

Imperial College London

Faculty of Medicine

Institute of Clinical Sciences

MRC London Institute of Medical Sciences

**Cohesin alters immune response gene
expression primarily by transcriptional
burst frequency modulation**

IRENE ROBLES REBOLLO

Supervised by

Prof Matthias Merkenschlager and Prof Michael Stumpf

A thesis submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy of Imperial College London

Statement of Originality

I, Irene Robles Rebollo, declare that this thesis, which I have presented for consideration for my PhD degree, is the result of my own study and research and any work carried out by others is acknowledged below and appropriately referenced in the text.

Dr Sergi Cuartero set up *Rad21* knock-out (KO) macrophages as a system. He also produced the spike-in RNA sequencing (RNAseq) and global run-on sequencing (GROseq) data, and Dr Felix Weiss produced the RNAseq from Rad21-TEV macrophages. Initial analysis on these datasets was performed by Dr Gopuraja Dharmalingam, including the fold change of ribonucleic acid (RNA) content, from the MRC-LMS Bioinformatics Facility. Differentially expressed (DE) analysis and gene set enrichment analysis (GSEA) of human acute myeloid leukaemia (AML) was performed by Dr Yi-Fang Wang, from the MRC-LMS Bioinformatics Facility. She also calculated the fraction of GROseq reads mapping to different genomic regions as well as RNA polymerase II (RNA pol II) pause ratios. Moreover, she normalised the spike-in RNAseq by replicate and spike-in. Part of the results in Chapter 1 were published in Cuartero et al. 2018 under the lead of Dr Sergi Cuartero and Prof Matthias Merkenschlager. This work provided this thesis background and was already ahead by the time I started in the lab. Therefore, this chapter is my view-point, extension and validation of this paper's conclusions.

Prof Matthias Merkenschlager, Dr Sergi Cuartero, Dr Ludovica Bruno and Dr Teresa Buenaventura Perera provided dead mice for dissection. Dr Teresa Buenaventura Perera also provided Lineage Sca1+ c-Kit+ (LSK) for single molecule fluorescence in situ hybridisation (smFISH). During his master, Adrià Cañellas initiated smFISH experiments in WT and *Rad21* KO macrophages. I kept one of his replicates for smFISH validation. Dr Chad Whilding and Dr Dirk Dormann, from the Microscopy Facility, helped to optimise acquisition and analysis of confocal images. Burst quantification strategy was developed by Dr Mohammad Mahdi Karimi,

from the MRC-LMS Bioinformatics Facility. Dr Ya Guo mapped macrophages published Hi-C data. There is a high number of smFISH experiments contained in this thesis. I carried out all image acquisition and analysis, but some experiments were carried out by others under my supervision. Alexandra Chivu performed the *Peli1* intron/exon smFISHs while she was a summer student in the lab. Natasha Irani performed two replicates of the *Actb* intron/exon smFISHs while she was a master student and started smFISH experiments on the *Peli1-Sertad2-Spred2*. Sarah Wells dissected mice for me, and helped me with many smFISH experiments, in particular in the cell mix experiments and intronic smFISHs of *Hprt1* and *Fh1*. She also performed the real time quantitative PCR (RT-PCR)s, included in Figure 2.10.

Dr Sergi Cuartero prepared macrophages for single-cell RNA sequencing (scRNAseq). Sequencing was carried out at the CNAG-CRG in Barcelona. Gustavo Rodríguez and Elisabetta Mereu performed reads alignment and initial analysis. Dr Nils Eling from Jonh Marioni's lab performed the variability analysis on scRNAseq. Dr Rowan Brackston, from Theoretical Systems Biology at Imperial College London provided help, support and a template code to perform telegraph model simulations.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

More than a scientific report, this thesis has been an adventure. There have been good days, and there have been bad days. This is a tale of challenges, personal growth and a lot of beautiful images. I have not been alone, and I do not think I would have made it alone.

First of all, I want to thank Matthias Merckenschlager, who has given me his full support and guidance all these years. Thank you, Matthias, for keeping the enthusiasm for every small result, for always encouraging me to go further, and for keeping me under control, which is not always easy. I also want to thank Michael Stumpf, my co-supervisor, who believed in me in the first place (after I proved myself worth of Julia), and who always kept the doors of his lab open.

Sergi Cuartero planted the seed of my project and took care of me during the first steps of my PhD, becoming a role model for me. My sincere gratitude to Chad Whilding and Dirk Dormann who made my microscope dreams come true. Thanks also to Yi-Fang Wang, Mohammad Mahdi Karimi and Nils Eling for coding everything I could not. Thanks also to all my students, for their trust and everything they taught me. In order of appearance: Thanks to Alexandra Chivu, for her perseverance and being my first. Thanks to Natasha Irani, for her enthusiasm. And, thanks to Sarah Wells, for increasing my productivity massively during the year that she spent as a research assistant.

For starting and finishing the PhD roller-coaster by my side, thanks to Chiara Prodani, Helen Patterson, Matt Newton, Lisa Schneider, Domhnall McHugh and Vassili Kusmartsev. Special mention to Chiara who had the curse of being literally next to me, and to Matt and Helen, for always making me feel welcome. For pushing me to grow as a scientist, thanks to Felix Weiss. For the Christmas tree collaboration and partnership during coronavirus times, thanks to Tianyi Zhang. Thanks to Teresa Buenaventura and Ludovica Bruno for the time they took to bring

me dead mice, and all the problematic orderings. For the taming of my wild writing, thanks to Ediem Al-Jibury, Marta Oliviera and Nicolás Veland. To all members of the Lymphocyte Development group, thanks for every laugh and every tear wiped. You turned the lab into my second home. Thanks also to the Theoretical Systems Biology group, for being so patient with all my mathematical questions.

Thanks to Beatriz Sánchez because her friendship is the best that gift that London has given me. Thanks also to Nina Ruíz, Patricia Alonso and Angelo Barboni for their companionship during the last difficult months. For reminding me that London is an exciting place, thanks to Laura Casado, Laura Montero and Céline Collin. Thanks to my flatmates, to Arianna Renzini, for the dinosaurs, to Laura Dal Toso, for her honesty and to Nick Pawlowski, for our short arranged marriage. Thanks to the Amplified choir, for two and a half years of songs, pints and karaokes. Thanks to my fellow PubhD organisers, Domhnall, Anna, Kyle, Camilla and Sandro, for the commitment despite storms, and to all the poor humans that I persuaded to give a talk in exchange of a beer. Worth to mention Ibles, with whom I literally started a friendship because I needed a speaker. Thanks to Nuria Lamagrande, Patricia Benito and Aleix Ramón for always answering the phone. Thanks to Juan Montiel for always taking me home after a walk. Thanks to Adrián López for bringing Sonia y Selena at the right time, to Daniel Gallego, for being one of my favourite persons and for all the musicals. Thanks also to Thomas for not giving up.

I am who I am, and I am where I am because of the unconditional support and love of my family. Thanks to my sister, Cristina, for being my best friend. Thanks to my dad for giving me the best advice, and to my mum for wanting to take care of me even when I am fully grown up and live in a different country. *Y gracias también al resto de mi familia por estar siempre ahí, aunque estemos lejos. Siento que tengamos que esperar un poco para celebrar esto juntos.*

Finally, thanks to all the mice that were sacrificed so I could carry out my experiments. And thanks to Chocolate, for teaching me that a good attitude and perseverance can achieve cookies.

Abstract

Transcription is an intermittent process that occurs in bursts. The frequency and the number of messenger RNA (mRNA) produced by these bursts determine expression levels and shape cell-to-cell variability. Burst dynamics are modulated by productive encounters between enhancers and promoters, which can be facilitated by the cohesin complex through loop extrusion. In experimental systems, cohesin deficiency disrupts cell-type-specific expression programs more severely than constitutive gene expression. Cohesin loss of function increases the self-renewal of haematopoietic progenitor stem cells (HPSCs), and cohesin-related mutations are frequent in cancers, such as AML. Cohesin defects in human AML, mouse HSPC and mature myeloid cells lead to a transcriptional state characterised by the reduced expression of inducible inflammatory genes and the elevated expression of *Myc* targets and oxidative phosphorylation (OXPHOS) genes. In this thesis, I combine RNAseq, GROseq, scRNAseq, and smFISH to explore how cohesin depletion affects transcriptional parameters at steady-state and during cell-state transitions. Using primary macrophages as a model system, I find that loss of cohesin alters the dynamics of transcriptional bursts without measurably increasing cell-to-cell variability in scRNAseq. I show that expression changes triggered by LPS are primarily orchestrated through the modulation of burst frequencies, and this modulation is blunted in cohesin-deficient macrophages. An increase in burst frequency can have consequences not just for individual cells, but also for their neighbours. Specifically, the frequency of transcriptional *Ifnb1* bursts determines the number of cells that produce *Ifnb1*-encoded type I interferon (IFN), which through paracrine signalling can alter IFN-dependent expression in neighbouring cells. Co-culture with cohesin-deficient macrophages reduces LPS-induced gene expression in wild type (WT) macrophages, indicating that cohesin defects can buffer inflammatory responses and impair cell-state transitions due to the defective control of burst frequencies. I speculate that these features contribute to the somatic selection of cohesin mutations in cancer.

Abbreviations

3'UTR 3'untranslated region

3D 3-dimensional

4-OHT 4-hydroxytamoxifen

AML acute myeloid leukaemia

ATP adenosine triphosphate

ATPase adenosine triphosphatase

bDNA branched DNA

BMP bone marrow progenitor

bp base pairs

ChIP chromatin immunoprecipitation

cDNA complementary DNA

CNV copy number variations

CP CellProfiler (Carpenter et al. 2006)

CPM counts per million

cryoEM cryo-Electron Microscopy

CT CellTrackerTM

CTD carboxy-terminal domain

DE differentially expressed

DNA deoxyribonucleic acid

eRNA enhancer RNA

FACS flow-activated cell sorting

FISH fluorescent in situ hybridisation

FQ FISHquant (Mueller et al. 2013)

GSEA gene set enrichment analysis

gqPCR genomic DNA quantitative RT-PCR

GROseq global run-on sequencing

GTF general transcription factor

h hours

H3K27ac histone H3 lysine 27 acetylation

H3K4me1 histone H3 lysine 4 monomethylation

HAT histone acetyltransferase

HDAC histone deacetylase

HPSC hemaetopoietic progenitor stem cell

ICR imprinting control region

IF immunofluorescence

IFN interferon

Inr Initiator element

ISH in situ hybridisation

ISRE interferon-stimulated response element

kb kilobases

KO knock-out

LCM laser capture microdissection

lncRNA long non-coding mRNA

LPS lipopolysaccharide

LSK Lineage Sca1+ c-Kit+

Mb Megabases

min minutes

MPP multi-potent progenitor

mRNA messenger RNA

mtDNA mitochondrial DNA

ncRNA non-coding RNA

NELF negative elongation factor

nuDNA nuclear DNA

ODE Ordinary Differential Equation

OXPHOS oxydative phosphorylation

PAMP pathogen associated molecular pattern

PBS phosphate-buffered saline

PCR polymerase chain reaction

PIC pre-initiation complex

p-TEFb positive transcription elongation factor b

RNA ribonucleic acid

RNA pol II RNA polymerase II

RNAseq RNA sequencing

rpm revolutions per minute

RT room temperature

RT-PCR real time quantitative PCR

s seconds

scRNAseq single-cell RNA sequencing

SDS-PAGE Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SE super-enhancer

SEC super elongation complex

SMC structural maintenance of chromosomes

SMOC supramolecular organising centre

smFISH single molecule fluorescence in situ hybridisation

SNPs single-nucleotide polymorphisms

ssDNA single-strand DNA

SV structural variation

TAD topologically-associated domain

TAF TBP associated protein

TBP TATA-box binding protein

TF transcription factor

TLR toll-like receptor

TSS Transcription Start Site

UMI unique molecular identifiers

vs versus

WT wild type

Table of Contents

Statement of Originality	2
Copyright Declaration	4
Acknowledgements	5
Abstract	6
Abbreviations	7
Table of Contents	12
List of Figures	22
List of Tables	25
1 Introduction	28
The cohesin complex	29
Structure of cohesin	29
Changes in cohesin conformation	30
Interactions of cohesin with chromatin	30

Association of cohesin with chromatin	30
Cohesin loading and unloading	31
Chromatin loop extrusion by SMC complexes	31
Principles of genome folding	32
Chromatin structures	33
Folding forces	35
CTCF	35
Loop extrusion	36
Studying genome folding	37
Imaging techniques	37
Proximity ligation techniques	38
Ligation-free techniques	39
Modulation of genome folding	39
Disruption of genome contacts	40
The regulation of gene expression	41
Nucleosomes	41
Nucleosome remodelling	41
Histone modification	42
RNA pol II control	43
RNA pol II recruitment	43
RNA pol II pause release	44
Functional elements	46

Promoter regions	47
Enhancers	47
Enhancer transcription	48
Enhancers and chromatin loops	49
Measuring gene expression in individual cells	50
In Situ Hybridization techniques	50
Single-molecule fluorescence in situ hybridisation	51
Live cell imaging of gene expression	52
Single cell RNA sequencing techniques	53
Cell-to-cell variability	54
Transcriptional bursts	56
Modelling transcriptional bursts	56
Control of transcriptional bursts	57
The macrophage antimicrobial response as a model	59
Bone marrow-derived Macrophages	59
The antimicrobial response	59
Collective behaviour in immune responses	61
Cohesin in hematopoiesis and leukaemia	62
Results	63
2 Cohesin at the crossroad between inflammation and cancer	64
Cohesin mutations in Acute Myeloid Leukaemia	64
Studying the role of cohesin in gene regulation	66

Knocking out the cohesin subunit RAD21 in macrophages	66
Weakened immune response in <i>Rad21</i> KO macrophages	68
Cohesin deficiency leads to unpaired immune responses	68
Patterns associated to DE genes in Rad21KO macrophages	71
Inducible gene expression failure: IFNs	71
ISRE motif associates to genes with reduced expression	71
Rescue of late inducible genes by IFN γ in <i>Rad21</i> KO	71
Redistribution of cellular resources upon LPS and IFN γ	73
Repression failure: OXPHOS and Myc	74
Increased mitochondrial content in Rad21KO macrophages	75
Mitochondrial genes are up-regulated in <i>Rad21</i> KO macrophages	75
<i>Rad21</i> KO macrophages have increased mitochondria content	76
Mitochondria during macrophage differentiation	76
Mitochondrial regulators during macrophage differentiation	77
MYC-MAX-MXD1 axis during macrophage differentiation	78
Discussion	80

3 The quantitative impact of cohesin on transcription 83

Transcriptional control and cohesin	83
Global transcriptional changes in Rad21KO macrophages	84
Transcription is a major cause of gene expression changes in <i>Rad21</i> KO macrophages	84
Increase in total RNA content in <i>Rad21</i> KO macrophages	84
Reduction in RNA pol II pause ratios in <i>Rad21</i> KO macrophages	86

Single-molecule perspective of gene expression	86
Validation of bDNA smFISH approach	87
Increasing the throughput of image acquisition	88
Quantification of transcription bursts	89
Housekeeping gene expression: <i>Actb</i>	90
Inducible gene expression: <i>Il12b</i> and <i>Peli1</i>	92
Detailed timecourse expression data	92
Mature mRNA analysis: Different gene activation distributions	92
Nascent mRNA analysis: Reduced burst frequency in <i>Rad21</i> KO.	92
Repressed genes: <i>Fh1</i> and <i>Hprt</i>	95
Gene repression by LPS and IFN- γ	95
<i>Hprt</i> burst dynamics	95
Pre-treatment with IFN- γ restores the ability of <i>Rad21</i> KO macrophages to repress the burst frequency of <i>Fh1</i> upon LPS	97
The <i>Spred2-Sertad2-Peli1</i> locus	100
smFISH detection of eRNA	100
Enh eRNA bursting	102
Enh increases burst frequency, not TSS intensity upon LPS-activation.	102
Enh showed a slight increase in burst frequency in <i>Rad21</i> KO macrophages	102
Inducible genes are mainly regulated by burst frequency upon LPS activation	103
<i>Rad21</i> KO macrophages fail to increase the burst frequency of inducible genes upon LPS-activation	104

<i>Spred2</i> burst frequency is reduced upon LPS-activation.	105
<i>Spred2</i> burst frequency is higher in <i>Rad21</i> KO macrophages	105
Burst co-dependence	105
Assigning multiple TSSs to a single locus	105
TSS-TSS distances	107
Active TSS-TSS distances remain unchanged upon LPS-activation	107
Gene-enhancer burst distances are similar between WT and Rad2KO, but gene-gene distances are generally longer in <i>Rad21</i> KO . . .	108
Co-dependence between burst frequencies	109
Enh and the two inducible gene burst frequencies are co-dependent . . .	109
Enh bursts associate to <i>Peli1</i> or <i>Sertad2</i> bursts more often in WT than in <i>Rad21</i> KO macrophages	111
General features of TSS intensities and frequencies	111
Expression changes between <i>Rad21</i> KO and WT are explained by changes in burst frequencies	111
Overall, <i>Rad21</i> KO macrophages displayed bigger TSS intensities	113
Distance to a TAD boundary	113
Long-term cohesin depletion does not associate with boundaries	113
Short-term cohesin depletion associates with boundaries	115
scRNAseq analysis	115
<i>Rad21</i> KO defects in LPS-inducible expression are general	118
Similar gene overdispersions in WT and <i>Rad21</i> KO macrophages	119
Discussion	120

4	<i>Ifnb1</i> burst fraction directs collective behaviour in response to LPS	124
	Collective behaviour of immune responses	124
	<i>Ifnb1</i> expression	126
	<i>Ifnb1</i> production occurs only in a fraction of cells	126
	<i>Ifnb1</i> producers are halved in <i>Rad21</i> KO macrophages	126
	<i>Ifnb1</i> burst fraction is reduced in <i>Rad21</i> KO cells	127
	<i>Ifnb1</i> is induced in a subset of cells, whereas some <i>Ifnb1</i> -dependent genes are broadly expressed	128
	Identifying consequences of <i>Ifnb1</i> expression	129
	<i>Ifnb1</i> -dependent gene selection for smFISH	130
	<i>Ifnb1</i> -dependent expression in <i>Ifnb1</i> KO-mixed populations	130
	<i>Cxcl10</i> induction in WT macrophages	131
	<i>Cxcl10</i> induction fails in <i>Rad21</i> KO and <i>Ifnb1</i> KO macrophages	134
	Rescue of <i>Ifnb1</i> KO <i>Cxcl10</i> and <i>Mx1</i> induction via co-culture	134
	Co-culture with <i>Ifnb1</i> KO cells strengthen <i>Cxcl10</i> induction	136
	Inducible response in <i>Rad21</i> KO-WT co-cultures	136
	<i>Ifit1</i> and <i>Mx1</i> induction after 3h of LPS	136
	Differences between WT and <i>Rad21</i> KO <i>Ifnb1</i> -dependent genes induction are diluted in mixed populations	138
	WT- <i>Rad21</i> KO co-culture drops <i>Ifnb1</i> and <i>Il12b</i> induction in WT cells .	138
	Discussion	140

