was calculated analytically and the steady-state distribution of the number of mRNA per cell was approximated numerically (Supplementary Information, model description). The parameters for the phenomenological two-state model are the minimum on rate k_{on}^0 , the minimum on rate k_{on}^1 , the off rate k_{off} , the initiation rate μ and the constant c and Hill exponent h , which together control the nonlinear dependency of k_{on} on contact probability. The parameters for the apparent two-state model are the basal on rate k_{on}^{basal} , the enhanced on rate k_{on}^{enh} , the off rate k_{off} , the initiation rate μ , the ratio between the forward and backward rates of the regulatory steps β and the number of regulatory steps n . All of these parameters were considered to be free in the fitting procedure. The apparent two-state model was also fitted to the binned mean number of mRNA molecules inferred from the eGFP⁺ cell lines with the truncated version of the SCR (Fig. 4). In this case, three versions of the apparent two-state model were fitted to the data using log-transformed likelihood ratios. The parameter β (version 1) or k_{on}^{enh} (model 2) or both (model 3) were considered to be free parameters, whereas the other parameters were fixed to the best fit values obtained for the full-length SCR dataset. Using log-transformed likelihood ratios, the fit of the three versions was compared to the fit of the model for which all of the parameters were considered to be free. The mathematical description of the enhancer–promoter communication model, the derivation of the apparent two-state model, and the fitting procedures are explained in detail in the Supplementary Information (model description).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All capture-C, RNA-seq, Oxford Nanopore, tagmentation and population-based splinkerette PCR sequencing fastq files generated in this study have been uploaded to the Gene Expression Omnibus (GEO) under accession number GSE172257. The following public databases were used: BSgenome.Mmusculus.UCSC.mm9 (<https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Mmusculus.UCSC.mm9.html>), TxDb.Mmusculus.UCSC.mm9.knownGene (<https://bioconductor.org/packages/release/data/annotation/html/TxDb.Mmusculus.UCSC.mm9.knownGene.html>).

Code availability

Custom codes generated in this study are available at GitHub (https://github.com/zhanyinx/Zuin_Roth_2021 (cHiC, Nanopore, Insertion mapping); https://github.com/gregroth/Zuin_Roth_2021 (mathematical model); and https://github.com/vansteensellab/tagmap_hopping/tree/giorgetti (tagmentation-based mapping of PiggyBac insertions)).

50. Qi, Z. et al. An optimized, broadly applicable piggyBac transposon induction system. *Nucleic Acids Res.* **45**, e55 (2017).
51. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
52. Uren, A. G. et al. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat. Protoc.* **4**, 789–798 (2009).
53. Stern, D. L. Tagmentation-based mapping (TagMap) of mobile DNA genomic insertion sites. Preprint at *bioRxiv* <https://doi.org/10.1101/037762> (2017).

54. Schep, R. et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell* **81**, 2216–2230 (2021).
55. Voigt, F., Eglinger, J. & Chao, J. A. in *RNA Detection: Methods and Protocols* (ed. Gaspar, I.) 373–384 (Springer, 2018).
56. Tinevez, J.-Y. et al. TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
57. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).
60. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
61. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
62. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1303.3997> (2013).
63. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
64. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
65. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
66. 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 Res.* **27**, 479–490 (2017).

Acknowledgements We thank R. Mitra for sharing the piggyBac-splitGFP vector; A. Boettiger and J.Y. Xiao for discussions on modelling; M. Michalski and S. Andrews for capture-C primer

design; L. Gelman and J. Eglinger for help with microscopy and image analysis; and G. Natoli, D. Schübeler, H. Grosshans and E. Nora for discussions and comments on the manuscript. J.Z. was supported by a Marie Skłodowska-Curie grant (748091 ‘3DQuant’). Research in the Giorgetti laboratory is funded by the Novartis Foundation, the European Research Council (ERC) (759366 ‘BioMeTre’) and Marie Skłodowska-Curie Innovative Training Networks (813327 ‘ChromDesign’ and 813282 ‘PEP-NET’) under the European Union’s Horizon 2020 research and innovation program, and the Swiss National Science Foundation (310030_192642). Research in the Meister laboratory is supported by the Swiss National Science Foundation (IZCOZO_189884/31003A_176226 to P. Meister.). Research in the van Steensel laboratory is supported by ERC Advanced Grant 694466 ‘GoCADiSC’.

Author contributions L.G. and J.Z. conceived and designed the study. J.Z., J.C., E.P., M.K. and G.T. performed the experiments. G.R. wrote and analysed the mathematical model. H.K. provided assistance with flow cytometry. J.C., P. Meister. and S.S. contributed to setting up nCATS. S.S. also provided assistance with high-throughput sequence experiments. J.Z., J.C., J.R. and P. Mach. generated cell lines. M.E., C.L. and B.v.S. assisted with tagmentation-based mapping of insertions. G.R. and Y.Z. analysed the data, except for flow cytometry and single-clone insertion mapping (J.Z.) and smRNA-FISH (E.P.). L.G. wrote the paper with G.R., J.Z. and Y.Z., and input from all of the authors.

Competing interests The authors declare no competing interests.

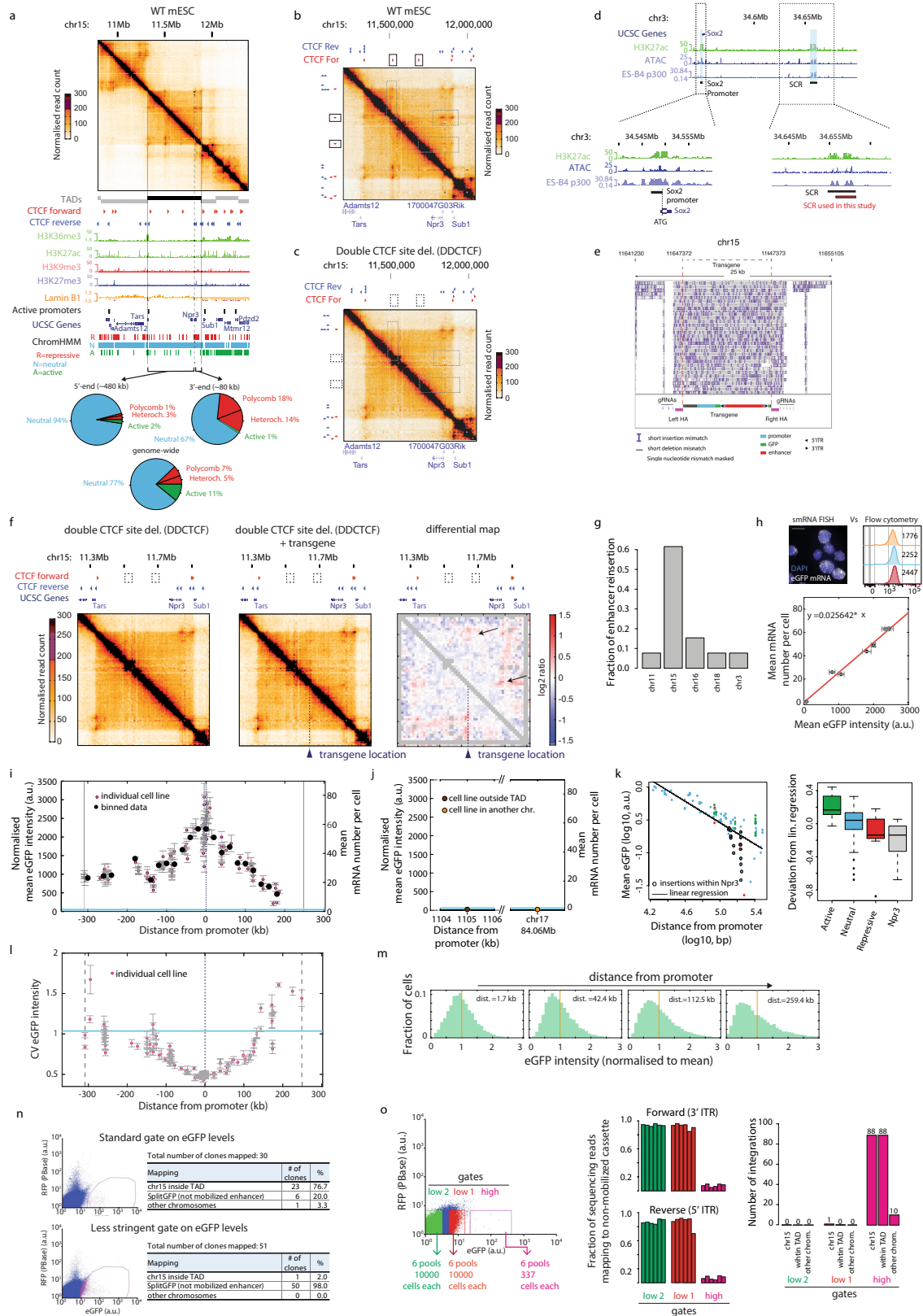
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-04570-y>.

Correspondence and requests for materials should be addressed to Luca Giorgetti.

Peer review information *Nature* thanks Mikhail Spivakov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