5 Summary and final discussion	142
Cohesin, inflammation and cancer	143
Cohesin, enhancers, and bursts	144
Burst frequency and paracrine signalling	146
Mutations and cell-to-cell communication	147
Conclusion	148
6 Materials and Methods	149
Materials and Methods	149
Bone marrow derived macrophages	149
Cell treatments	150
L929 conditional media production	150
LPS-activation	150
IFN- γ pre-treatment	150
CellTracker staining	150
Cell mixing	150
Genomic DNA extraction	151
RNA extraction	151
cDNA synthesis	152
Real time quantitative PCR	152
Western blot	154
Immunofluorescence	155
Flow-activated cell sorting (FACS)	157

Mitochondria quantification	157
Single molecule Fluorescent In Situ Hybridization	158
Validation of bDNA FISH	158
High throughput smFISH	159
High throughput smFISH image analysis	160
Cell, nuclei and mature mRNA identification	160
Identification and quantification of transcriptional bursts	160
Association of multiple gene burst to a particular locus	161
smFISH overdispersion	162
Telegraph model simulations	162
Expression and mutation data of human AML	164
Mutation gene sets	164
Patient gene expression processing	165
Expression data of mouse LSK cohesin knockdown	165
<i>Rad21</i> KO macrophage data	166
Macrophage RNAseq data	166
GROseq data	167
Bioinformatic analysis	167
LPS-inducible gene subset	167
Constitutive gene subset	167
LPS-repressed gene subset	168
Macrophage enhancers	168

Gene Set Enrichment Analysis	168
RNA pol II pause ratio calculation	168
Known Motif Promoter Enrichment Analysis	168
Voronoi maps	169
Single cell RNAseq	169
WT and <i>Rad21</i> KO macrophage scRNAseq	169
Variability analysis	170
Module score	170
Statistical analysis for comparisons	171
qPCR data comparisons	171
Comparison between mitochondria imaging data	171
qPCR data comparisons	171
Comparisons between the expression log2FoldChanges in one comparison vs the log2FoldChanges in another	171
Burst frequency and size comparisons	171

7 Supplementary Material 173

Bibliography 180

List of Figures

1.1	The cohesin complex	29
1.2	Extrusion in real time	32
1.3	The hierarchical folding of the genome	34
1.4	Loop extrusion	36
1.5	Studying genome folding	38
1.6	Nucleosome regulation	42
1.7	Transcription initiation	45
1.8	Brief history of RNA visualisation	51
1.9	Perspectives on cell-to-cell variability	55
1.10	Self-renewal vs differentiation during hemaetopoiesis	62
2.1	AMLs with cohesin mutations present a distinctive expression program that clusters with <i>FLT3</i> and <i>WT1</i> mutations, with increased OXPHOS and <i>Myc</i> targets and reduced inflammatory hallmarks	65
2.2	Cohesin-deficient macrophages	67
2.3	<i>Rad21</i> KO macrophages detected antimicrobial signals but failed to establish a strong immune response	69
2.4	Expression hallmarks are inverted between macrophages establishing an antimicrobial response and cells with cohesin defects	70

2.5	IFN γ pre-treatment rescues part of late inducible genes in <i>Rad21</i> KO macrophages	72
2.6	Redistribution of cellular resources upon LPS and IFN γ	73
2.7	<i>Rad21</i> KO macrophages failed to repress genes with c-Myc and NRF1 motifs	74
2.8	Increased relative mtDNA content in <i>Rad21</i> KO macrophages	77
2.9	Increased mitochondrial mass in <i>Rad21</i> KO cells	78
2.10	Macrophage increase on mitochondrial content during differentiation and expression of potential regulators	79
3.1	Global transcription changes in <i>Rad21</i> KO macrophages	85
3.2	Establishing bDNA smFISH in the lab	87
3.3	Imaging and quantifying transcriptional bursts	89
3.4	smFISH analysis revealed <i>Actb</i> expression differences between WT and <i>Rad21</i> KO macrophages	91
3.5	Single-molecule perspective of inducible gene expression in WT and <i>Rad21</i> KO macrophages	93
3.6	<i>Il12b</i> and <i>Peli1</i> transcriptional bursts	94
3.7	<i>Rad21</i> KO cells fail to increase burst frequencies of inducible genes	96
3.8	Transcriptional burst of a <i>Myc</i> target (<i>Hprt</i>) and an OXPHOS gene (<i>Fh1</i>)	98
3.9	The <i>Spred2-Sertad2-Peli1</i> genomic region	101
3.10	Enh eRNA detection	102
3.11	Burst fraction and burst frequency of enhancer Enh and locus genes in WT and <i>Rad21</i> KO macrophages	103
3.12	Locus burst co-localisation	106
3.13	Assigning TSSs to locus 1 or 2	107

3.14	Distances between detected active TSSs	108
3.15	Co-dependence of burst frequencies	110
3.16	Expression changes between <i>Rad21</i> KO and WT are explained by changes in burst frequencies, but <i>Rad21</i> KO displayed bigger TSS intensities in the genes studied	112
3.17	Long and short-term cohesin depletion display different associations between DE genes and distance to a TAD boundary	114
3.18	Upon LPS activation, deregulated LPS inducible and repressed genes tend to be far away from a TAD boundary	116
3.19	Reduction in LPS response is widespread in <i>Rad21</i> KO macrophages	117
3.20	<i>Rad21</i> KO macrophages did not show an increase in gene overdispersions	118
3.21	Validation of scRNAseq residual dispersion by smFISH	119
4.1	The fraction of <i>Ifnb1</i> -producers and their <i>Ifnb1</i> levels are reduced in <i>Rad21</i> KO macrophages. <i>Ifnb1</i> -producers express higher levels of <i>Peli1</i> and <i>Stat1</i>	125
4.2	<i>Ifnb1</i> burst frequency is reduced in <i>Rad21</i> KO macrophages	127
4.3	<i>Ifnb1</i> and <i>Il12b</i> are induced in a fraction of macrophages whereas <i>Ifnb1</i> -dependent genes are broadly induced upon LPS	129
4.4	<i>Ifnb1</i> -dependent genes for smFISH	130
4.5	Identifying follower and leader cells in a mix population	132
4.6	Rescue of <i>Cxcl10</i> and <i>Mx1</i> expression in <i>Ifnb1</i> KO via co-culture with WT and <i>Rad21</i> KO cells	133
4.7	Co-coculture with <i>Ifnb1</i> KO leads to an increase in <i>Cxcl10</i> expression upon LPS-activation	135

4.8	<i>Rad21</i> KO and WT co-cultures leads to a partial rescue of <i>Ifnb1</i> -dependent genes in <i>Rad21</i> KO. <i>Ifnb1</i> -dependent genes are reduced in WT macrophages via co-culturing with <i>Rad21</i> KO macrophages.	137
4.9	WT macrophages co-cultured with <i>Rad21</i> KO showed a reduction in the fraction of cells expressing <i>Ifnb1</i> and in the magnitude of <i>Il12b</i> expression upon LPS-activation	139
5.1	Macrophage enhancers around <i>Peli1</i>	145
7.1	Images of <i>Hprt</i> bursts	175
7.2	Images of <i>Fh1</i> bursts	176
7.3	Lowly express genes tend to be up-regulated whereas highly express genes tend to be down-regulated in <i>Rad21</i> KO	177
7.4	Increase inducible gene expression in <i>Ifnb1</i> -producer cells by replicate	178
7.5	<i>Mx1</i> and <i>Cxcl10</i> counts in <i>Ifnb1</i> KO, WT, <i>Rad21</i> KO and mix populations by replicate	179

List of Tables

2.1	Differential expression analysis of genes involved in IFN signalling in <i>Rad21</i> KO macrophages vs WT	71
2.2	NRF1 motif enrichment	75
2.3	Myc motif enrichment	75
3.1	RNA content comparisons from spike-in RNAseq data	84
3.2	Optimisation of smFISH image acquisition	88
3.3	<i>Actb</i> transcription measures in WT and <i>Rad21</i> KO macrophages	90
3.4	Summary of <i>Il12b</i> and <i>Peli1</i> transcription measures by smFISH in WT and <i>Rad21</i> KO macrophages	94
3.5	<i>Hprt</i> and <i>Fh1</i> transcriptional burst measures in WT and <i>Rad21</i> KO macrophages	99
3.6	TSS intensities and burst frequencies measured at the <i>Spred2-Sertad2-Peli1</i> locus in WT and <i>Rad21</i> KO macrophages	104
4.1	Differential expression analysis of <i>Ifnb1</i> by RNAseq	126
4.2	Summary of transcriptional burst measures of <i>Ifnb1</i> timecourse in WT and <i>Rad21</i> KO macrophages	128
6.1	Genomic qPCR primers	152
6.2	RT qPCR primers	153

6.3	bDNA FISH probes	159
7.1	Distance to nearest TAD boundary of genes studied for burst frequency and size	173
7.2	Number of technical replicates for replicates for mix and unmix samples in <i>Il12b</i> and <i>Ifnb1</i> smFISH experiments	174
7.3	Differential expression analysis in LPS-activated <i>Rad21</i> KO macrophages with and without IFN priming by RNAseq	174

Chapter 1

Introduction

The transcription of most metazoan genes is a discontinuous process that occurs in bursts (Halpern et al. 2015; Raj et al. 2006a). Burst size and frequency tune expression levels and determine the shape of mRNA distributions. Enhancer-promoter proximity can alter transcriptional burst parameters (Bartman et al. 2016; Deng et al. 2012; Fukaya et al. 2016). The cohesin protein complex can embrace and extrude chromatin fibres, forming domains of local chromatin interactions and controlling contact frequencies between gene regulatory elements in 3-dimensional (3D) nuclear space (Davidson et al. 2019; Fudenberg et al. 2016; Haarhuis et al. 2017; Jaeger et al. 2020; Kim et al. 2019; Rao et al. 2014, 2017; Schwarzer et al. 2017). In certain experimental settings, cohesin deficiency disrupts cell-type-specific expression programs more severely than constitutive gene expression (Cuartero et al. 2018; Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017; Seitan et al. 2013). Still, the relationship between genome folding and the control of gene expression is not fully understood.

To dissect cohesin's role in transcriptional regulation, I work on primary quiescent cells, bone marrow-derived macrophages. Macrophages are innate immune cells with a well-characterised antimicrobial response. Detection of lipopolysaccharide (LPS), a component of Gram – bacteria, initiates a cascade of events that transforms their transcriptional landscape (Bhatt et al. 2012; Kaikkonen et al. 2013; Ostuni et al. 2013).

In this thesis, I explore how cohesin depletion affects gene regulation and transcriptional parameters of macrophages while they transition towards an LPS-activated state, and examine the downstream consequences for the cell population as a whole.

The cohesin complex

Structure of cohesin

Cohesin is a structural maintenance of chromosomes (SMC) complex. SMC complexes appear in all organisms and can embrace chromatin fibres, they are fundamental for 3D genomic organisation, and have functions that range from chromatin compaction to gene regulation. Cohesin was initially identified as the complex that keeps sister chromatids together until chromosome segregation at the onset of anaphase (Michaelis et al. 1997; Uhlmann et al. 2000), but it also participates in deoxyribonucleic acid (DNA) repair, genome folding and gene regulation.

SMC complexes are ring-shaped. They are formed by two SMC subunits, an α -kleisin and an Scc3 unit. SMC proteins have two globular ATPase (adenosine triphosphatase) domains at their N- and C- terminals separated by a long-coiled coil domain. This coiled-coil domain folds

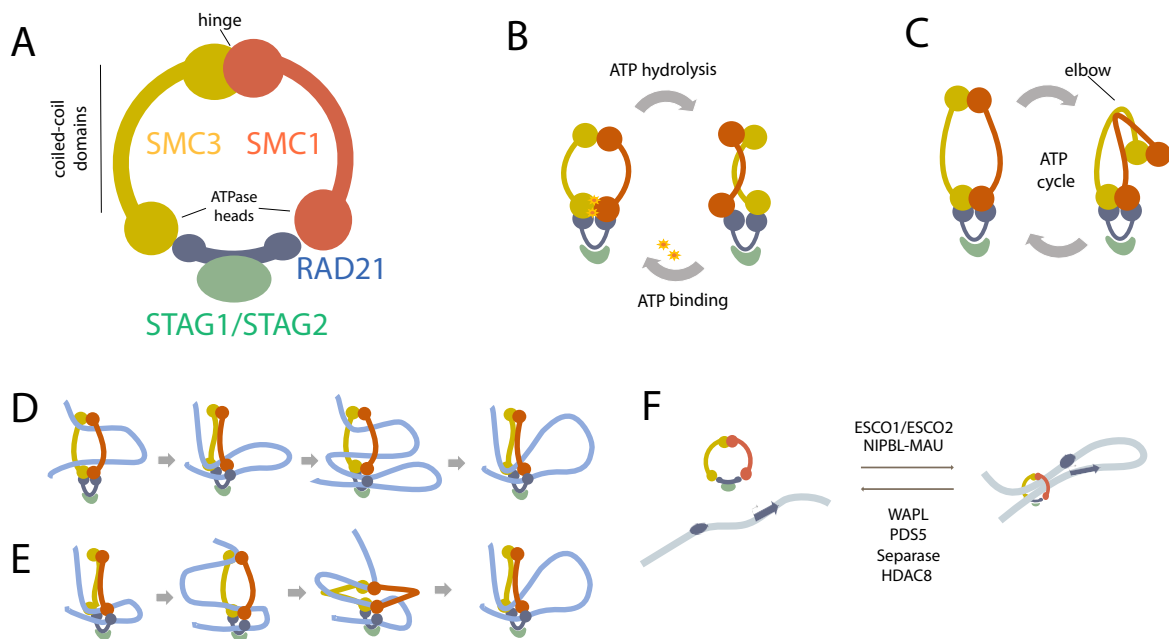


Figure 1.1: **The cohesin complex.** A) SMC1, SMC3, RAD21 and STAG1 or STAG2 are the core components of the cohesin complex. SMC1 and SMC3 dimerise at the hinge and interact with the RAD21 subunit through their ATPase heads, SMC hinge and ATPase head are separated by a coiled coil domain. B) ATP hydrolysis rotates SMC ATPase heads. C) SMC's coiled coil domains can bend over themselves at the elbow. D) Cohesin extrusion through the capture/pumping model. E) Cohesin extrusion through the scrunching model. F) Cohesin chromatin loading and unloading components. Figure based on Cacciatore and Rowland 2019 and Hassler et al. 2018.

over itself at the hinge. SMC proteins form V-shaped heterodimers with the monomer's hinges meeting at the V's vertex. At the other end, the α -kleisin interact with the SMC's ATPase heads, closing the ring. The central region of α -kleisin is disordered and acts as a recruitment platform. In the case of cohesin, the two SMC subunits are SMC1 and SMC3, and its α -kleisin subunit is RAD21. The Scc3 units in vertebrates can be either STAG1 or STAG2 (Gruber et al. 2003) (Figure 1.1A). Cohesin-STAG1 and Cohesin-STAG2 carry redundant but also distinct functions (Kojic et al. 2018). STAG2 is located in the human X chromosome and it is the most common cohesin-related mutation in cancer, cells without STAG2 can rely on STAG1 to proliferate without abnormal chromosome segregation (Van Der Lelij et al. 2017).

Changes in cohesin conformation

The movement of SMC adenosine triphosphatase (ATPase) heads is coupled to ATP hydrolysis cycle. When adenosine triphosphate (ATP) is bound, SMC heads face each other. ATP hydrolysis forces SMC ATPase heads to rotate, prompting the juxtaposition of the coiled-coil domains (Cacciatore and Rowland 2019; Diebold-Durand et al. 2017) (Figure 1.1B). Coiled-coil domains have an elbow that enables SMC complexes to fold over themselves. Elbow movement can be driven as well by the rotation of the SMC heads (Bürmann et al. 2019) (Figure 1.1C).

Interactions of cohesin with chromatin

Association of cohesin with chromatin

There are two modes of cohesin binding to DNA. Topological binding involves opening and closing the cohesin ring around DNA, whereas non-topological binding does not (Srinivasan et al. 2018). There are two proposed interfaces for DNA entry and exit gates: the hinge and the SMC1-RAD21 interphase (Buheitel and Stemmann 2013; Srinivasan et al. 2018).

SMC3 acetylation by ESCO1 and ESCO2 promotes stable cohesin association to chromatin (Figure 1.1F). ESCO1 is involved in 3D genome organisation whereas ESCO2 is more active during the S-phase. ESCO1 has constitutive activity and co-localise with cohesin and CTCF at boundaries of chromatin loops (Alomer et al. 2017). HDAC8 can deacetylate SMC3 (Deardorff et al. 2012).

Cohesin loading and unloading

Cohesin loading into chromatin is dependent on NIPBL, MAU and ATP hydrolysis (Ciosk et al. 2000; Davidson et al. 2019) (Figure 1.1F). cryo-Electron Microscopy (cryoEM) has solved the structure of the cohesin-NIPBL complex. There, NIPBL adopts a bent structure that interacts predominantly with RAD21, STAG1, as well as the hinge and ATPase heads (Chao et al. 2017; Shi et al. 2020). When NIPBL associates to the cohesin ring, DNA is entrapped within the STAG subunit, the α -kleisin and SMC ATPase heads (Shi et al. 2020).

Cohesin unloading depends on deacetylation and the interaction of cohesin with WAPL and PDS5 (Figure 1.1F). WAPL and PDS5 dissociate cohesin from chromatin. Without WAPL and PDS5, the residence time of cohesin in chromatin increases (Gassler et al. 2017; Haarhuis et al. 2017; Wutz et al. 2017). At the onset of anaphase, cohesin is removed by separase, which proteolytically cleaves α -kleisin to enable the segregation of sister chromatids. (Michaelis et al. 1997; Uhlmann et al. 2000).