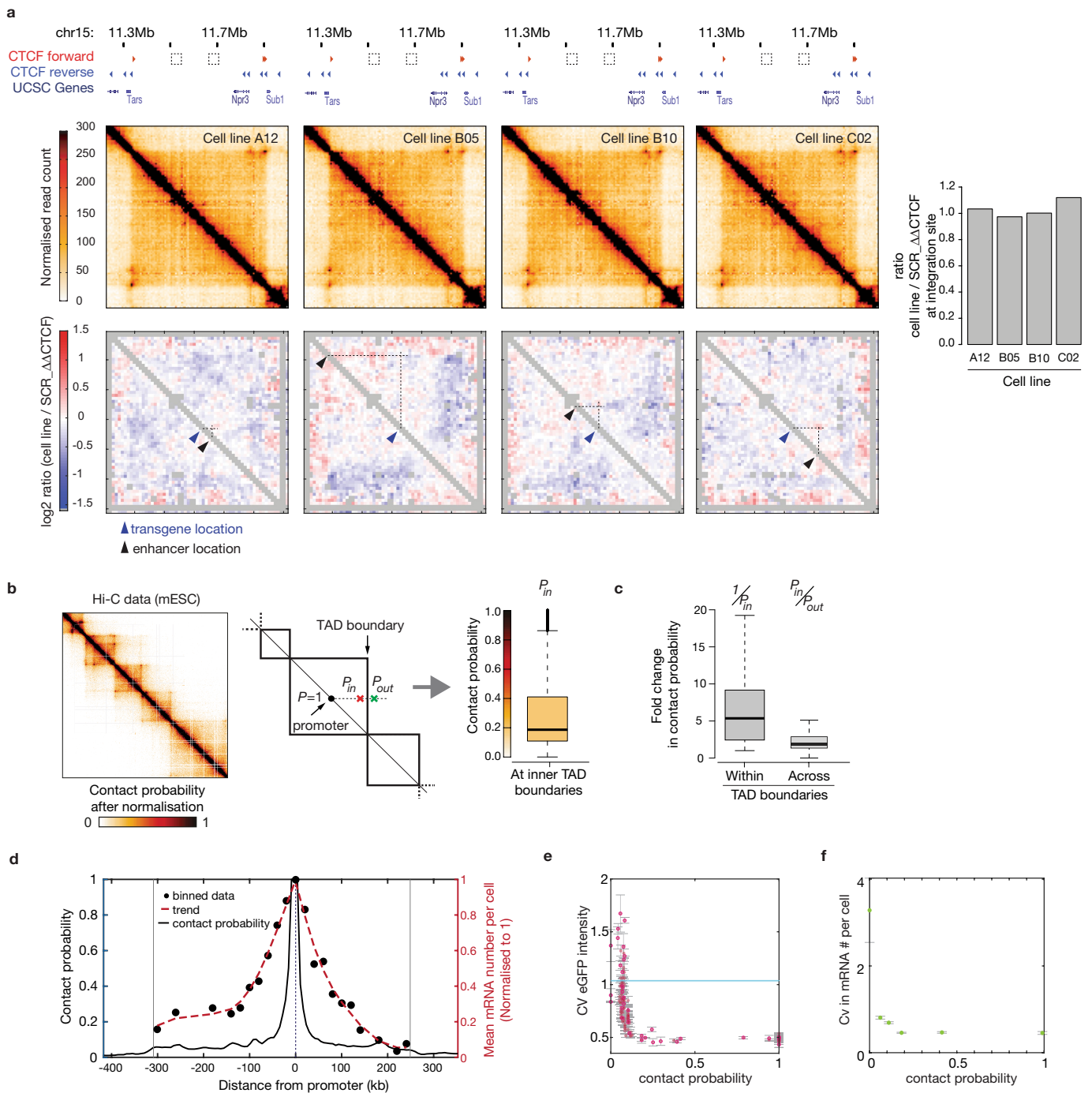
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Enhancer action is modulated by genomic distance from the target promoter and constrained by TAD boundaries.

a. Top: capture-C contact map at 6.4 kb resolution in wild-type (WT) mES cells in a 2.6 Mb region centred around the neutral TAD on chromosome 15 we used for the experiments. Vertical grey lines: TAD boundaries. Bottom: genomic datasets and ChromHMM analysis showing that the chosen TAD is devoid of active and repressive chromatin states, with the exception of 80 kb at the 3b at which is enriched in repressive chromatin states. b. Close-up view of panel a, highlighting the presence of CTCF-mediated chromatin loops (dotted boxes) in WT mES cells. c. capture-C contact map at 6.4 kb resolution for the same region as panel b in the cell line with double CTCF site deletions. CTCF deletions lead to loss of CTCF-mediated chromatin loops (dotted boxes). d. Top: UCSC snapshot of the endogenous *Sox2* locus and *Sox2* control region (SCR). Bottom: close-up views showing the regions of the *Sox2* promoter, the SCR region found in ref.²⁹ and the SCR used in the transgene construct. e. IGV snapshot showing nanopore sequencing reads mapped to a modified mouse genome including the transgene integration. Reads spanning from genomic DNA upstream the left homology arm to genomic DNA downstream the right homology arm confirmed single insertion of the transgene. f. capture-C maps at 6.4 kb resolution of the mES cell line with double CTCF sites deletion (left) and the founder mES cell line with transgene insertion (centre). Right: differential contact map. Grey pixels correspond to 'noisy' interactions that did not satisfy our quality control filters (see Methods). Transgene insertion induces new mild interactions with CTCF sites at the 3. and 5a extremities of the TAD (arrows). g. Barplot showing the fraction of piggyBac-SCR reinsertions genome-wide determined by Illumina sequencing of splinkerette PCR products from a pool of cells after PBase expression. See Methods for a detailed description of the protocol. h. Top: Representative smRNA-FISH image and flow cytometry profiles over different passages in a cell line where the SCR was mobilized in the immediate vicinity of the ectopic *Sox2* promoter. Scale bar, 10 μ m. Bottom: Linear relationship between the mean eGFP intensity and the average number of eGFP mRNAs measured using smRNA-FISH for seven single cell lines ($R^2 = 0.9749$, $p < 0.0001$, t-test). Error bars on the x-axis: standard deviation of three measurements performed on different days, as in Fig. 1h. Error bars on the y-axis: standard deviation of three technical replicates. i. Normalized mean eGFP intensities levels in individual eGFP+ cell lines are plotted as a function of the genomic position of the SCR in individual eGFP+ lines. Data from 127 individual cell lines (light red dots) from a single experiment are presented as mean \pm standard deviation (n=3 measurements performed in different days, as in Fig. 1g). Average eGFP values calculated within equally spaced 20 kb bins

(black dots) are shown. Mean mRNA numbers per cell were inferred from eGFP counts using calibration with smRNA-FISH, see Extended Data Fig. 1h. Shaded light blue area indicates the interval between mean \pm standard deviation of eGFP levels in three promoter-only cell lines. j. Same plot as Fig. 1h showing the only two SCR insertions we detected outside the TAD boundaries (brown dot) and on another chromosome (yellow dot). k. Left: Log10 average eGFP expression (from Fig. 1h) as a function of log10 absolute genomic distance between transgene position and SCR reinsertion. Points are colour-coded as in panel A (ChromHMM active, neutral, and repressive states). Black line denotes linear regression. Black circles denote SCR reinsertions within the *Npr3* gene body. Right: deviations of eGFP expression levels from the linear regression correlate with chromatin states called using ChromHMM (n: active = 16; neutral = 83; *Npr3* = 17; repressive = 7). Reinsertion of SCR within active or repressive regions respectively increases or decreases enhancer activity compared to neutral regions. Box plot: centre line denotes the median; boxes denote lower and upper quartiles (Q1 and Q3, respectively); whiskers denote 1.5x the interquartile region (IQR) below Q1 and above Q3; points denote outliers. l. Coefficients of variation (CV) of eGFP levels measured by flow cytometry plotted against SCR insertion locations in eGFP+ cell lines (light red dots). Data are presented as mean \pm standard deviation (n = 3 measurements in different days). Shaded light blue area indicates the interval between mean \pm standard deviation of eGFP level CVs in three promoter-only cell lines. m. Representative eGFP distributions (normalized to mean eGFP level) in clones with increasing absolute genomic distance (1.7 kb, 42.4 kb, 112.5 kb, and 259.43 kb) between the mobilized enhancer and the ectopic *Sox2* promoter. Vertical line indicates normalized mean eGFP levels. n. FACS plot showing standard (top) and less stringent (bottom) gates on eGFP levels used for single cells sort and insertion analysis of corresponding clonal cell lines. o. Left: FACS plot showing the gates used to sort pools of cells for tagmentation-based mapping of PiggyBac-enhancer insertions. For gates "low 1" and "low 2", six pools of 10000 cells were sorted while for gate "high", six pools of 337 cells were sorted. Gate "high" corresponds to the standard gate used to isolate eGFP positive cell lines for the mobilization experiments. Centre: Barplot showing the fraction of sequencing reads mapping to non-mobilized enhancer cassette determined by tagmentation-based mapping from the different pools sorted in gates "low 1", "low 2" and "high". See Methods for a detailed description of the protocol. Right: Numbers and genomic locations of confident insertion sites (identified as those with at least one read for both 5' and 3' mapping from the different pools sorted in gates "low 1", "low 2" and "high" eGFP gates.