Chromatin loop extrusion by SMC complexes

Chromatin extrusion is a motor-driven activity consisting of an SMC complex translocating across the DNA (see Banigan and Mirny 2020 for a recent review). In this process, chromatin entrapped within the SMC ring generate and enlarge loops. Recent single-molecule microscopy studies have visualised extrusion in vitro, first with condensin, a structural homolog of cohesin (Ganji et al. 2018) and then with cohesin (Davidson et al. 2019; Kim et al. 2019). In these experiments, a DNA molecule is tethered between two beads, SMC complexes are added, and DNA loops are visualised as a local accumulation of DNA. When flow is applied, the loops stretch out and become U-shaped (Figure 1.2).

Cohesin extrusion requires NIPBL, MAU and ATP. Interestingly, while cohesin moves along the DNA, MAU and NIPBL remain attached to the loop boundary. NIPBL-cohesin association during extrusion suggests that NIPBL not only loads cohesin, but also actively promotes cohesin extrusion activity, probably through stimulating SMC ATPase activity. Crosslinking experiments suggest that loop extrusion does not require opening of the cohesin ring (Davidson et al. 2019).

In vitro studies have shown that loop extrusion by cohesin is symmetrical (two-sided). Cohesin extrusion can proceed for more than 10 kilobases (kb) at 0.5 - 2 kb per seconds (s) spending ~ 1 ATP/s (Davidson et al. 2019; Kim et al. 2019). While the exact mechanism is subject to debate, both capture/pumping and scrunching models are plausible (Banigan and Mirny 2020). In the capture/pumping models, ATP hydrolysis rotates SMC heads, causing their engagement and disengagement, pushing DNA into the kleisin compartment (Figure 1.1D). In the scrunching models, the SMC elbow bends and reels DNA into a loop (Figure 1.1E). NIPBL and the Scc3 units STAG1 and STAG2 have been proposed to act as a DNA anchor during extrusion (Banigan and Mirny 2020; Cacciatore and Rowland 2019). Cohesin extrusion by monomers and dimers have been reported by in vitro studies (Davidson et al. 2019; Kim et al. 2019).

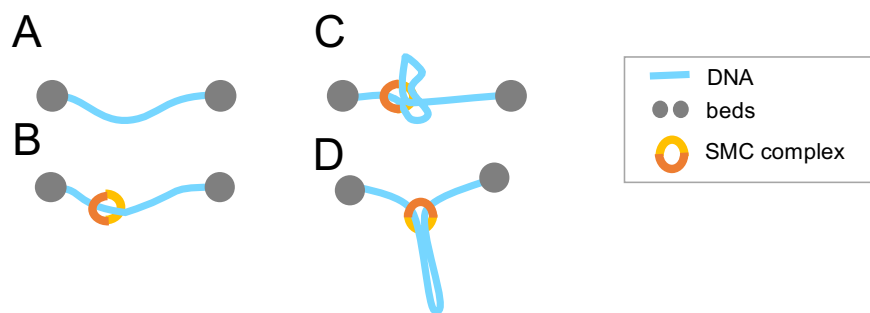


Figure 1.2: **Extrusion in real time.** Single molecule tracking experiments that visualise extrusion generally consist of A) entrapping a piece of DNA by the ends between two beds. B) Then, a SMC complex is loaded. C) If the complex can perform extrusion, a local accumulation of DNA occurs. If flow is applied, DNA loop will acquire an U shape with the SMC complex present at the loop boundary.

Principles of genome folding

During interphase, chromosomes decondense occupying the entire nuclear space. Despite appearing disordered, interphase chromatin folds into dynamic but organised structures. The 'Fractal globule' is a mathematical framework that describes how chromatin occupies the nuclear space. A fractal globule is a type of self-organised structure that folds hierarchically. In a fractal globule, monomers aggregate into small globules, which act as monomers to form again globules which will again act as monomers for more giant globules. Monomers form globules again and again until only one globule-of-globules-of-globules remains. Mathematical models

indicate that fractal globules facilitate dynamic reshaping of polymers while avoiding entangling (Grosberg et al. 1993; Lieberman-Aiden et al. 2009). Genome folding is hierarchical and shaped by compartmentalisation and loop extrusion.

Chromatin structures

Chromosomes interact primarily with themselves, occupying discrete spaces within nuclei termed **chromosome territories** (Figure 1.3A). DNA in situ hybridisation (ISH) experiments demonstrate the presence of chromosome territories a hundred years after they were hypothesised (Cremer and Cremer 2010; Schardin et al. 1985). Their relative position displays cell-type preferences. Active transcription sites reside on the surfaces of chromosome territories (Shah et al. 2018). These surfaces (about 20% of the chromosome volume) intermingle and make specific contacts with adjacent chromosome territories (Branco and Pombo 2006; Kempfer and Pombo 2019; Quinodoz et al. 2018).

Compartments are self-organising regions with similar chromatin states and histone marks (Lieberman-Aiden et al. 2009). Euchromatin aggregates in the A compartment, whereas heterochromatin aggregates in the B compartment. B compartments are generally located near the nuclear lamina and within the interior of chromosome territories. A compartments reside on the surface of chromosome territories (Figure 1.3A). Compartmentalisation is mutable. During differentiation, at least 36% of the genome switches compartments (Dixon et al. 2015). Compartment structures can be detected in single cells (Wang et al. 2016a).

A **topologically-associated domain (TAD)** is a self-interacting region of 40 kb to 3 Megabases (Mb) where the probability of interaction of sequences within the same TAD is higher than with sequences in neighbouring TADs (Figure 1.3A and B, Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). TADs are cis-regulatory units where most enhancer-promoter interactions occur. TAD-like structures are present in many different organisms and they often represent units of syntenic conservation (Lazar et al. 2018; Rudan et al. 2015). The features of TADs are conserved across vertebrates, but are not universal (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014; Szabo et al. 2019). For example, transcription inhibition disrupts TADs in fruit flies but not in mammalian cells. Furthermore, CTCF is found at the boundaries of vertebrate TADs (Kaaib et al. 2018; Rao et al. 2014), but it is absent at *Drosophila* boundaries which are

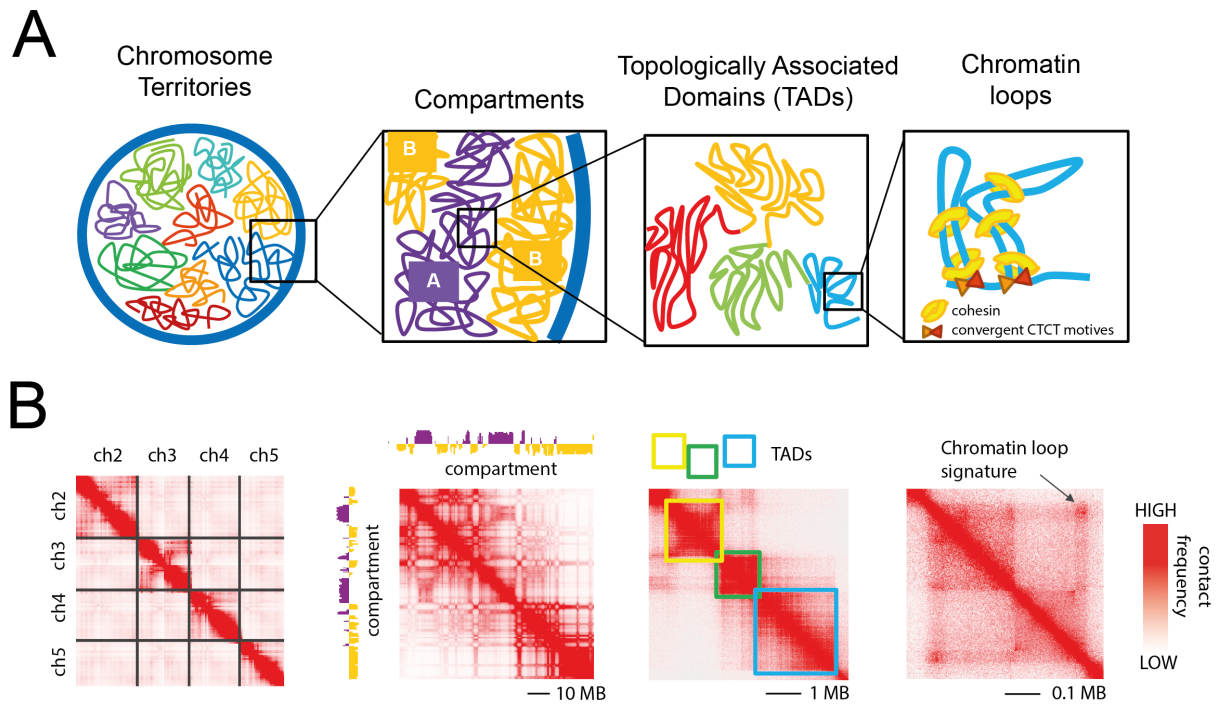


Figure 1.3: **The hierarchical folding of the genome.** A) Chromosomes occupy discrete nuclear spaces named chromosome territories. In these territories, chromatin segregates into the A (active) or B (inactive) compartment. On a smaller scale, chromatin folds into globular regions named TADs. The smallest folding units are chromatin loops, formed by an extrusion factor stopped at specific sites by an insulator. B) Examples of DNA contacts generated by these structures in Hi-C. Colour indicates contact frequency. Dark squared regions identify TADs, where interaction probability inside is higher than outside. The two boundary regions of a loop exhibit a high contact probability with each other. Hi-C maps from Rao et al. 2017 visualised with JuiceBox.

enriched with other insulator elements (Sexton et al. 2012).

DNA conformation is dynamic in single cells, but TADs are defined by population studies, but they are recurrently formed in single-cells. In contrast, inter-TAD contacts are highly variable (Nagano et al. 2013). TADs in the A compartment engage in more extended interactions than B TADs (Su et al. 2020). Genes that share a TAD correlate their gene expression, and active genes are more likely to be found at TAD boundaries (Nora et al. 2012). Loop domains are TADs that depend on cohesin to retain their structure. Compartment domains do not disappear if cohesin is depleted (Rao et al. 2017). **Chromatin loops** are local structures with their boundaries in proximity. Vertebrate boundaries have cohesin and CTCF in convergent orientations (Kaaib et al. 2018; Rao et al. 2014). TADs can accommodate multiple chromatin loops, one, or none.

Folding forces

There are two interplaying forces driving genome architectural organisation: compartmentalisation and loop extrusion. **Compartmentalisation** is defined as the chromatin tendency to cluster with chromatin with similar histone marks. Gene activity drives compartmentalisation. SMC complexes mediate **loop extrusion**, which is delimited by boundary elements. Compartmentalisation and loop extrusion can synergise or counteract one another (Nuebler et al. 2018). If loop extrusion is disrupted, loop domains disappear but compartmentalisation is retained and enhanced, (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017). Conversely, the maternal zygotic genome has TADs and loops, but lacks compartments (Flyamer et al. 2017).

CTCF

CTCF and cohesin co-localize in the genome (DeMare et al. 2013; Parelho et al. 2008; Wendt et al. 2008) and their disruption affects genome 3D organisation. CTCF and cohesin colocalise at loop boundaries (Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017), and they control genome folding through loop extrusion (Fudenberg et al. 2016; Nasmyth 2001; Sanborn et al. 2015a).

CTCF or CCCTC-binding factor is a conserved 11 zinc finger protein (Kim et al. 2015). Initially, CTCF was described as an enhancer insulator (Bell et al. 1999). CTCF binds preferentially to unmethylated sequences and methylation can explain 41% of differential CTCF binding (Wang et al. 2012). Cohesin is the major interaction partner of CTCF (Kim et al. 2015). Nevertheless, cohesin depletion does not interfere with CTCF binding, whereas CTCF depletion alters cohesin anchoring along chromatin (Rao et al. 2014; Wendt et al. 2008). Therefore, CTCF positions cohesin along the genome, not the other way around. CTCF C-terminal interacts with cohesin STAG and RAD21 subunit (Li et al. 2020). STAG subunit interacts with CTCF through its stromalin conservative domain. Loop boundaries require STAG-CTCF interaction. Mutations in the STAG-CTCF interface alter CTCF's activity as a barrier to loop extrusion.

Loop extrusion

Loop extrusion is mediated by cohesin and CTCF. Cohesin extrudes chromatin, and when it contacts CTCF bound to the DNA in a specific orientation, extrusion stops (Figure 1.4). According to this model, CTCF defines loop boundaries, but cohesin extrusion forms the loops. This process can facilitate promoter-enhancer interactions.

Experimental evidence for loop extrusion comes from different sources. In the absence of cohesin and CTCF, compartment interactions become stronger, while most contact domains and chromatin loops disappear (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017). In the absence of WAPL and PDS5 (which remove cohesin from chromatin), extended loops are found genome-wide (Gassler et al. 2017; Haarhuis et al. 2017; Wutz et al. 2017). Therefore, cohesin residence time on chromatin regulates loop length. The residence time of cohesin and CTCF is in the order of minutes (Hansen et al. 2017), suggesting a rapid loop formation. Re-introduction of cohesin can re-build loops in less than an hour (Rao et al. 2017). Loop boundaries have CTCF motifs in convergent orientation (Rao et al. 2014). CTCF motif deletion can disrupt TAD boundaries and lead to TAD fusion (Dowen et al. 2014), and CTCF motif inversion can switch CTCF loop boundary partner (Guo et al. 2015; Sanborn et al. 2015b; Wit et al. 2015). Furthermore, single-molecule studies have confirmed cohesin extrusion *in vitro* on naked DNA (Davidson et al. 2019) and nucleosomal DNA (Kim et al. 2019). Finally, CTCF N-terminal

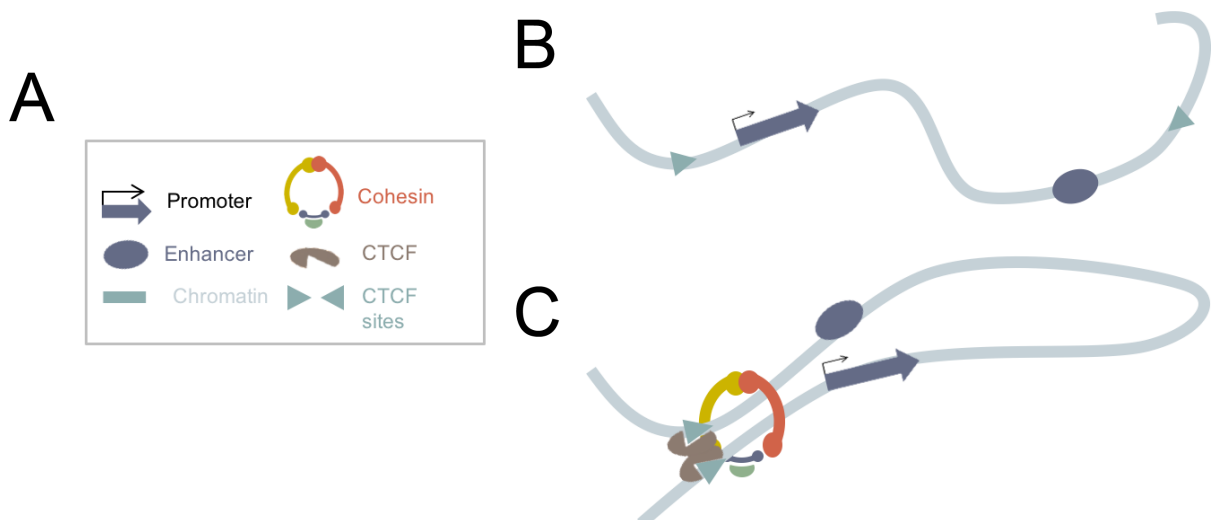


Figure 1.4: **Loop extrusion.** CTCF motifs define loop boundaries, whereas cohesin forms loops by extruding chromatin fibres, increasing the contact probability of DNA elements within the loop. A) Legend. B) Chromatin without CTCF and cohesin. C) Chromatin with CTCF and cohesin.

interacts with cohesin STAG2 and RAD21. Mutations in this domain allow cohesin to surpass CTCF barriers (Li et al. 2020).

In the absence of cohesin the positioning of these TAD-like structures is randomised in single-cell, and therefore TADs disappear when populations are averaged (Bintu et al. 2018). In this sense, the correct positioning of TADs depends on cohesin and CTCF, but they can be formed without them. Extrusion has been reported by other SMC complexes, for example, condensin. In fact, chromatin extrusion was first reported in condensin in single-molecule studies (Ganji et al. 2018) Condensin II is enriched at TAD boundaries with TFIIC, (Yuen et al. 2017).

Studying genome folding

A broad range of techniques has uncovered the principles of genome folding during interphase. There are three main categories of techniques; imaging, ligation and ligation-free methods (See Kempfer and Pombo 2019 for a recent review).

Imaging techniques

Imaging techniques can visualise DNA positions in single cells. DNA ISH allows the visualisation of individual loci within fixed nuclei (Gall and Pardue 1969; Pardue and Gall 1969), and its throughput is generally limited to the number of fluorescent labels that can be detected at once, but multiplexing is possible. Multiplexing DNA ISH has revealed single-cell conformations of a human chromosome 21 1.3 Mb region (Bintu et al. 2018; Müller et al. 2002). Traditionally, Nick-translation generates DNA ISH probes by labelling plasmids containing stretches genomic regions, but other strategies based on polymerase chain reaction (PCR) amplification are used (Beliveau et al. 2012).

Live-cell imaging of DNA relies on the DNA recognition of a fluorescent protein. Signal amplification requires multiple binding events and often requires a certain degree of DNA-editing (Lucas et al. 2014). CRISPR technologies have been adapted to visualise DNA. Fluorescent tags can associate to the Cas9 or the guide RNA (Fu et al. 2016; Shao et al. 2016; Wang et al. 2016b) (Figure 1.5A).

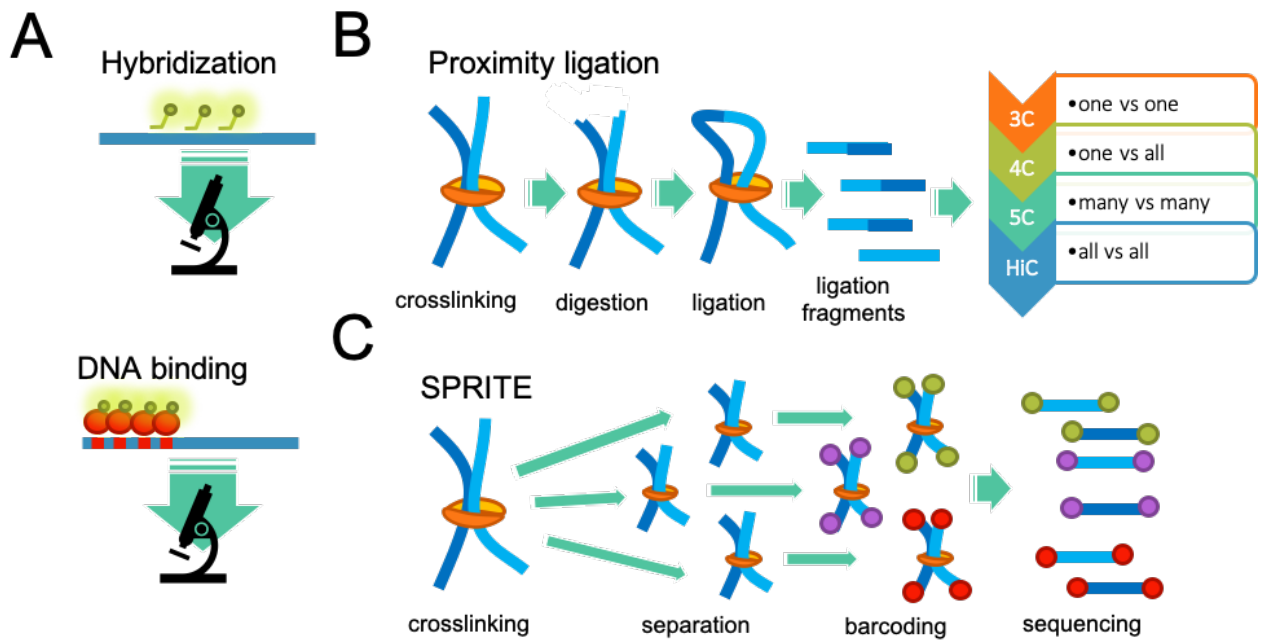


Figure 1.5: **Genome folding techniques.** A) Imaging techniques rely on hybridisation between complementary sequences or a DNA binding protein carrying a fluorophore. B) Proximity ligation techniques are based on cross-linking chromatin, followed by digestion by a restriction enzyme and a subsequent DNA ligation. The probability of ligation increase by their spatial proximity C) Alternatively, in SPRITE, cross-linked chromatin can be sorted and barcoded. After sequencing, cosegregation frequency quantifies sequence contacts

Proximity ligation techniques

Ligation methods rely on the increased probability of DNA ligation between sticky ends that are spatially close. Proximity ligation techniques start with a cross-linking to preserve chromatin interactions. Then, a restriction enzyme digests DNA, leaving sticky ends. Sticky ends hybridise with other sticky ends in their vicinity. Finally, a DNA ligase ligates them and, ligation products are purified and sequenced. Ligation frequency becomes a measure of contacts between two sequences (Figure 1.5B).

Pioneer experiments started in 2002. The first technique was called Chromosome Conformation Capture (3C, Dekker et al. 2002) which quantifies contact frequency between two loci ("one vs one"). Circular Chromosome Conformation Capture (4C, Zhao et al. 2006) identifies all interaction partners of a particular sequence ("one vs all"). Chromosome Conformation Capture Carbon Copy (5C, Dostie et al. 2006) uses designed primers to quantify contacts within a region of interest ("many vs many"). 5C revealed the presence of TADs (Nora et al. 2012).

Hi-C methods (Lieberman-Aiden et al. 2009) are able to quantify genome wide-contacts by purifying and sequencing ligation products ("all vs all") (Figure 1.5B). Since its inception, Hi-C has been optimised. In particular, cross-linking within nuclei reduced noise, achieving 1 kb resolution (in-situ Hi-C, Rao et al. 2014). Combined with chromatin immunoprecipitation (ChIP) it could inform about interactions mediated by a specific protein or complex (Fang et al. 2016; Fullwood et al. 2009; Mumbach et al. 2016).

Hi-C technologies can also be applied to single cells to study rare cell types or DNA folding variations. Digestion, ligation efficiency and ligation fragment retrieval per cell limits Hi-C performance. Besides, one restriction fragment can only lead to two ligation products in a diploid cell (Flyamer et al. 2017; Nagano et al. 2013, 2015). Furthermore, maternal and paternal chromosomes might have distinct structures within a single cell. For example, Dip-C, which use single-nucleotide polymorphisms (SNPs) to differentiate maternal and paternal chromosomes, have studied single-cell conformations within single cells upon imprinting (Tan et al. 2018).

Ligation-free techniques

There are alternative methods to infer chromatin conformation. The GAM technology sequences nuclei laser cryosections. The cosegregation frequency of two sequences are used as a measure of their contact frequency. GAM does not rely on fixation and can detect interactions between many chromatin fragments at once (Beagrie et al. 2017).

In the SPRITE method, cross-linked chromatin complexes are separated into 96-well plates. Then, complexes are barcoded and sequenced. Barcode cosegregation in two sequences indicates a DNA contact (Figure 1.5C). This method can detect long-range interactions that evade ligation technologies (Quinodoz et al. 2018). ChIA-Drop is similar to SPRITE, but integrating microfluidic systems to deliver barcodes (Zheng et al. 2019).

Modulation of genome folding

TAD positions are shared across species and cell-types. However, their contacts can be modulated (Dixon et al. 2012). Differentiation can change compartmentalisation as well as intra-TAD interactions (Beagan et al. 2017; Bonev et al. 2017; Dixon et al. 2015; Hanssen et al. 2017; Kaaij

et al. 2018). Intra-TAD interactions are important for cell-type-specific interactions between promoters and enhancers. Acquisition of active histone marks and CTCF binding correlates with an increase in genomic contacts, whereas the acquisition of repressive marks correlates negatively with contacts (Dixon et al. 2012).

Cells can modulate DNA contacts by altering the recruitment and activity of CTCF and cohesin. For example, during neural differentiation, cohesin unloading factors, *Wapl* and *Pds5a*, decrease their expression, leading to an increase in intra-TAD contacts and a decrease in inter-TAD contacts (Bonev et al. 2017). During the differentiation of pluripotent cells to neural progenitors, *Ctcf* expression decreases. As a result, there is a lower CTCF occupancy around pluripotency genes, accompanied by a loss in interactions between pluripotency genes and their regulatory elements (Beagan et al. 2017). Moreover, imprinting can cause differential 3D genome organisation between homologous chromosomes (Dixon et al. 2012). DNA methylation can alter loop boundaries, as CTCF binds preferentially to unmethylated sequences. An example of this is the expression of *Igf2* from only the paternal chromosome (Hark et al. 2000; Yoon et al. 2007).

Disruption of genome contacts

Mutations that re-shuffle genome sequences are known as structural variations (SVs) and they occur de novo in 5-19% of births (Spielmann et al. 2018). SVs rewire genes and regulatory elements, and they often result in malformations, because they can modify spatiotemporal expression patterns in embryos (Franke et al. 2016; Giorgio et al. 2015; Lupiáñez et al. 2015; Will et al. 2017). The most common SVs are duplications, deletions and inversions. Duplications that involve gene regions are known as copy number variations (CNV) and can lead to gene dosage problems. Intra-TADs SVs that involve non-coding regions can also alter gene dosage by increasing (duplications) or reducing (deletions) the number of regulatory inputs for a gene. Inter-TAD duplications that include boundaries can result in the formation of a new TAD, termed 'neo-TAD'. Cooks syndrome can be caused by a duplication that incorporates a duplicated *Kcnj2* into a neo-TAD. As a result, *Kcnj2* receives input from regulatory elements from its neighbouring gene, *Sox9*. This modifies *Kcnj2* expression pattern to resemble a *Sox9*-like pattern (Franke et al. 2016).

Boundary deletion can cause fusion between TADs (Arzate-Mejía et al. 2020) and adoption of new regulatory elements from the fused TAD (Mateo et al. 2019). It is a common cancer strategy to turn proto-oncogenes into oncogenes (Hnisz et al. 2016). In another example, deletion of CTCF-cohesin sites can extend interactions between a strong enhancer and genes in the adjacent domain, driving their up-regulation (Hanssen et al. 2017). Nevertheless, TAD boundary deletion does not always cause aberrant expression. At the *Sox9-Kcnj2* locus, a boundary deletion results in TAD-fusion with no major regulatory changes (Despang et al. 2019).

In summary, SVs modify DNA contacts, which lead to loss and gain of regulatory inputs. Rewiring of regulatory elements alters gene dosage and spatiotemporal expression patterns. The extent of regulatory changes often determines SV pathogenicity (Lupiañez et al. 2015). However, some studies find a limited effect of gene contacts for the control of gene expression. Balancer chromosomes are highly rearranged fruit fly chromosomes maintained in heterozygosis. They contain multiple SVs, including nested inversion, duplications and deletions. However, chromosome rearrangements did not explained expression changes between WT and balancer chromosomes (Ghavi-Helm et al. 2019).

The regulation of gene expression

Nucleosomes

Chromatin consists of DNA wound around histone octamers called nucleosomes. Their presence and modifications determine which DNA sequences are accessible for recognition and by whom. Nucleosome location, composition and modification is regulated by chromatin remodellers and histone modifiers (Lee and Young 2013) (Figure 1.6).

Nucleosome remodelling

Nucleosome associated DNAs is generally inaccessible by transcription factors (TFs). Remodelling complexes can reposition nucleosomes, and they are often recruited by pioneering TFs to allow DNA recognition by additional TFs and the transcriptional machinery (Figure 1.6A).

Chromatin remodelling complexes are ATP-driven machinery that mobilises and restructures nucleosomes. Their subunit composition determines their nucleosome preference and interaction partners. Generally, they are DNA translocases that pump DNA along histone surfaces. ISWI subfamily spaces nucleosomes regularly, limiting chromatin access. On the other hand, the SWI/SNF subfamily facilitates accessibility by translocating and ejecting nucleosomes. Additionally, other subfamilies like INO80 and CHD are also able to alter the histone octamer composition (Clapier et al. 2017) (Figure 1.6B).

Histone modification

Histones are highly conserved proteins in eukaryotes. They are composed by a globular C-terminal domain which forms part of the nucleosome core and a flexible N-terminal domain that mediates interactions with other complexes and DNA. Post-translational modifications can occur in both domains, mediated by a diverse set of enzymes, including acetylases, methyltransferases, sumoylases, ubiquitinylases, phosphorylases, ADP ribosylases, and deaminases (Lawrence et al. 2016). Enzymes that add post-translational modifications are histone writers and enzymes that remove them are histone erasers (Figure 1.6C).

These modifications can alter DNA-histone interaction and nucleosomal compaction. For example, one essential modification is acetylation. Acetylation is mediated by histone acetyltransferases (HATs) and deacetylation by histone deacetylases (HDACs) which often form part

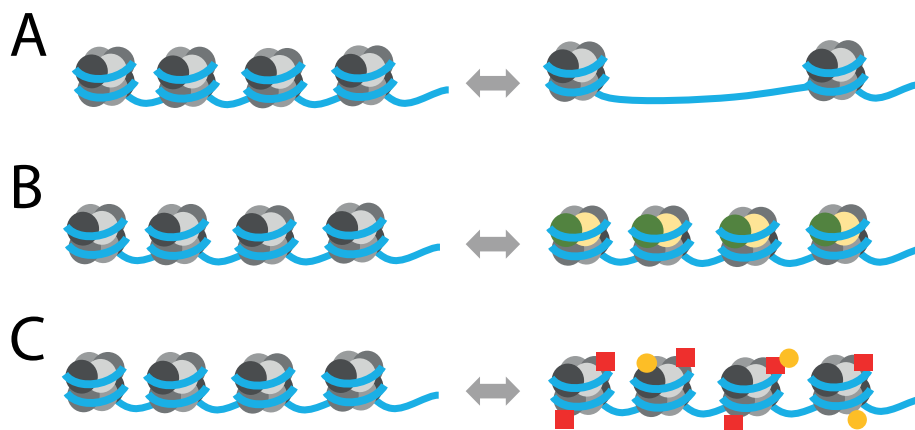


Figure 1.6: **Nucleosome regulation.** A) Nucleosome remodelling complexes can alter the distribution of nucleosomes along the DNA, or B) alter nucleosome subunit composition. C) Histone writers and eraser add and remove postranslational modifications.

of larger complexes. Histones surface has positive charges which interact with the negative charges in the DNA backbone. Histone acetylation neutralises these charges, loosening up this interaction. HATs function as transcriptional coactivators. On the contrary, HDAC action reverses HAT acetylation, and they usually form part of repressor complexes (Bannister and Kouzarides 2011).

Furthermore, different protein post-translational modifications recruit different transcriptional machinery. Histone binding domains, like chromodomains or PHD fingers, recognise different post-translational modifications. Also, distinctive sets of histone marks associate with different functional sequences. For example, H3K9me3 mark is associated with repressed chromatin, whereas H3K27ac is associated with chromatin at cis-regulatory regions like active enhancers (ibid.).

RNA pol II control

Gene expression implies the transcription of DNA into RNA by RNA polymerases. *Escherichia coli* has only one RNA polymerase, whereas higher eukaryotes have three and plants, five. Each RNA polymerase synthesises RNAs with particular functions. RNA pol II synthesises capped RNAs, like mRNAs that can be translated into protein by ribosomes. Enhancers and promoters control where and when transcription by RNA pol II occurs. The general transcription machinery is recruited at promoters, whereas enhancers recruit different sets of specific TFs depending on the cellular context, and their interactions are dictated by their distance in 3D, not by their distance in the linear sequence. TFs affect transcription indirectly by associating with general co-activators and co-repressors, like Mediator, p300 or histone modifiers (Lee and Young 2013). Expression of new TFs can propagate and amplify cell-state transitions (Bhatt et al. 2012).

RNA pol II recruitment

RNA pol II recruitment depends on pre-initiation complex (PIC) assembly (Gupta et al. 2016) (Figure 1.7A). This megadalton complex directs RNA pol II to gene promoters. It is composed of general transcription factors (GTFs), RNA pol II and Mediator.

GTFs are TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. TATA-box binding protein (TBP) and TBP associated proteins (TAFs) form TFIID. TBP is half-moon shaped and recognises promoter elements. TAFs are interchangeable and modulate TBP binding. TAFs hold promoter-specific functions. TFIIB also binds to promoter elements, and it is involved in Transcription Start Site (TSS) selection and transcription orientation. TFIIA assists TFIIB and TFIID, facilitating their interaction with DNA. Then, RNA pol II is recruited to the PIC by TFIIF. To initiate transcription, TFIIH needs to be recruited by TFIIE. TFIIH has a helicase and a kinase module. The helicase module allows RNA pol II access to single-strand DNA (ssDNA) using ATP to open the transcription bubble. The kinase module phosphorylates RNA pol II's carboxy-terminal domain (CTD) at serine 5 position, a sign of transcription initiation (Sainsbury et al. 2015) (Figure 1.7B).

The Mediator complex is a large coactivator present in eukaryotic cells. Signal-specific TFs recruit Mediator at enhancers and cohesin enables Mediator interaction with its target promoters, which are transient (El Khattabi et al. 2019; Jaeger et al. 2020; Soutourina 2018). Mediator connects the context-dependent regulation with the general transcription machinery. Once bound, Mediator assists the assembly of the general transcription machinery at promoters and facilitates transcriptional elongation. Mediator depletion activates compensatory mechanisms in mammalian cells that reduce RNA pol II pausing but fails to rescue cell-type-specific expression (Jaeger et al. 2020).

RNA pol II pause release

Once TFIIH opens the transcription bubble, transcription can start, and RNA pol II escapes PIC. However, without additional regulatory steps, RNA pol II can only transcribe 20-50 bp before pausing (Figure 1.7C). Productive elongation requires an additional checkpoint denominated pause release. Pausing blocks productive expression from a gene (Adelman and Lis 2012; Core and Adelman 2019). RNA pol II recruitment occurs faster than RNA pol II release, and cells can regulate pausing timing (Bartman et al. 2019). Genes which have to activate their expression rapidly in response to stimuli tend to have paused RNA pol II at their promoters.

CTD post-translational modifications couple to different stages of the transcriptional cycle. The CTD is a disordered 7-aminoacid repetitive sequence at the C-terminal of the RNA pol II

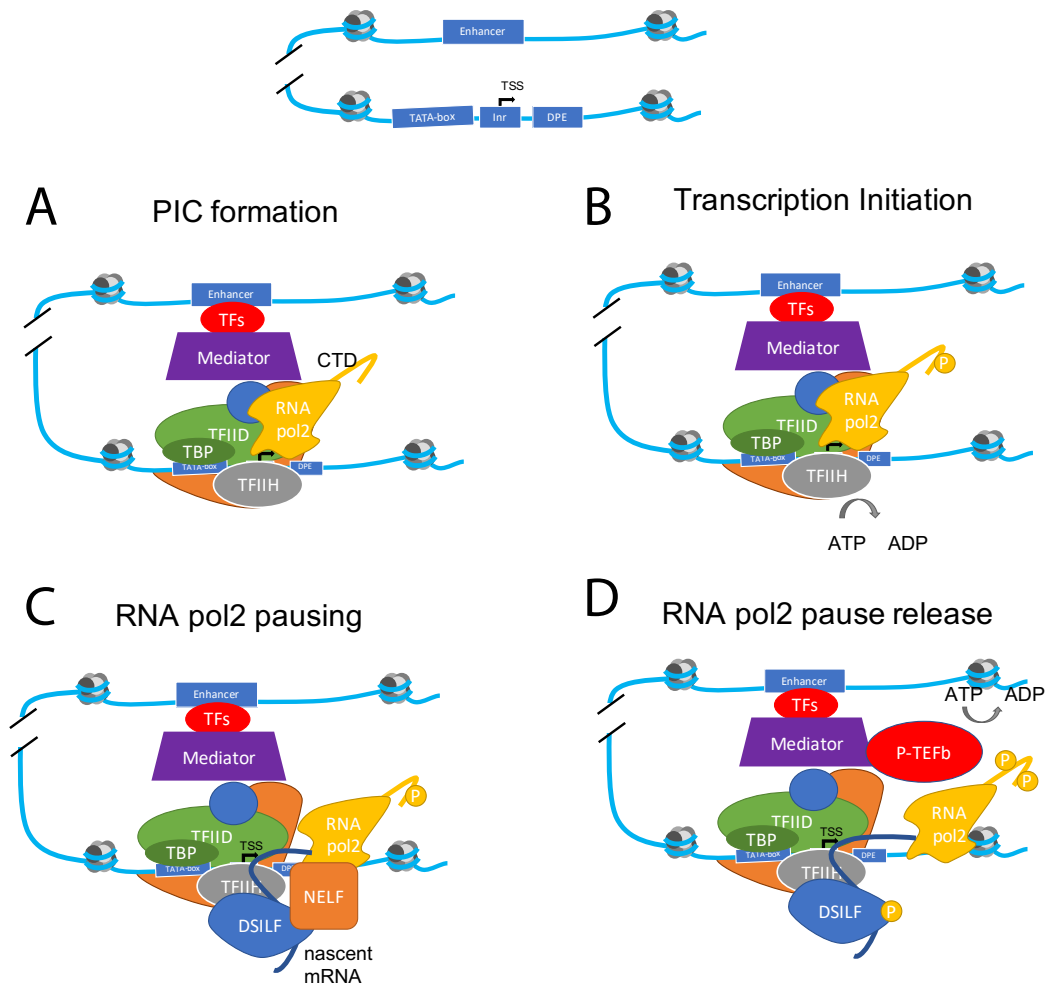


Figure 1.7: **Transcription initiation.** A) Recruitment of RNA pol II to gene promoters depends on the formation of the PIC. B) Transcription initiation is triggered by TFIIF. TFIIF opens the transcription bubble and phosphorylates RNA pol II at its CTD domain. C) RNA pol II transcribes few nucleotides downstream TSS before pausing. Paused RNA pol II is established by NELF, recruited by DSILF. D) RNA pol II pause release occurs when DSILF is phosphorylated by pTEFb, leading to a conformational change that releases NELF.