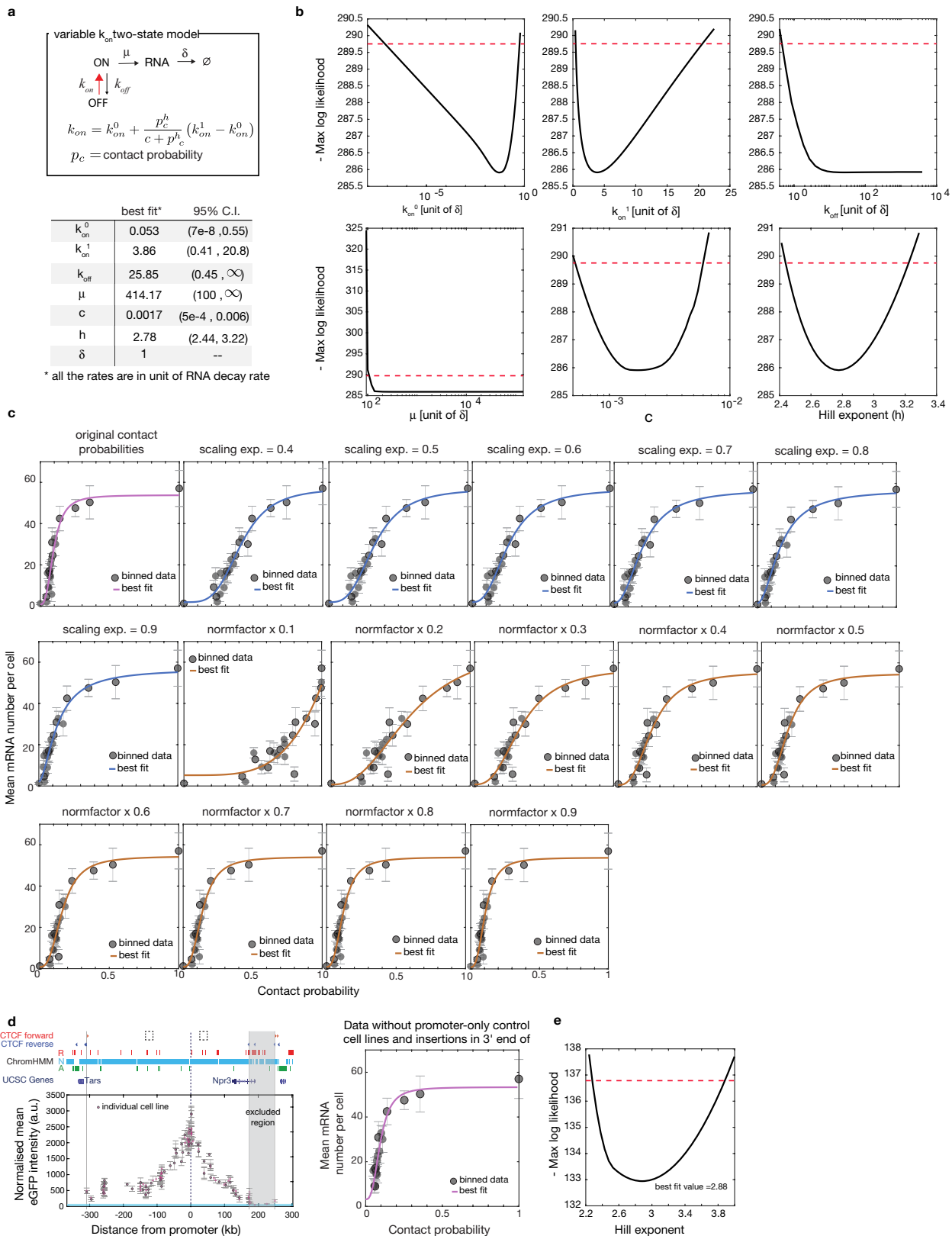
Article



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Analysis of chromosome structure around the transgenic locus and genome-wide in mES cells. a. Top: capture-C maps (6.4 kb resolution) of four cell lines where the SCR (black arrow) has been reinserted at different distances from the promoter (blue arrow). Bottom: differential contact map between individual cell lines and the founder line. Grey pixels: correspond to 'noisy' interactions that did not satisfy quality control filters (see Methods). Right: barplot showing the change in average interaction probabilities between the SCR reinsertion and the cassette, calculated using a square of 5 bins (6.4 kb resolution) centred at the cassette SCR reinsertion interaction. b. Left: example of Hi-C heatmap in mES cells at 6.4 kb resolution. Centre: scheme depicting how the probability of interaction between a promoter and the region immediately before the nearest TAD boundary (P_{in} , 12.8 kb i.e. two 6.4 kb bins before the boundary called using CaTCH^{6b}) and after the nearest TAD boundary (P_{out}) are calculated. Right: distribution of contact probability between all active promoters in mES cells and the closest inner TAD boundary (P_{in}) (n = 9655). Box plot description as in Extended Data Fig. 1k. c. Box plots showing the distribution of contact probability changes within the TAD and across the closest TADs boundary for

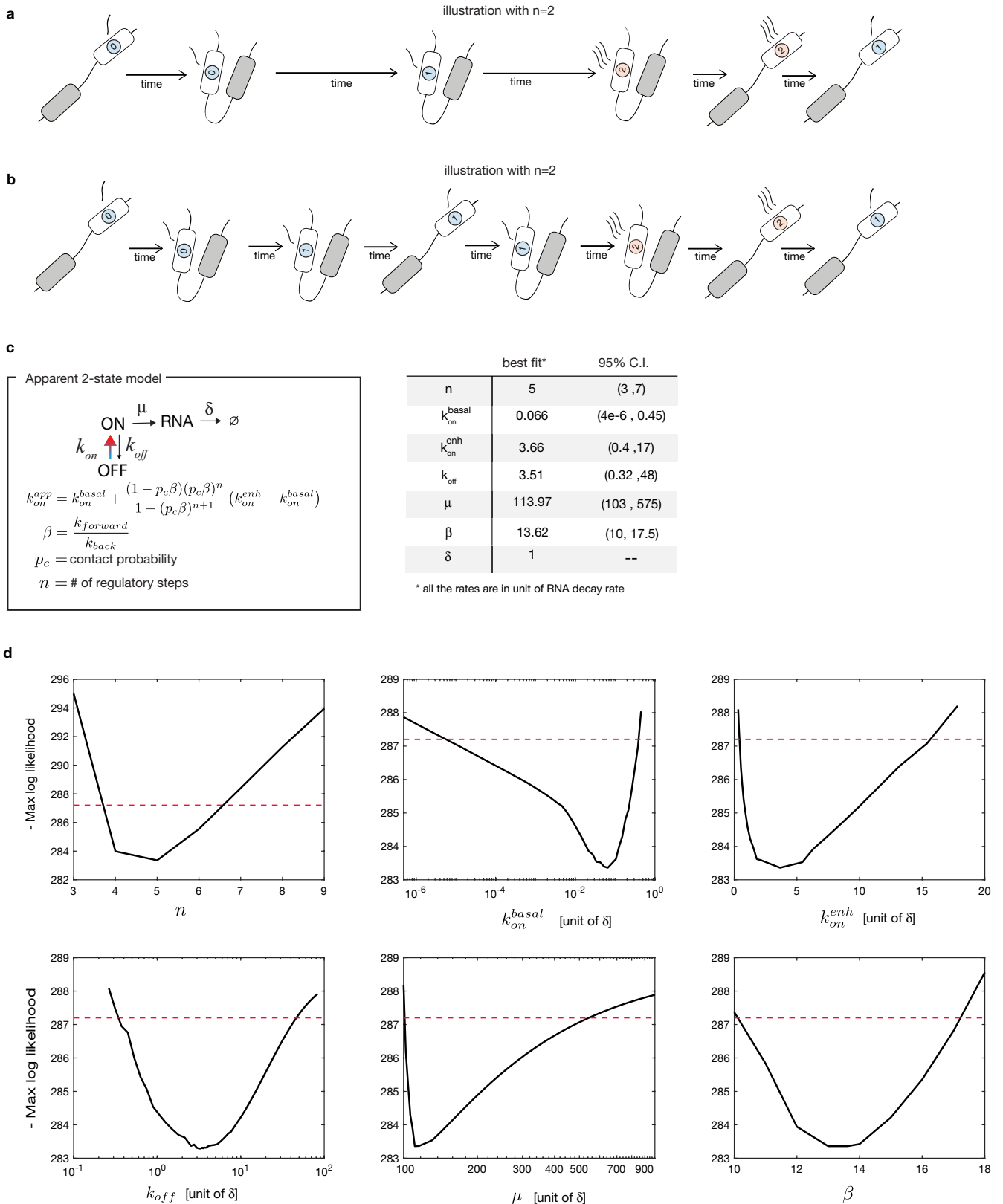
all active promoters in mES cells (n = 9655) whose contact probability outside the TAD is higher than 0.001 (n = 834). Box plot description as in Extended Data Fig. 1k; outliers not shown. d. Contact probabilities of the founder line from the location of the ectopic *Sox2* transgene (black line) and normalized averaged mean number of mRNAs per cell (highest value = 1) generated in individual eGFP+ lines by the SCR mobilization are plotted as a function of its genomic position (dashed red line). The average is calculated within equally spaced 20 kb bins as in Fig. 1h (black dots). e. Coefficients of variation (CV) of eGFP levels measured by flow cytometry plotted against contact probabilities between the ectopic *Sox2* promoter and the locations of SCR insertions. Data are presented as mean values +/- standard deviation (n = 3 measurements in different days). Shaded light blue area indicates the interval between mean +/- standard deviation of eGFP level CVs in three promoter-only cell lines. f. Coefficients of variation (CV) of mRNA number per cell measured by smRNA-FISH plotted against contact probabilities between the ectopic *Sox2* promoter and the locations of SCR in the cell the lines shown in Fig. 2c, d. Data are presented as mean values +/- standard deviation (n = 3 technical replicates).