Rpb1 subunit. Phosphorylation of Serine 5 by TFIIF facilitates promoter escape and recruits DSILF, a heterodimer of *Spt6* and *Spt5*, which mediates cap formation at the nascent RNA 5' end. DSILF recruits the negative elongation factor (NELF) complex, present in animals (Core and Adelman 2019). This factor stabilises paused RNA pol II by restraining its movement and by altering interactions at its active centre (Vos et al. 2018) (Figure 1.7C). Loss of NELF does not only affect pausing, but it also affects PIC formation stability, allowing nucleosomes to take over TSSs (Core and Adelman 2019).

The positive transcription elongation factor b (p-TEFb) releases paused RNA pol II. The p-TEFb complex contains a CDK9 kinase subunit that can phosphorylate DSILF *Spt5* subunit and RNA pol II's CTD. *Spt5* phosphorylation triggers a conformational change that dissociates NELF and turns *Spt5* into a recruitment platform for RNA-processing and positive elongation factors (Figure 1.7D). RNA pol II CTD phosphorylation by CDK9 occurs at Serine 2 residues, and can only occur after TFIIF phosphorylation. Sequential CTD modifications ensure the correct succession of transcription processing events (Core and Adelman 2019).

p-TEFb often forms part of larger complexes such as 7SK or the Super Elongation Complex (SEC). p-TEFb is recruited by Mediator and coactivators, like BRD4, but it is also targeted by TFs to modify gene expression, like AIRE, c-Myc or the HIV *Tat* protein (Adelman and Lis 2012; Core and Adelman 2019; Rahl et al. 2010). BRD4 can recruit p-TEFb to acetylated histones (Core and Adelman 2019). Specialised thymic stromal cells need to present self-antigens to maturing T-cells to promote self-tolerance. To do so, they express AIRE, which recruits p-TEFb at different promoters, expanding their expression repertoire (Adelman and Lis 2012). It also forms part of a transcription safeguard mechanism, as Mediator depletion activates a feedback loop that increases *Cdk9* expression, leading to a decrease in RNA pol II release rates (Jaeger et al. 2020).

Functional elements

The classic genomic regulatory elements are promoters and enhancers. In summary, promoters are the start sites of gene transcription, and enhancers are regulatory regions that modulate transcription from promoters.

Transcription can occur bidirectionally at enhancers and promoters (De Santa et al. 2010; Kim et al. 2010), but at promoters, sense transcripts are stabilised. Still, the distinction of promoters and enhancers is not well defined as they can have interchangeable functions. For example, intragenic enhancers can act as alternative TSSs; some enhancers can act as weak promoters and some promoters as strong enhancers (Kowalczyk et al. 2012; Mikhaylichenko et al. 2018). Therefore, functional sequences have various degrees of enhancer or promoter potential (Andersson and Sandelin 2019). What distinguishes an mRNA from an enhancer RNA (eRNA) production is the presence of RNA pol II elongation factors (Sartorelli and

Lauberth 2020). It is thought that pause release control failure destabilises transcribing RNA pol II, causing eRNA to be shorter (Core and Adelman 2019).

Promoter regions

PIC assembly occurs at the 'core promoter' and determines TSS selection and transcription direction. Even though there are some common elements, promoter sequences are diverse, and not all promoters contain all elements (Andersson and Sandelin 2019). Different promoter architectures use different strategies to assemble the PIC.

Promoter elements include initiator (Inr), the TATA-box, TFIIB recognition element (BRE) or the downstream core promoter element (DPE) (Kadonaga 2012). Inr is a small consensus sequence that covers the TSS. TATA-box and BRE sequences are upstream and serve as recognition and assembly sites for TFIID and TFIIB. TATA-box is recognised by the half-moon subunit TBP (Sainsbury et al. 2015). TAF6 and TAF9 TFIID subunits recognise DPE, located downstream the TSS.

Enhancers

Enhancers are cis-regulatory DNA sequences that can activate or repress the expression of their target genes independently of their orientation. They can regulate more than one gene, and more genes can be regulated by more than one enhancer. They can be upstream or downstream their target gene or even in intronic regions. They are responsible for the context-dependent control of gene expression (Schoenfelder and Fraser 2019). They act as recruitment sites of RNA pol II, Mediator, and specific TFs (Koch et al. 2011). Hence, enhancers can operate by the locally concentrating transcription machinery, enabling chromatin accessibility, PIC assembly and recruitment of co-activators. (Pennacchio et al. 2013).

Enhancers are identified by a repertoire of methods, including sequence conservation (Pennacchio et al. 2006), DNase I sensitivity sites (Dorschner et al. 2004), and presence of p300 (Visel et al. 2009), H3K27ac (Creyghton et al. 2010) and H3K4me1 (Heintzman et al. 2007). GROseq bidirectional transcription sites can also be used to identify them (Melgar et al. 2011). However,

not all enhancer sequences are marked by all these methods, and not all sequences marked as enhancers regulate genes.

Most enhancers are redundant, and a gene is more often control by more than one enhancer. Redundancy prevents loss of regulatory control due to mutations. Often, an observable phenotype requires the disruption of multiple enhancers (Osterwalder et al. 2018). Additionally, redundant inputs can be a substrate for the evolution of new regulatory patterns. More often, disease enhancer-related mutations are gain-of-function or alter gene-enhancer distances. An example is the case of *Ihh*. This gene regulates bone formation, and its misregulation causes developmental malformations. For example, *Ihh* enhancer duplication events lead to gene up-regulation and tissue misexpression (Will et al. 2017). Enhancers can also behave synergically and have additive effects (Perry et al. 2011).

Genomic regions where enhancers cluster are known as super-enhancers (SEs). They can drive higher levels of expression than a typical enhancers. TFs and transcriptional machinery densely populate SE, including cohesin and NIBLP. SEs are also susceptible to perturbation. They regulate cell-type-specific processes and are acquired by cancer cells to promote oncogene expression (Hnisz et al. 2013; Whyte et al. 2013). SE can generate phase-separated condensates, 3D membraneless structures which concentrate chemically compatible elements and can compartmentalise biochemical reactions.

Enhancer transcription

Transcription is associated with enhancer activity. Enhancer transcripts have been shown to precede their regulated gene transcripts in mouse primary neurons (Kim et al. 2010), macrophages (De Santa et al. 2010; Kaikkonen et al. 2013) and *Drosophila* (Andersson et al. 2014). Enhancer transcription is important to maintain an open chromatin that is accessible to TFs (Kaikkonen et al. 2013).

Short non-coding RNAs (ncRNAs) synthesised at open and active enhancers are denominated eRNAs. They are produced by the RNA pol II, capped, retained in the nuclei, and generally not polyadenylated nor spliced. Although eRNA production is widespread, it is inconclusive whether they carry general regulatory functions (Sartorelli and Lauberth 2020).

A variety of specific functions have been assigned to particular eRNAs. Typical eRNAs are

unstable and they often work in cis, by recruiting Mediator (Lai et al. 2013), HATs (Bose et al. 2017), BRD4 (Rahnamoun et al. 2018), or other transcription factors like YY1 (Sigova et al. 2015). eRNAs can also promote RNA pol2 pause release by NELF sequestration (Schaukowitch et al. 2014) and can be involved in enhancer-promoter looping (Lai et al. 2013; Li et al. 2013). Long polyadenylated eRNAs can regulate in trans, as they have a longer half-lives. As an example, an eRNA from the MyoD enhancer region (chromosome 1) translocates to the *Myogenin* promoter (chromosome 7) to regulate its activation during muscle differentiation (Tsai et al. 2018).

Enhancers and chromatin loops

Chromatin architecture governs promoter-enhancer interactions by blocking or facilitating their contacts. They can be preexistent or can be generated de novo upon stimuli (Schoenfelder and Fraser 2019).

CTCF can act as an insulator, blocking enhancer-promoter interactions. For example, the imprinting control region (ICR) contains several CTCF binding sites. It is located between the *H19* and *Igf2*. Paternal ICR is methylated, and CTCF can only bind the unmethylated maternal copy. This binding creates a series of chromatin loops and impedes communication between an *H19*-proximal enhancer and *Igf2*. As a consequence, *Igf2* is only expressed from the paternal chromosome (Hark et al. 2000; Yoon et al. 2007). Often, deletion of a CTCF boundary leads to the adoption of new regulatory elements by a given gene (Giorgio et al. 2015; Hanssen et al. 2017; Mateo et al. 2019). In *Escherichia coli*, lac operon looping associates to gene repression by increasing locally repressor concentration (Becker et al. 2013).

Chromatin loops can facilitate productive contacts between enhancer and promoters within and outside the loop. *Scd2* promoter-enhancer distance is shorter in mouse fetal liver cells (where it is expressed) compared to other surrounding cell-types (Liu et al. 2020). Artificial tethering of an enhancer and a promoter results in gene activation (Bartman et al. 2016; Deng et al. 2012), and loop formation can bring together regulatory regions in the adjacent domains, an effect described in *Drosophila* as domain skipping (Yokoshi et al. 2020). Furthermore, the ring-shaped cohesin complex bridge Mediator to target promoters (Jaeger et al. 2020; Kagey et al. 2010). Cohesin depletion does not lead to major changes in histone acetylation, but genes

in the enhancer vicinity tend to be DE (Cuartero et al. 2018; Ing-Simmons et al. 2015). eRNAs can interact with cohesin and contribute to looping (Li et al. 2013; Tsai et al. 2018).

YY1 is another architectural protein involved in promoter-enhancer interactions. Similar to CTCF, it is zinc finger protein that can form homodimers, essential for cell viability. However, whereas CTCF preferentially binds insulator sequences, YY1 preferentially binds enhancers and promoters in mammalian cells (Weintraub et al. 2017). In neural progenitors, it cooperates with a preexisting CTCF network to mediate cell-specific promoter-enhancer interactions (Beagan et al. 2017).

Measuring gene expression in individual cells

Bulk population methods measure differences in means. However, two very different distribution shapes can share the same mean. Intracellular processes shape gene expression distributions. Therefore, to understand biological mechanisms is vital to characterise what happens in individual cells. Furthermore, single-cell studies avoid the Simpson paradox, which occurs when two populations that have two different distributions are considered as one, and therefore, any measurement performed in the sum of both would not represent either.

In Situ Hybridization techniques

The first ISH experiments were performed in 1968 by Joseph G. Gall and his student Mary Lou Pardue. Initially, ISH was based on sequence complementarity between RNA probes and a target DNA. In a world without cloning, the most available RNAs probes were the most common RNAs, the ribosomal RNAs. Gall and Pardue were studying the ribosomal DNA amplification during meiosis in *Xenopus* oocytes. The first probes were ^3H -labeled ribosomal RNA (Gall and Pardue 1969) and ^3H -labeled ribosomal DNA (Pardue and Gall 1969). A year later, they applied the same technique to visualise mouse satellite DNA (Pardue and Gall 1970). The detection method of choice was radioactivity because fluorescent microscopy was in its infancy.

Fluorescence was first introduced indirectly, via an antibody that recognised RNA-DNA hybrids (Rudkin and Stollar 1977). The first cloning experiments were published in 1972 (Jackson et al.

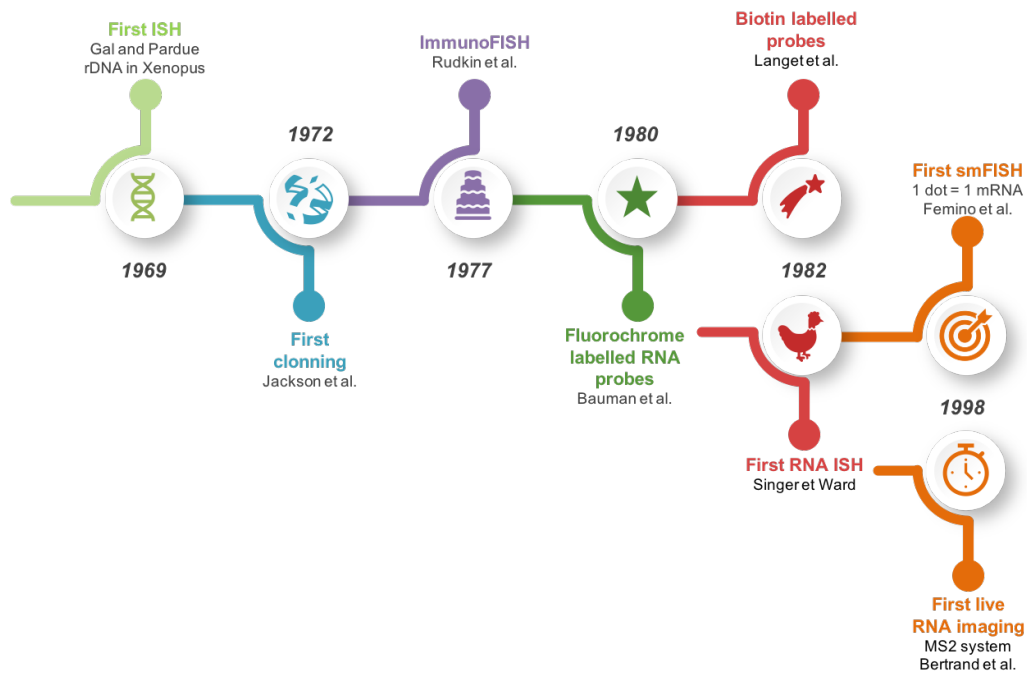


Figure 1.8: **Brief history of RNA visualisation.** ISH was developed for the visualisation of DNA sequences. First experiments relied on radioactive detection, which was then replaced for fluorescent methods. Improvements related to probe cloning and detection allowed the visualisation of individual RNA molecules. Protein recognition of inserted RNA loops allowed live-cell imaging of RNA.

1972), but it was not until the end of the decade when plasmid inserted sequences gave more flexibility to probe production. Soon, fluorochrome labelled RNA probes were used (Bauman et al. 1980) followed by biotinylated RNA probes (Langer-Safer et al. 1982). An overview of the beginnings of ISH is described in Gall 2016 (Figure 1.8).

Single-molecule fluorescence in situ hybridisation

The first attempt to visualise RNA was published in 1982 where they studies the expression of chicken *Actin*. However, images were unquantifiable, and they were only allowed to distinguish cell morphology (Singer and Ward 1982). Distinguishing individual molecules came later in 1998 when Femino et al. used several probes against different segments of the *Actb* gene to calculate hybridisation efficiency in a rat cell line. They also used intronic probes to visualise nascent transcripts at TSSs (Femino et al. 1998), smFISH was born.

Recent advances include the use of sequential hybridisations to increase the signal to noise ratio (Battich et al. 2013) and multiplexing. Multiplexing techniques can be applied to DNA

(Nagano et al. 2013), RNA (Chen et al. 2015) or both (Su et al. 2020). Multiplexing requires sequence barcoding and several cycles of hybridisation, imaging and label removal. Each round has a different set of probes that hybridise to different sequences and light up in different colour. Then, an image analysis software will decode each spot identity based on their colour pattern through the rounds.

Live cell imaging of gene expression

All ISH experiments are cross-sectional studies with no time dimension. The MS2 system allowed live-cell imaging of gene expression. This system is based on the recognition of a viral coat protein of RNA loops. It requires two constructions, one containing the MCP coat protein coupled to a fluorescent protein and the insertion of MS2 RNA loops inserted on the gene 3'untranslated region (3'UTR). Transcript expression recruits a chimeric MCP to its RNA loops, leading to the appearance of a fluorescent dot. Initially, the system was used to track particles in budding yeast (Bertrand et al. 1998). As an alternative, the bacteriophage PCP coat protein and its recognised PP7 RNA loops can be used (Chao et al. 2008). The MS2 and the PP7 system can be combined to track the expression of two genes simultaneously (Chen et al. 2018; Fukaya et al. 2016). Another strategy to visualise expression in live cells consists of RNA aptamers. Aptamers are RNA constructs that bind synthetic dyes which are brighter and smaller than their biological protein counterparts, and they tend to have food names such as spinach or mango (Dolgosheina et al. 2014; Paige et al. 2011).

Other approaches to visualise gene expression were based on destabilised protein reporters such as luciferase or GFP (Nicolas et al. 2018; Ochiai et al. 2019; Suter et al. 2011). However, signal intensity was the outcome of several cellular processes including transcription, mRNA nuclear export, translation and mRNA half-life. Nowadays they come accompanied by more direct measures of burst such as smFISH or the MS2 system.