Extended Data Fig. 3 | See next page for caption.

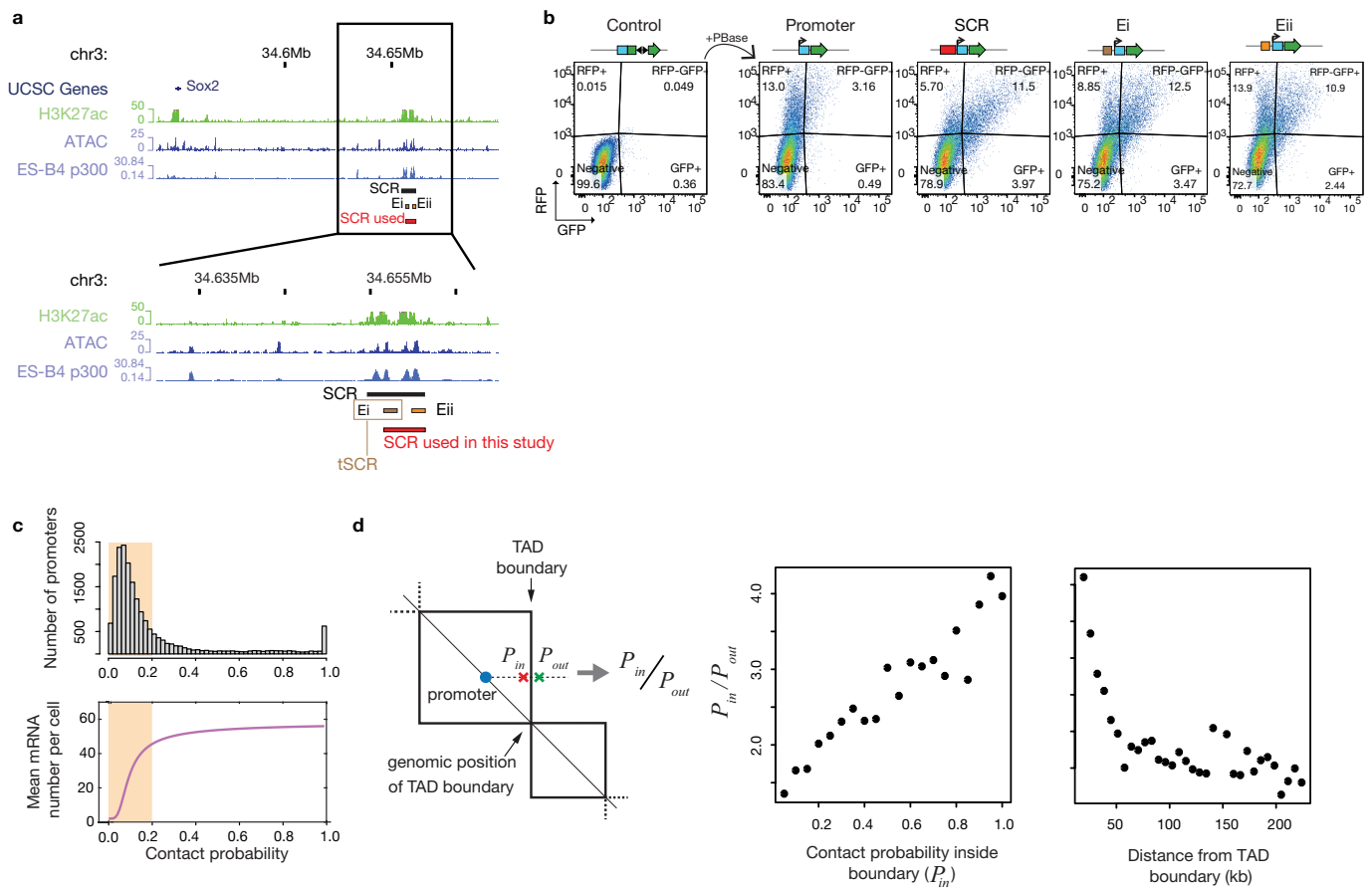
Extended Data Fig. 3 | Phenomenological two-state model fitting and robustness analysis. a. Parameter values and 95% confidence intervals for the best fitting phenomenological two-state model. The rates are in the unit of RNA decay rate (δ). b. Profile likelihood functions for all the parameters of the phenomenological two-state model. The red dashed line shows the threshold used to calculate the 95% confidence intervals (see Supplementary Model description for more details). c. Best fit of the phenomenological two-state model under different perturbations of the contact probabilities. Panels with blue curves show the best fit transcriptional responses when the scaling exponent of the contact probabilities was artificially set to 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. The scaling exponent of the original contact probabilities is 0.77. Panels with orange curves show the best fit transcriptional responses when contact probabilities were artificially increased by a factor $1/x$ with $x = 0.1, \dots, 0.9$ with step of 0.1. Data are presented as average eGFP values

calculated within equally spaced 20 kb bins \pm standard deviation ($n =$ number of cell lines per bin), as in Fig. 1h. d. Left: Normalized mean eGFP intensities in individual eGFP+ cell lines are plotted as a function of the genomic position of the SCR. Data from 135 individual cell lines (light red dots) are presented as mean \pm standard deviation ($n = 3$ measurements performed on different days, as in panel g). Shaded grey area indicates the genomic regions that were excluded from the fit shown in the right panel. Right: Best fit of the phenomenological two-state model in the absence of the promoter-only control cell line and the cell lines with insertions that landed beyond the first CTCF site at the 3' of the TAD (region highlighted in the left panel). Data are presented as average eGFP values calculated within equally spaced 20 kb bins \pm standard deviation ($n =$ number of cell lines per bin). e. Profile likelihood function for the Hill coefficient for the fit described in panel d.



Extended Data Fig. 4 | Fit of the mechanistic enhancer–promoter model and robustness analysis. a. Schematic description of the dynamics of the mechanistic model (here with two regulatory steps ($n=2$) for illustration). This case illustrates a scenario where, the enhancer–promoter interaction is long enough to allow the completion of the 2 regulatory steps and transiently increases the promoter activity. b. In an alternative scenario, the interactions are shorter but frequent enough to allow the completion of the 2 regulatory

steps and transiently increase the promoter activity. c. Parameter values and 95% confidence intervals for the best fitting apparent two-state model. The rates are in the unit of RNA decay rate (δ). d. Profile likelihood functions for all the parameters of the apparent two-state model. Red dashed lines show the threshold used to calculate the 95% confidence intervals (see Supplementary Model description for more details).



Extended Data Fig. 5 | Dependence of transcription levels and insulation on enhancer strength. a. Top: UCSC genome browser snapshot of the endogenous *Sox2* locus and *Sox2* control region (SCR). Bottom: close-up view showing the SCR (black) identified in ref.²⁹ and the enhancer regions used in the transient reporter assays shown in panel b. Full-length enhancer is in red (same as in Fig. 1); truncated versions are in brown (Ei) and orange (Eii). Experiments in Fig. 5 were performed with Ei. b. Flow cytometry analysis of mES cells transiently transfected with PBBase-RFP and different versions of split eGFP plasmids carry either no enhancer, or the full-length SCR (red, see panel a), or the first (brown-Ei) or second (orange-Eii) SCR subregions in front of the *Sox2* promoter. Transcription levels generated upon co-transfection with PBBase are higher in the presence of the full-length SCR compared to truncated versions. Numbers in each quadrant represent the % of cells either negative or RFP, GFP and RFP-GFP

positive. c. Top: distribution of contact probabilities between all active promoters in mES cells and the nearest inner TAD boundaries, calculated as in Extended Data Fig. 2b. Bottom panel: Model prediction for the mean eGFP mRNA numbers per cell plotted against contact probabilities shown as a comparison (same as Fig. 2e). Shaded areas correspond to promoters with contact probability with the closest TAD boundary below 0.2. d. Left panel: scheme of how the probabilities of interaction between promoter and the region before (P_{in}) and after the TAD boundary (P_{out}) are calculated, same criteria as in Extended Data Fig. 2b. Central panel: promoters with higher contact probabilities with TAD boundaries experience stronger drops of contact probability across boundaries. Right panel: promoters closer to TAD boundaries experience a stronger drop of contact probability across boundaries.

Pia Mach

Oetlingerstrasse 5, 4057 Basel, Switzerland
Email: pia.mach@fmi.ch; Phone: +41 77 531 9023

Education

- 2023 **Ph. D. in Molecular Biology**
Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
Thesis project: "The role of loop extrusion in the dynamics of chromosome folding"
Thesis supervisor: Dr. Luca Giorgetti
- 2018 **M. Sc. in Biochemistry**
Interfakultäres Institut für Biochemie, Eberhard Karls University, Tübingen, Germany
Thesis title: "Identification and characterisation of a novel arginine methylation (H4R78) in human cell lines"
Thesis supervisor: Prof. Tony Kouzarides, Gurdon Institute, University of Cambridge, Cambridge, United Kingdom
- 2015 **B. Sc. in Biochemistry**
Interfakultäres Institut für Biochemie, Eberhard Karls University, Tübingen, Germany
Thesis title: "Optimisation of the encodable editing system"
Thesis supervisor: Prof. Thorsten Stafforst