One drawback of this live technique is that they all require a certain degree of gene engineering, which limits their application.

Single cell RNA sequencing techniques

Next-generation sequencing technologies have been adapted to single cells in the last decade. By dissecting cell population complexity, they can reveal rare populations of cells, cell-state transitions (Trapnell et al. 2014), regulatory relationships (Chan et al. 2017) and developmental trajectories (Setty et al. 2016). Specialised protocols and analysis pipelines can identify allele-specific expression (Larsson et al. 2019; Ochiai et al. 2019) and predict a cell’s future state based on the ratio of intronic to exonic reads (La Manno et al. 2018).

The first step in scRNAseq consists of cell isolation. Cell isolation techniques include limiting dilution, FACS, laser capture microdissection (LCM), and microfluidics (Chen et al. 2019a). The next steps are cell lysis and reverse transcription. During reverse transcription, each read is ligated to a cell-specific barcode. These barcodes allow identifying all the transcripts coming from a certain cell, even when sequencing is performed in bulk. The current recommendation is that the sequencing depth per cell should be enough that each expressed gene has at least one read (Zhang et al. 2020).

Similarly, to control for PCR amplification bias, each transcript within a cell can be barcoded with universal molecular identifiers unique molecular identifiers (UMI). However, UMI barcodes insertion occurs in one end of mRNAs, so are not used in situations where intronic information is needed, for example, to quantify nascent transcription or splicing isoforms.

Challenges in the analysis of scRNAseq data include matrix sparsity, the presence of latent variables and the curse of dimensionality. Latent variables are biological variables that are not considered interesting for a particular experiment, such as differences regarding cell cycle, and can be corrected (Buettner et al. 2015). The curse of dimensionality affects data points represented by high number variables, such as scRNAseq where the expression of several genes is measured in each cell. As you increase the number of variables, the available space grows exponentially while the distances between points do not, requiring more data points to populate the space densely enough and identify significant differences. Reducing the number of variables can be done using Principal Component Analysis (PCA) or feature selection (Hwang et al. 2018).

In contrast to imaging techniques, scRNAseq implementation is more straightforward and scalable. Moreover, they can provide a higher-dimensional picture of gene expression and can po-

tentially detect any transcript in a cell, whereas pre-designed probes introduce bias in smFISH. However, on the downside, scRNAseq have a higher rate of dropout events, and multiplexed smFISH is catching up in terms of dimensionality (the highest number of different genes detected by multiplexed smFISH is 1,137 (Su et al. 2020) whereas, there is no potential limit for scRNAseq). Nevertheless, the use of very specialised equipment limits multiplex smFISH application to a limited number of labs.

Cell-to-cell variability

Cell-to-cell variability is referred to as gene expression noise. Noise has two components, intrinsic and extrinsic noise. Intrinsic noise is inherent to the low number of molecules present in a cell (the number of gene copies ranges from two to four, for example) and it shapes population distributions depending on how transcripts are produced (Munsky et al. 2012). On the other hand, extrinsic noise arises from baseline differences between cells that impact their chemical reactions rates (differences in the number of RNA pol II, for example). The sum of both components shapes gene expression distributions.

Attempts to estimate both types of noise from mRNA distributions have been made (Lenive et al. 2016). In some cases, sources of extrinsic noise can be identified, such as cell cycle, common cell ancestry, cell localisation or cellular components (Snijder and Pelkmans 2011). The preexisting variation between cells may be advantageous in some contexts. For example, increasing the number of possible phenotypes can increase possibilities of tumour cells to resist to an anticancer drug. In another example, HPSC diversity bias lineage choice (Haas et al. 2018).

Variability can be analysed from multiple perspectives that can be cell-centric or gene-centric. Cell-centric analyses may treat each cell as an individual, or compare individual cells to each other. Analyses that treat cells as individuals study how individual cells distribute cellular resources (Figure 1.9A). To compare cells to each other, each cell can be defined as a vector with all its gene expression values. In this respect, cells can be considered as multidimensional points. Therefore, the similarity between cells can be assessed as to their distance on the multidimensional expression space or in a dimensionality reduced space (Figure 1.9B). These approaches are often combined with clustering methods like k-means or k-nearest-neighbours

to identify subpopulations of cells, differentiation trajectories or branching points during development (Setty et al. 2016; Trapnell et al. 2014).

On the other hand, gene-centric analyses have gene distributions as main characters. They use variability measures to characterise mRNA distribution shapes. Most distribution variability measures are mean dependent in discrete setups, like the coefficient of variation or the FANO factor. Eling et al. 2018 developed a metric called overdispersion that corrects from mean dependence allowing comparisons between distributions independent of whether a gene has the same mean expression in both populations or not (Figure 1.9D). When considering all the overdispersion in all genes (Figure 1.9C), we can access the general patterns of expression distributions between two populations. Using this approach, Martinez-Jimenez et al. 2017 identified an increase in gene overdispersions during immune responses of ageing mice. Cross-comparisons between gene distributions have also been used to identify gene regulatory circuits using, for example, information theory measures (Chan et al. 2017).

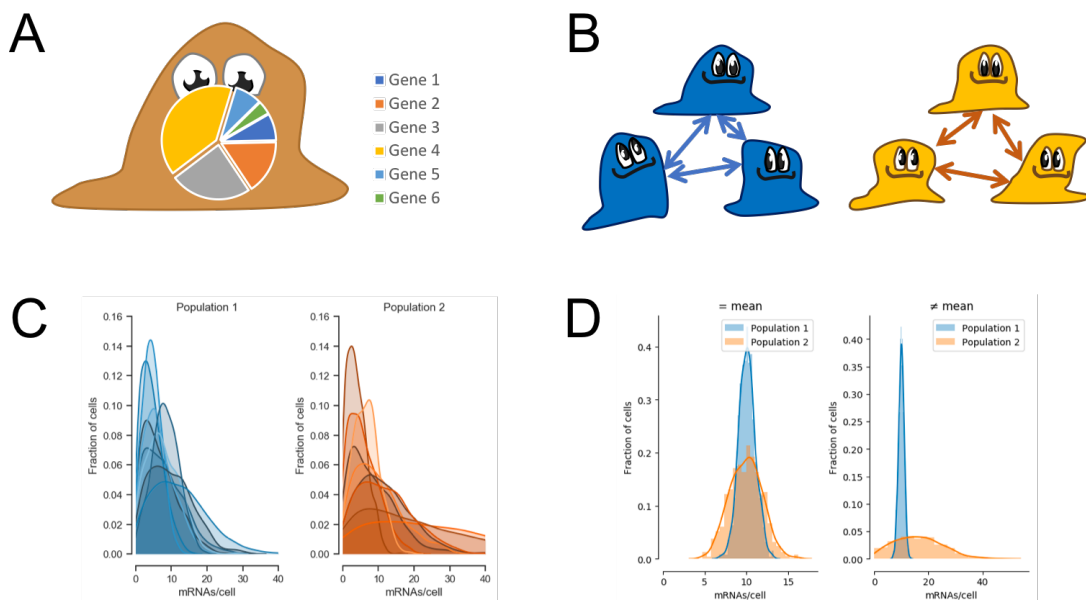


Figure 1.9: **Perspectives on cell-to-cell variability.** A) Distribution of cellular resources in individual cells. B) Quantification of expression differences as distances among a cells. C) Analysis on all gene distribution. D) Analysis on the distribution of a single gene.

Transcriptional bursts

Gene expression involves the transcription of individual mRNA molecules. As the total number of mRNA molecules can be counted, gene expression in a population of cells is discrete. However, transcription is not continuous. If gene expression were continuous, then transcription would randomly generate mRNA at a certain rate, following a Poisson distribution. Although certain gene distributions look Poisson (Zenklusen et al. 2008), in most cases, transcription occurs in a firework-like fashion where periods of mRNA production are interspersed. Transcriptional bursts have been observed from bacteria to intact mammalian tissues (Chubb et al. 2006; Golding et al. 2005; Halpern et al. 2015; Raj et al. 2006a). These transcription pulses are a major cause of intrinsic noise, as they involve a low number of molecules (two alleles in a diploid cell and the number of available RNA polymerases, for example), causing dramatic fluctuations within a single cell and leading to broader gene expression distributions within the population. (Raj and Van Oudenaarden 2008). Cell compartmentalisation, ploidy and certain gene network regulatory motifs mitigate noise fluctuations (Alon 2007; Halpern et al. 2015; Stoeger et al. 2016).

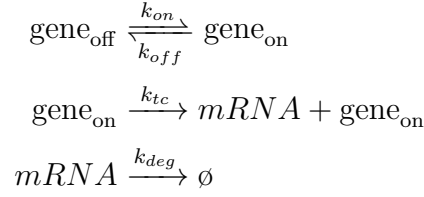
In bacteria, bursting phenomena are associated with positive torsion accumulation caused by elongating polymerases. This tension inhibits transcription, and it is alleviated by DNA gyrase excision activity. Increasing gyrase activity makes gene expression look Poisson. Poisson expression distributions are shaped by one constant, the transcription rate. Therefore, if expression distributions look Poisson, then transcription does no longer occur in bursts (Chong et al. 2014).

Modelling transcriptional bursts

Several mathematical models have been proposed to explain transcriptional bursting. Some of them include multiple regulatory inputs (Tunnacliffe and Chubb 2020). However, increased complexity tends to lead to data overfitting. For this reason, the most commonly used is the telegraph model, which is the simplest mathematical model that explains discontinuous periods of expression (Peccoud and Ycart 1995; Raj et al. 2006a).

The telegraph model is a two-state model. Here, the promoter of a gene can switch between an on and an off state with rates k_{on} and k_{off} . When a gene is on, it can produce mRNA

molecules at a rate k_{tc} . Then, those mRNA molecules are degraded at a rate k_{deg} .



Where $\text{gene}_{\text{on}} + \text{gene}_{\text{off}} = n$, and n is the total number of alleles, the cell ploidy.

Although the telegraph model is useful to explain cell-to-cell variability, it does not account for all aspects of regulation. In Bartman et al. 2019, they added RNA polymerase recruitment and pause release to the telegraph model. According to their definition, burst is a state, that does not necessarily lead to the production of nascent mRNA molecules. The authors discern between several mechanistic models by combining several experimental techniques, RNA pol II elongating simulations and system perturbations by the use of drugs. As a result, they conclude that RNA pol II recruitment follows burst initiation and that during erythroid differentiation, cells modulate RNA pol II pause release and burst initiation.

Control of transcriptional bursts

Cross-sectional studies measure two transcriptional burst properties. Burst frequency is how often alleles burst and, burst size is the average number of mRNA molecules produced by a burst. Bursting dynamics are gene-specific (Suter et al. 2011), and most studies focus on how transcriptional burst frequency and size are affected under different perturbations.

Positive DNA torsion caused by transcription-associated DNA unwinding can inhibit transcription and generate bursty expression (Chong et al. 2014). Even though it probably contributes to bursting expression in higher organisms, positive torsion accumulation is unlikely to be the only responsible mechanism for bursts. An accepted view is that core promoter elements control burst size, whereas enhancers control burst frequency (Larsson et al. 2019). Notch enhancers are an exception, which regulate burst size during *Drosophila* development (Falo-Sanjuan et al. 2019; Lee et al. 2019).

Bigger burst sizes are associated with an increased number of TF binding sites (Raj et al. 2006a) and the presence of TATA boxes (Donovan et al. 2019). TF binding can trigger burst initiation. In living yeast, Single-Molecule Tracking (SMT) studies have shown a correlation between stable GAL4 binding and the beginning of a new burst (ibid.). TF dwell times correlate with burst sizes, and the fraction of bound receptors is related to burst frequency in cells responding to estrogen signalling (Donovan et al. 2019; Stavreva et al. 2019)

Enhancers drive context-dependent gene regulation by modulating burst frequency (Larsson et al. 2019). Enhancers were first linked to burst frequency in 1995 (Walters et al. 1995). Burst frequency modulation mediates sensitive, and rapid transitions between gene expression states (Li et al. 2018), and cell-type-specific programs change mainly burst frequencies (Larsson et al. 2019). For example, galactose signalling increases the burst frequency of GAL4 target genes in yeast (Donovan et al. 2019). Histone acetylation modulates burst frequency during circadian rhythm (Nicolas et al. 2018) and neural activation (Chen et al. 2019b). promoter acetylation correlates with the fraction of cells expressing a transcript in dendritic cells (Shalek et al. 2014). Genes located in the A compartment show higher burst frequencies but similar burst sizes to genes located in the B compartment (Su et al. 2020).

Most studies show that promoter-enhancer contacts direct burst initiation (Bartman et al. 2019, 2016; Chen et al. 2018; Deng et al. 2012; Fukaya et al. 2016), although enhancer activation of *Sox2* does not rely on spatial proximity (Alexander et al. 2019). Promoter-enhancer distances can modulate burst frequency and size (Yokoshi et al. 2020). When two genes share an enhancer and their genomic distances are similar, their bursts coordinate. If one of these distances increases, the gene that is closer to the promoter will exhibit an increased burst frequency. Insulator presence reduces this crosstalk (Fukaya et al. 2016), and coregulated genes tend to co-segregate (Schoenfelder et al. 2010). Expression changes modulated by stimulus preferentially regulate burst frequencies (Bartman et al. 2016; Fritzsche et al. 2018; Fukaya et al. 2016; Larson et al. 2013; Molina et al. 2013; Senecal et al. 2014), because it enables fast and sensitive responses (Li et al. 2018).

BET proteins regulate burst initiation (Bartman et al. 2019) and pol2 release (Bartholomeeusen et al. 2012; Winter et al. 2017; Yang et al. 2005). Elongation factors associate to high burst frequencies, whereas PRC2 associates to genes with big burst sizes but low frequencies (Ochiai et al. 2019). However, this association is context-dependent.

In summary, burst size associates with the strength of RNA pol II promoter recruitment, whereas regulatory inputs preferentially regulate burst frequency. Variability increases as burst size increase, and frequency becomes smaller. Therefore, gene activation is associated with a noise reduction, whereas gene deactivation increases it.

The macrophage antimicrobial response as a model

Bone marrow-derived Macrophages

Macrophages are innate immune cells, involved in the first line of defence against infection, antigen presentation, tissue-regeneration and homeostasis. They orchestrate inflammation by releasing factors that recruit and modulate immune responses of other cell-types. Macrophages are a heterogeneous group of professional phagocytes. They have two possible origins: fetal liver and circulating monocytes (Varol et al. 2015). During development, the first wave of macrophages migrates from the fetal liver to their destination tissue. In their destination tissue, fetal macrophages adapt locally and become tissue-resident macrophages. Tissue-resident macrophages populations persist to adulthood through self-renewal. Monocyte-derived macrophages are short-lived and quiescent. Their regeneration relies on the self-renewal and differentiation of HPSCs. Monocyte-derived macrophages complement the function of tissue-resident macrophages on demand, for example, in the event of an infection.

In this thesis, we use bone marrow-derived macrophages. Macrophage colony-stimulating factor (M-CSF) promotes macrophage commitment. Primary bone marrow cells cultured for seven days in M-CSF enriched media generate macrophages. These cells resemble monocyte-derived macrophages (Davis 2019), are quiescent, and provide a cell cycle-independent system to study the impact of cohesin on gene regulation.

The antimicrobial response

Innate immune cells have an extensive battery of pattern recognition receptors that recognise a repertoire of pathogen associated molecular patterns (PAMPs). One of these PAMPs is LPS), a complex glycolipid present in gram - bacteria cell wall, which is recognised by the toll-like

receptor (TLR) 4. TLRs are transmembrane receptor characterised by an extracellular Leucine-Rich Repeat domain that binds PAMPs and cytosolic Toll/Interleukin-1 Receptor homology domain that interacts with intracellular signalling scaffolds. The TLR family is diverse across species, and individual TLRs are involved in a variety of signalling pathways (Fitzgerald and Kagan 2020).

LBP and CD14 facilitate TLR4 interaction with LPS. LBP is a soluble extracellular protein that can extract single molecules of LPS from microdroplets. An LBP-LPS complex associates with CD14, a TLR4 co-receptor. From CD14, the LPS molecule is transferred to MD2. Then, MD2-LPS binds to TLR4, leading to its TLR4-MD4-LPS heterodimerisation and dimerisation of TLR4 intracellular domains. (Gioannini et al. 2004; Park et al. 2009; Poltorak et al. 1998). The TLR4 dimerised TIR domains are recognised by a set of scaffold proteins, such as TIRAP/MAL and TRAM. Scaffold association leads to the rapid assembly of the supramolecular organising centre (SMOC)(Fitzgerald and Kagan 2020). SMOCs composition is context-dependent, increasing response repertoire and specificity. SMOCs assembly success imposes a threshold for signalling pathway activation(Kagan et al. 2014). Protein cascades involving protein recruitment and post-translational modifications mediate the signal amplification. The protein cascade leads to the eventual modification of nuclear TF activity. For example, TLR4 activation leads to the nuclear translocation of transcription factors such as AP-1, $\text{NF}\kappa\beta$ or IRFs. As a result, macrophages activate and change their gene expression.

Gene expression changes can be classified by their dependence on de novo protein synthesis or by their timing. In the first classification, primary response genes do not rely on de novo protein synthesis, whereas secondary response genes do. Thus, their expression occurs as a consequence of the translation of primary response genes (Herschman 1991). Bhatt et al. 2012 classified immune response genes based on their expression timing and subcellular association. Very early response genes have CpG islands, and RNA pol II paused at their promoter in unstimulated macrophages. 10 out of 16 of the genes induced after 15 min of stimulation were transcription factors that control subsequent expression changes. Late response genes often display large fold-changes and require additional steps such as nucleosome remodelling or RNA pol II recruitment.

Antimicrobial signals transform the enhancer landscape. In the context of macrophage response to LPS, enhancers are classified as constitutive, poised, latent or repressed depending

on their accessibility and histone mark dynamics (Ostuni et al. 2013). Constitutive enhancers have H3K4me1 and H3K27Ac histone mark and are accessible without stimulation. Poised enhancers are also accessible, but they only gain H3K27Ac histone mark in LPS-activated macrophages. Latent enhancers are unmarked and inaccessible in unstimulated macrophages, but they gain accessibility and acetylation upon stimulation. Repressed enhancers look like constitutive enhancers in unstimulated cells, but they lose H3K27Ac in response to antimicrobial signals. Different stimuli activate a slightly different subset of latent enhancers, accompanied by different specific TF motif enrichments. LPS latent enhancers are enriched for AP-1 and JunB motifs, whereas TGF β are enriched for SMAD-like motifs. Enhancer activation is associated with the transcription of their eRNAs, and their acetylation and transcription precede gene expression changes (Kaikkonen et al. 2013).