Publications

- 2023 **Mach, P.** and Giorgetti, L. Integrative approaches to study enhancer-promoter communication. *Current Opinion in Genetics and Development* 80, p. 102052, doi.org/10.1016/j.gde.2023.102052 (2023).
- 2022 **Mach, P.***, Kos, P. I.*, Zhan, Y.*, Cramard, J., Gaudin, S., Tünnermann, J., Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., Giorgetti, L. Cohesin and CTCF control the dynamics of chromosome folding. *Nature Genetics* 54, 1907-1918, doi:10.1038/s41588-022-01232-7 (2022). *equal contribution
- Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., **Mach, P.**, Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., Giorgetti, L. Nonlinear control of transcription through enhancer-promoter interactions. *Nature* 604, 571-577, doi:10.1038/s41586-022-04570-y (2022).
- Ng, T. L., Olson, E. J., Yoo, T. Y., Weiss, H. S., Koide, Y., Koch, P. D., Rollins, N. J., **Mach, P.**, Meisinger, T., Bricken, T., Chang, T. Z., Molloy, C., Zürcher, J., Chang, R. L., Mitchison, T. J., Glass, J. I., Marks, D. S., Way, J. C., Silver, P. A. High-Content Screening and Computational Prediction Reveal Viral Genes That Suppress the Innate Immune Response *mSystems* 7, e0146621, doi:10.1128/msystems.01466-21 (2022).

- 2019 Pandolfini, L., Barbieri, I., Bannister, A. J., Hendrick, A., Andrews, B., Webster, N., Murat, P., **Mach, P.**, Brandi, R., Robson, S. C., Migliori, V., Alendar, A., d’Onofrio, M., Balasubramanian, S., Kouzarides, T. METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Molecular Cell* 74, 1278-1290, doi:10.1016/j.molcel.2019.03.040 (2019).
- 2017 Schmohl, L., Bierlmeier, J., von Kügelgen, N., Kurz, L., Reis, P., Barthels, F., **Mach, P.**, Schutkowski, M., Freund, C. & Schwarzer D. Identification of sortase substrates by specificity profiling. *Bioorganic & Medicinal Chemistry*, doi:10.1016/j.bmc.2017.06.033 (2017).
- Heep, M., **Mach, P.**, Reautschnig, P., Wettengel, J. & Stafforst, T. Applying Human ADAR1p110 and ADAR1p150 for Site-Directed RNA Editing—G/C Substitution Stabilizes GuideRNAs against Editing. *Genes* 8, 34, doi:10.3390/genes8010034 (2017).

Fellowships

- 2019 – 2022 **Horizon 2020 Marie-Skłodowska-Curie Training Network ChromDesign**
Early Stage Researcher in the ChromDesign network, www.chromdesign.eu

Conferences

- 2023 **FMI-BR Symposium on Human Genetics, Basel, Switzerland:** “Approaches to map function and gene association of non-coding regulatory regions”, Oral presentation
Gordon Research Seminar & Conference, Ventura, United States: “Cohesin and CTCF control the dynamics of chromosome folding”, Oral presentation
- 2022 **EpIC Conference, Granada, Spain:** “Control of chromosome folding dynamics by cohesin and CTCF revealed by live-cell imaging”, Oral presentation
EMBO Workshop Nuclear structure and dynamics, Montpellier, France: “Cohesin and CTCF control the dynamics of chromosome folding”, Oral presentation
Nuclear Architecture & Function Symposium, International Nucleome Consortium, Toulouse, France: “Control of chromosome folding dynamics by cohesin and CTCF revealed by live-cell imaging”, Oral presentation
- 2019 **IMB Conference Chromosome Territories and Nuclear Architecture, Mainz, Germany:** “Measuring chromosome dynamics in the presence an absence of cohesin”, Poster presentation

Research Experience

- 2018 **Research Intern**, Group of Prof. Tony Kouzarides, Gurdon Institute, Cambridge, United Kingdom
Project: “Characterization of a novel H4R78 modification in human cell lines”
- 2017 **Research Intern**, Group of Prof. Pamela Silver, Wyss Institute, Boston, United States
Project: “Effects of uncharacterized viral proteins on the innate immune pathways”
- 2017 **Research Intern**, Group of Prof. Dirk Schwarzer, Eberhard Karls University, Tübingen, Germany
Project: “Testing alternative sorting motifs for sortase A”
- 2015 – 2016 **Research Assistant**, Group of Prof. Thorsten Stafforst and Prof. Dr. Ralf Jansen, Eberhard Karls University, Tübingen, Germany
Project: “Visualization of mRNA molecules by fluorescence *in situ* hybridization”

Teaching Experience

- 2017, 2020 **Supervision of undergraduate students**,
& 2023 trained and supervised Bachelor and Master students for thesis projects
- 2015 **Teaching Assistant**, supervised the Biochemistry I Practical Course for Biochemistry undergraduate Program, Eberhard Karls University, Tübingen, Germany

Techniques and Skills

Laboratory techniques

Molecular cloning, Cell and tissue culture, Genome engineering (CRISPR/Cas9, transposase- and recombinase-mediated techniques), Biochemical assays, next-generation sequencing-based techniques (RNA-seq, ChiP-seq, (capture) Hi-C, 4C-seq, Nanopore sequencing), Microscopy-based assays (Immunofluorescence, Fluorescence *in situ* hybridization), Live-cell imaging (Single-particle tracking and Single-molecule localization microscopy), Microscope maintenance assays

Analysis skills

Sequencing data analysis in R, Microscopy image analysis in Fiji and Python

Languages German (native), English (fluent), French (advanced)