Collective behaviour in immune responses

A population of outwardly similar cells can contain several hidden states. These hidden states can be transient, but they can affect how cells react to stimuli. NF κ B activation involves the nuclear translocation of the p65 subunit, which leads to gene expression changes. This response is cyclic and coordinated at high doses of TNF α . However, at lower doses, only a fraction of cells manage to perform the translocation, and it becomes much less coordinated. Stochastic differences in regulatory circuits cause small differences in response thresholds. Therefore, due to the slightly different signal thresholds presented by each cell, the population as a whole can transform a single-cell digital response into an analogue response that can adapt to a range of input levels (Tay et al. 2010). LPS response genes display higher levels of heterogeneity than house-keeping genes. Some of them display bimodal expression in bone marrow-derived dendritic cells, where some cells express a transcript and others not (Shalek et al. 2013). Preventing cell communication reduces expression of maturation markers and bimodality among some of the early response inflammatory genes (Shalek et al. 2014).

Responding cells can also alter gene expression by other cells, acting as alert sentinels, secreting factors that target their neighbours' receptors. Macrophages in isolation display a reduced cytokine production (Xue et al. 2015). For example, during viral infection *Ifnb1* is only expressed in a fraction of cells (Rand et al. 2012; Schmid et al. 2016), even if the infection is wide-spread. However, its membrane receptor, *Ifnar*, is present in all cells. So, as soon as some early response

cells begin to secrete *Ifnb1*, the neighbours respond to *Ifnb1*. Cell-to-cell communication allows signal amplification and response coordination (Rand et al. 2012). *Ifnb1* producers present a characteristic high expression of antiviral core genes which is dependent in autocrine interferon feedback, as cells lacking *Ifnar*, an IFN receptor, uncouple IFN expression to the genes in this module (Shalek et al. 2014).

Cohesin in hematopoiesis and leukaemia

Blood cells are generated from HPSCs in the bone marrow. This stem cell population has to differentiate, generating a variety of cell lineages, and to renew themselves, so the progenitor population is maintained. Excessive differentiation leads to HPSCs depletion, whereas excessive self-renewal leads to neoplasms (Figure 1.10). The balance between differentiation and self-renewal is not an isolated system, as diverse stimuli can modulate it. Some stimuli, such as inflammation, push HPSCs towards differentiation, impairing self-renewal (Pietras 2017). This stimulus-driven differentiation has been used to develop cancer treatments. For example, HDAC inhibitors promote leukemic cell differentiation through the activation of IFN-1 pathway (Salmon et al. 2020).

Classic cancer driver mutations affect direct regulators of proliferation, apoptosis or differentiation. However, there is a subset of recurrent mutations that affect general transcription

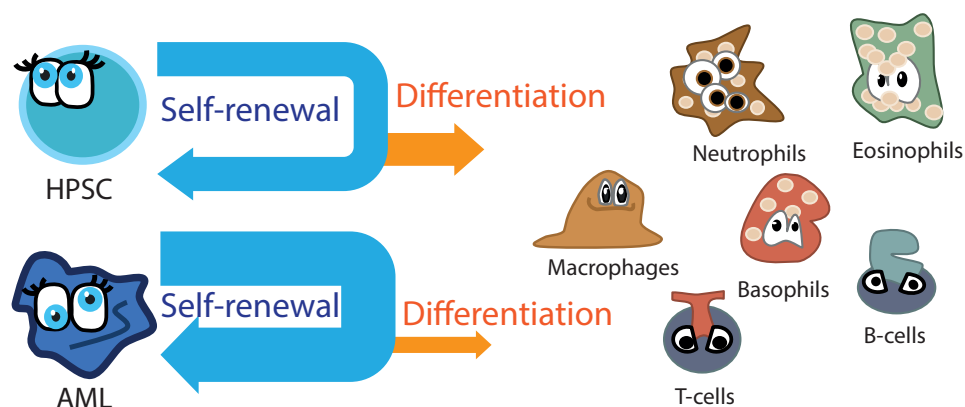


Figure 1.10: **Self-renewal vs differentiation during hemaetopoiesis.** HPSCs can perform a symmetric cell division an produce more HPSCs or divide asymmetrically and differentiate into multiple immune cell lineages. The balance between self-renewal and differentiation swings towards self-renewal in blood cancers.

regulators. Among them, cohesin mutations are recurrent in myeloid neoplasms and constitute early events that are early tumourigenic events not associated with abnormal karyotypes (Cancer Genome Atlas Research Network 2013; Corces-Zimmerman et al. 2014; Cuartero et al. 2019; Kon et al. 2013).

As cohesin is required for normal chromosome segregation, the absence of genome instability is surprising. However, mitosis can progress with 13% of cohesin wild-type levels (Heidinger-Pauli et al. 2010), and disease-linked mutations only involve partial loss of cohesin function (Kon et al. 2013). Therefore, the cancer-driving force of cohesin mutations is ascribed to abnormal gene regulation.

Cohesin depletion can increase self-renewal and abrogate differentiation of hematopoietic stem cells (Fisher et al. 2017; Galeev et al. 2016; Mullenders et al. 2015; Viny et al. 2015). Ageing associates to chronic inflammation, alters the behaviour of HPSCs, and increases the chance to develop neoplastic syndromes (Kovtonyuk et al. 2016). Late in life, only a clonal subset of the initial birth HPSCs population remains. In this context, mutations that increase clonal proliferation and promote resistance to differentiation stimuli will confer a clonal advantage. Accordingly, cohesin mutant HPSCs become predominant in ageing mice (Chen et al. 2019c), promoting a pre-leukemic population (Corces-Zimmerman et al. 2014). Phenotypes differ among mutations in different cohesin subunits. SMC3 haploinsufficiency is not sufficient to cause leukaemia in mice. However, SMC3 haploinsufficiency enhances a leukaemic phenotype in combination with FLT3-ITD mutation, a common leukaemic mutation (Viny et al. 2015). In contrast, conditional knock-out of STAG2 in HPSCs was enough to generate leukaemia in mice (Viny et al. 2019). Cohesin complexes have either STAG1 or STAG2, so in the absence of STAG2, STAG1 can carry out cohesin cell cycle functions, but gene regulation becomes defective. Removal of both, STAG1 and STAG2 impairs proliferation and results in karyotypic abnormalities (Van Der Lelij et al. 2017; Viny et al. 2019). STAG2 is the cohesin-related mutation most common in cancer. STAG2 cancer mutations cluster at the STAG2-CTCF interaction surface, suggesting that cancer mutations target cohesin-CTCF association (Li et al. 2020).

In summary, cohesin malfunction leads to an altered gene regulation that pushes HPSCs towards self-renewal, promoting a pre-leukemic state.

Chapter 2

Cohesin at the crossroad between inflammation and cancer

Cohesin mutations in Acute Myeloid Leukaemia

AML is a poor prognosis blood cancer that originates in bone marrow myeloid lineage cells. Less than 1% of the total diagnosed cancers are AML, but its incidence rate has increased by 29% in the last decade (*AML statistics*, 2020). Recurrent mutations arise in distinct gene subsets. Despite its essential functions during the cell cycle, mutations in members of the cohesin complex are frequently reported (Cancer Genome Atlas Research Network 2013), but they do not associated with abnormal karyotypes (Corces-Zimmerman et al. 2014; Cuartero et al. 2019; Kon et al. 2013) or fatality rate (Thol et al. 2014).

In the BEAT AML cohort (Tyner et al. 2018), from 408 patients analysed, 49 contained cohesin mutations (12%), with STAG2 being the most frequently mutated subunit (5.15% of all cases). Dr Yi-Fang Wang classified AML by mutation subgroup and performed a DE analysis, followed by a GSEA. Cohesin mutated AMLs presented a distinctive expression program that clustered with AML mutations in *FTL3* and *WT1*. This program was characterized by the down-regulation of inflammatory hallmarks (Cuartero et al. 2018) and up-regulation of OXPHOS, *Myc* targets, DNA repair and mTOR signalling (Figure 2.1).

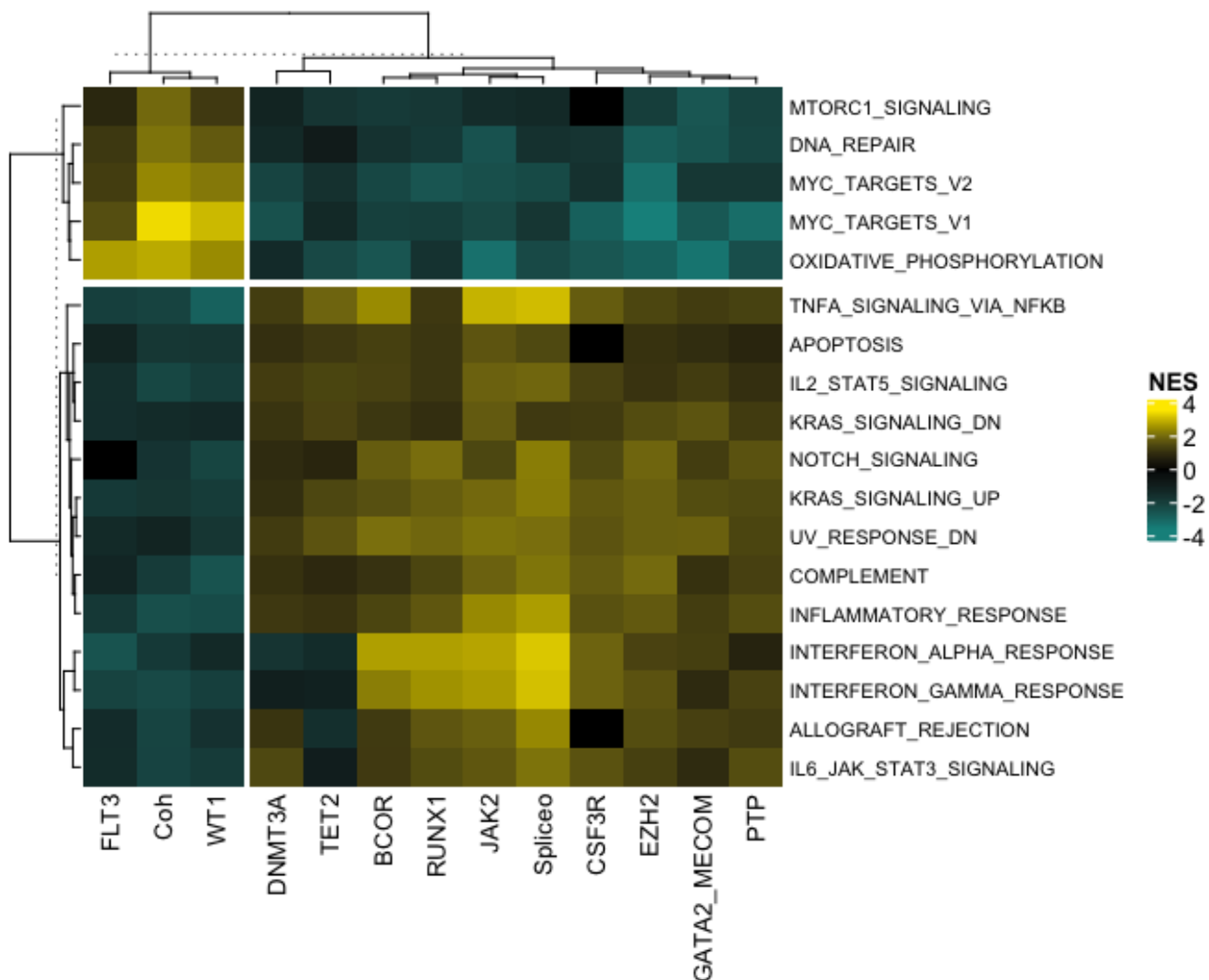


Figure 2.1: AMLs with cohesin mutations present a distinctive expression program that clusters with *FLT3* and *WT1* mutations, with increased OXPHOS and *Myc* and reduced inflammatory hallmarks. Clustering of GSEA analysis of AMLs (Tyner et al. 2018) classified by subgroups of mutations. MSigDB hallmarks (Liberzon et al. 2015) displayed on the y-axis. Coh, cohesin mutations. Spliceo, mutations in the spliceosome. Figure courtesy of Yi-Fang Wang.

Studying the role of cohesin in gene regulation

Complete loss of cohesin is incompatible with cell cycle, but partial levels of cohesin can carry out normal chromosome segregation (Heidinger-Pauli et al. 2010), and increases self-renewal of HPSCs, abrogating their differentiation. (Fisher et al. 2017; Galeev et al. 2016; Mullenders et al. 2015; Viny et al. 2015). Thus, the cohesin complex can act as a tumour suppressor. The aim of our study is to understand cohesin role in gene regulation. Since cohesin is involved in chromosome segregation, to avoid any confounding effect due to defective mitosis, we work on a non-dividing primary cell type, BMD macrophages (after 7 days of differentiation, 89% cells are in G1 (Cuartero et al. 2018) which have a well-characterised response to microbial signals (Bhatt et al. 2012; Kaikkonen et al. 2013; Ostuni et al. 2013).

Knocking out the cohesin subunit RAD21 in macrophages

To knock out cohesin, we used a conditional creERT2 system in primary cells. Our mice had a Cre recombinase fused to an estrogen receptor and *Rad21* exons 6 and 5 floxed by lox-P sites. The chimeric protein remained in the cytoplasm until 4-hydroxytamoxifen (4-OHT) was added to the media. This triggered a conformational change in the ERT2 receptor that led to its translocation to the nucleus. Once in the nucleus, Cre can find the lox-P targets and cut, knocking out the cohesin's subunit (Seibler et al. 2003; Seitan et al. 2011). In the absence of 4-OHT, mice with floxed *Rad21* alleles display a normal phenotype. We extracted bone marrow progenitors from *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{wt/wt}* mice and *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{lox/lox}* mice (Seibler et al. 2003; Seitan et al. 2011), and differentiated them into macrophages for 7 days.

On day 4, when cells start to display a macrophage-like morphology, we treated the cells with 4-OHT (Figure 2.2). 24 hours (h) later, only about 1 in 10 *Rad21* floxed alleles remained intact, and 72 h later, less than 1 in 100 remained unexcised (Figure 2.2B). As control, we use also 4-OHT treated macrophages with the chimeric construction but a WT *Rad21* allele. At day 7, *Rad21* KO macrophages only had 5-30% of the RAD21 protein levels of WT macrophages (Figure 2.2C and D). To visualise protein depletion at the single-cell level, we performed RAD21 immunofluorescence (IF). In the WT, fluorescent signal concentrates in the nuclei, which was absent from the nuclei of *Rad21* KO macrophages (Figure 2.2E and F).

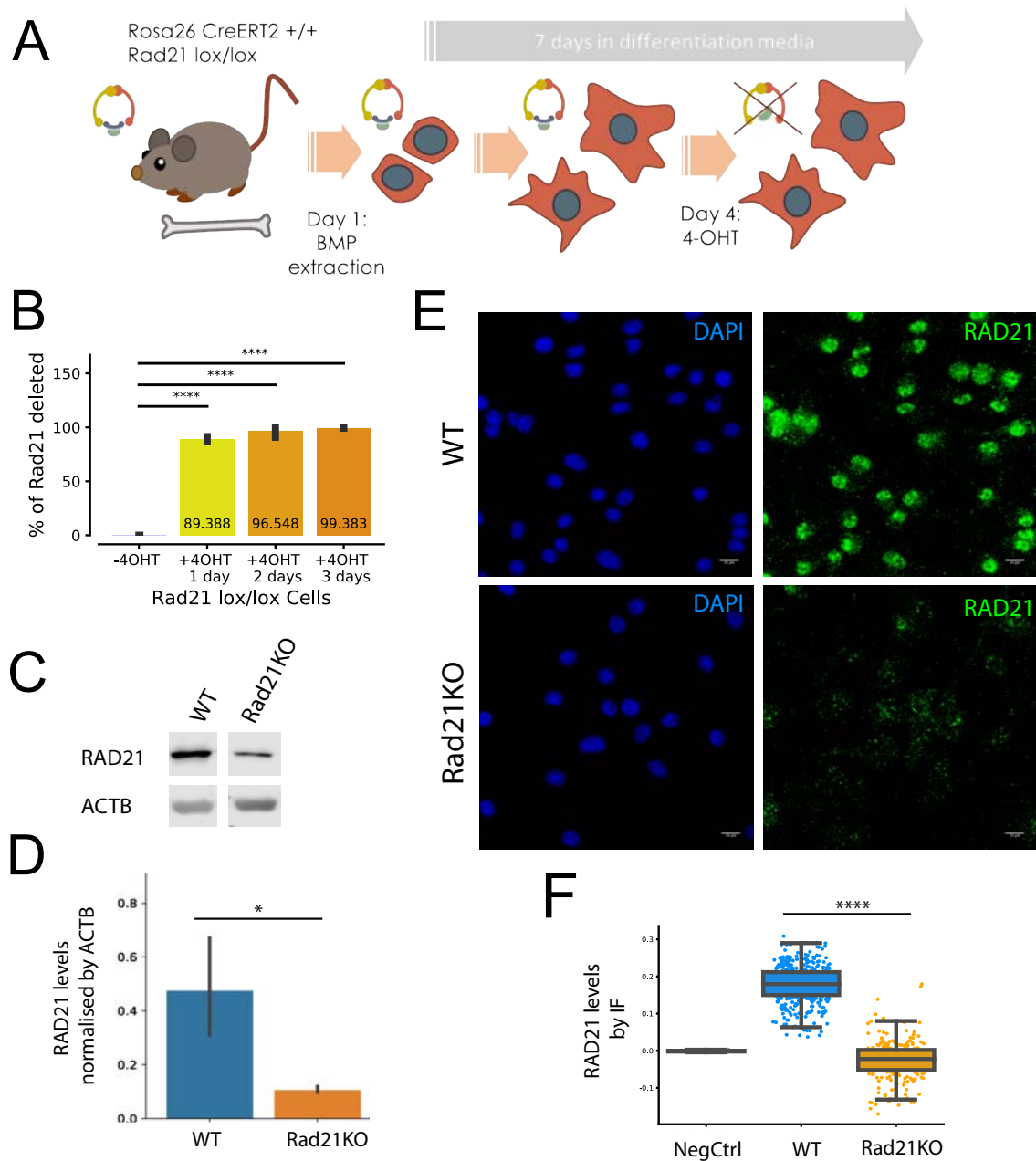


Figure 2.2: **Cohesin-deficient macrophages.** A) Bone marrow progenitors were extracted from mouse tibia and femur and then differentiated into macrophages for 7 days. These mice have the gene *Rad21*, a cohesin subunit, loxP-flxed and a chimeric Cre enzyme coupled to an ERT2 receptor in the ROSA26 locus. At day 4, 4-OHT was added to the media, leading to CreERT2 nuclear translocation and knockout of *Rad21* alleles. Experiments are performed at day 7. B) Genetic *Rad21* allele depletion dynamics by genomic DNA quantitative PCR (gpPCR). As a control, day 7 differentiated macrophages not treated with 4-OHT were used (-4OHT). Statistical test: One way ANOVA. C) RAD21 and ACTB western blot images and D) quantification. Statistical test: t-test. E) Rad21 IF images. F) Rad21 IF quantification. See *Materials and Methods* for details. Statistical test: t-test. NegCtrl were WT cells where Rad21 primary antibody was missing from the IF. 726 cells imaged in total. For each experiment, at least 3 biological replicates were performed.

Weakened immune response in *Rad21* KO macrophages

To test how cohesin deficient cells react to microbial signals, we treated our macrophages with LPS. LPS is a component of Gram negative bacteria that binds to TLR4 receptor, initiating a signalling cascade that activates macrophages (Figure 2.3B). To unveil gene expression changes, RNAseq was performed (Cuartero et al. 2018). The transcriptional landscape of *Rad21* KO macrophages reproduced several features of cohesin mutated AMLs. Hallmark GSEA analysis showed that "Myc targets", "OXPHOS" and "Fatty acid metabolism" terms were up-regulated, whereas inflammatory hallmarks were down-regulated (Liberzon et al. 2015; Subramanian et al. 2005). In contrast, neither "MTORC1 signalling" nor "DNA repair" terms were up-regulated in *Rad21* KO macrophages (Figure 2.3A).

In untreated macrophages, cohesin depletion impact on constitutively expressed genes. In contrast, LPS inducible genes (defined in Bhatt et al. 2012) are massively down-regulated. Differences between *Rad21* KO and WT increase as LPS-activation progresses. There is also a repression failure of LPS responsive genes. At 8h after LPS, 53% of LPS-repressed genes were up-regulated in *Rad21* KO macrophages compared to WT (Figure 2.3B and C).

This weakened immune response has two possible explanations. Either *Rad21* KO macrophages fail to detect LPS signals, or they fail to establish an appropriate transcriptional response. Upon detection of antimicrobial signals, the p65 subunit of NF κ B translocates to the nucleus where it modulates gene expression. We measured the ability of macrophages to translocate p65 as a proxy for the detection of LPS. *Rad21* KO macrophages performed p65 translocation 1h after LPS. However, 2h after LPS, WT macrophages displayed p65 translocation, but the percentage of positive macrophages dropped in *Rad21* KO. (Figure 2.3D). Therefore, *Rad21* KO macrophages can detect LPS, but they cannot establish an appropriate response.

Cohesin deficiency leads to unpaired immune responses

Given the previous results, we hypothesize that cohesin defects drive cells into a cell-state opposed to the transcriptional program activated by inflammatory signals, like LPS. To do so, we compared GSEAs of cohesin defects to gene expression changes during macrophage LPS-activation (DEseq2, WT macrophages, 8h LPS + INF γ vs untreated). We included in our

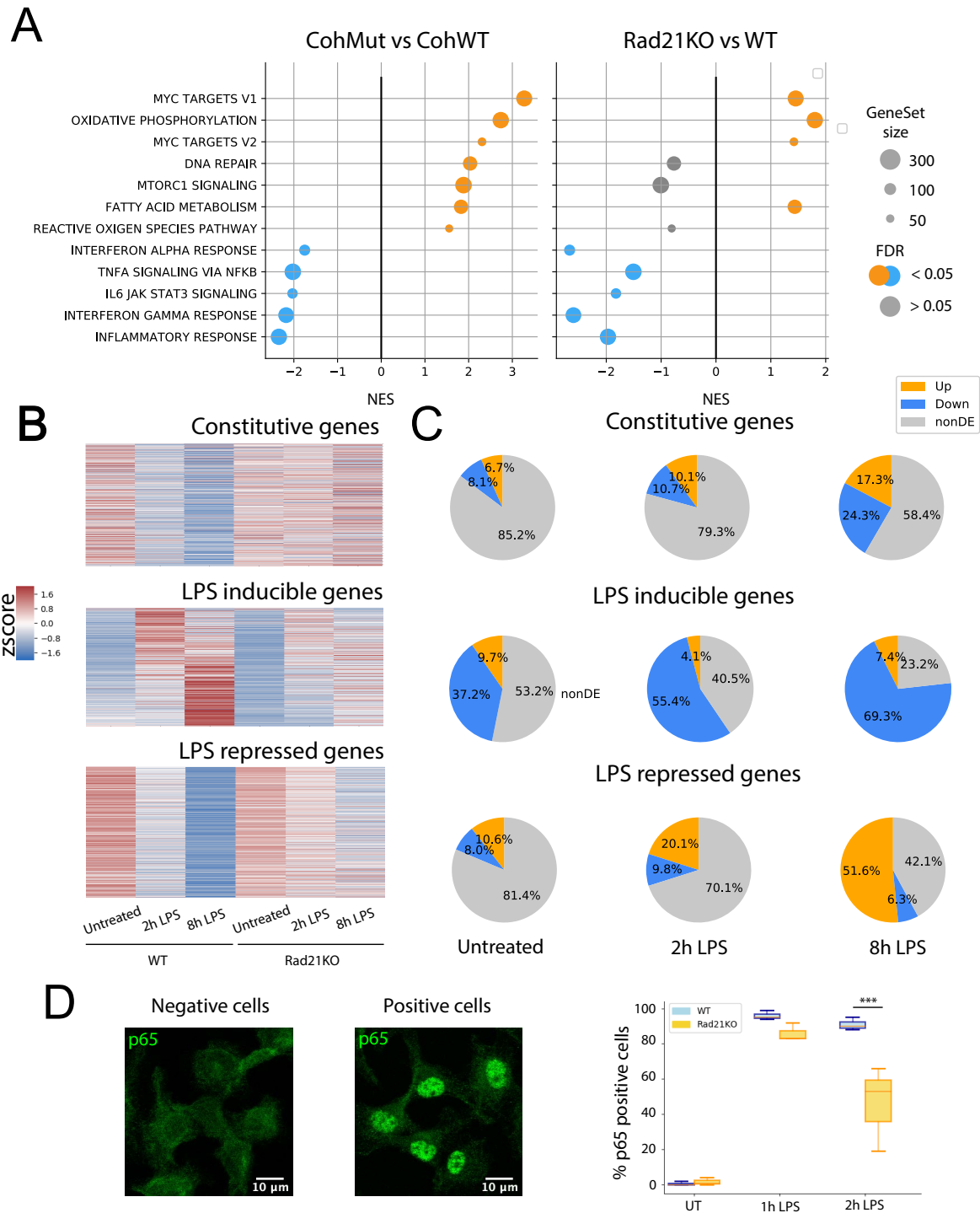


Figure 2.3: *Rad21* KO macrophages detected antimicrobial signals but failed to establish a strong immune response. A) GSEA comparison between selected terms of cohesin mutated AML vs cohesin WT AML (CohMut vs Coh WT) and *Rad21* KO vs WT macrophages (DEseq2 comparison, data from Tyner et al. 2018). B) Z-score (expression value minus mean expression of a gene across conditions, and then divided by the standard deviation across conditions) of log2 + 1 FPKM of constitutive, inducible and repressed genes. C) Proportion of DE genes in each group comparing *Rad21* KO vs WT D) NF κ B translocation example and their percentage after LPS induction. A cell was considered as positive if the correlation between DAPI and the immunofluorescence signal was greater than 0.5. Statistical test: 2-way Anova with Sidak Correction test. 3 biological replicates. 4645 cells imaged.

analysis a system halfway between the cycling and heterogeneous human AML and the quiescent and homogeneous macrophages: LSK cells. LSKs comprise a population of HPSCs and multi-potent progenitor (MPP), precursors of most AMLs and macrophages. We used LSK expression data where cohesin subunits *Smc1* and *Stag2* were silenced via an inducible shRNA (Mullenders et al. 2015). shRNA of SMC1 and STAG2 in LSK cells lead to the same deregulation of pathways observed in *Rad21* deficient macrophages and in cohesin mutated AMLs. Upon pre-treatment with IFN- γ and LPS, macrophages activated inflammatory and IFN-related hallmarks which were down-regulated in cohesin deficient systems. Macrophage activation led to the down-regulation of "Myc targets" and "OXPHOS", which were up-regulated in cohesin-deficient systems (Figure 2.4).

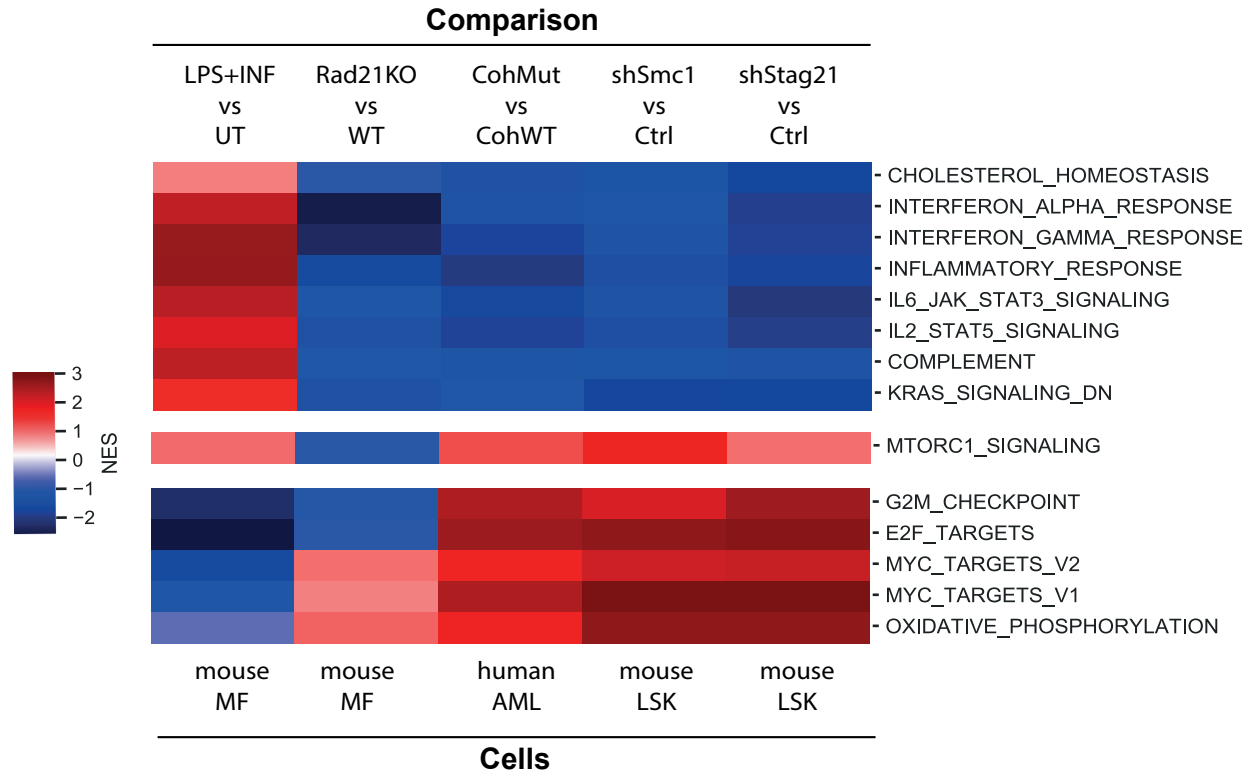


Figure 2.4: **Expression hallmarks are inverted between macrophages establishing an antimicrobial response and cells with cohesin defects.** LPS+INF vs UT, GSEA results of DESeq2 analysis between 8h LPS + 24h IFN- γ treated macrophages and untreated macrophages. *Rad21* KO vs WT, GSEA results of *Rad21* KO macrophages and WT macrophages. Macrophage data from Cuartero et al. 2018. CohMut vs CohWT, cohesin mutated leukaemias vs non-cohesin mutated leukaemias, human AML tumor biopsy data from Tyner et al. 2018, GSEA, mutation classification and DESeq2 analysis performed by Yi-Fang Wang. shSmc1 vs Ctrl and shStag2 vs Ctrl, GSEA results of a shRNA against *Smc1* and *Stag2*, two subunits of the cohesin complex, vs cotrol shRNA in mouse LSK, data from Mullenders et al. 2015. NES, Normalised enrichment score.

Patterns associated to DE genes in *Rad21*KO macrophages

Inducible gene expression failure: IFNs

ISRE motif associates to genes with reduced expression

The promoters of the genes that display a reduced expression in *Rad21* KO macrophages were enriched with the interferon-stimulated response element (ISRE) (Figure 2.5A). This motif activates the expression of IFN stimulated genes when bound by the multimeric transcription factor complex ISG3. The ISG3 complex is formed as a consequence of JAK/STAT signalling due to the activation of IFN type 1 receptor (Ivashkiv and Donlin 2014; Schneider et al. 2014) (Figure 2.5B). The majority of genes involved in the formation of IGF3 were down-regulated at baseline in *Rad21* KO macrophages (Table 2.1).

Rescue of late inducible genes by IFN γ in *Rad21* KO

IFN- γ polarises macrophages into an M1 pro-inflammatory phenotype. We hypothesised that priming *Rad21* KO macrophages with IFN- γ will rescue part of the inducible gene expression program. In Bhatt et al. 2012, they classified LPS-inducible genes by their timing and their subcellular association. Classes A1, A2, B and C have peaked within the first 90 min after LPS, whereas classes D, E and F increased expression until the end of their experiments. IFN- γ priming partially rescued the inducibility of classes D, E and F classes in *Rad21* KO macrophages with respect to baseline expression (Figure 2.5C) and in comparison to WT (Figure 2.5D).

Gene	log2FoldChange	adjusted p-value
<i>Irf9</i>	-0.55	9.67e-5
<i>Ifnar</i>	-0.8	5.87e-4
<i>Stat1</i>	-1.25	3.14e-11
<i>Stat2</i>	-1.63	2.9e-10

Table 2.1: DE analysis of genes involved in IFN signalling in *Rad21* KO macrophages vs WT.

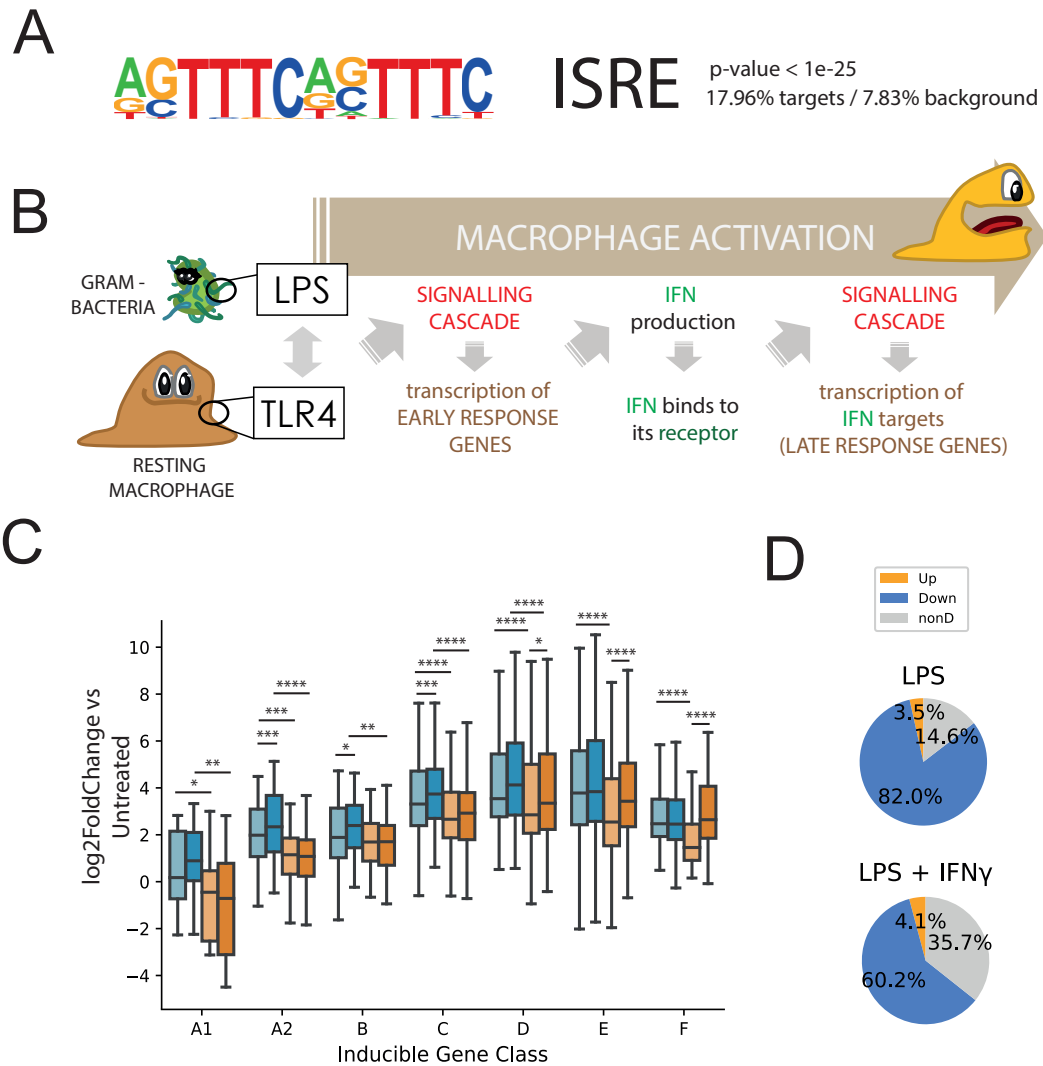


Figure 2.5: IFN γ pre-treatment rescues part of late inducible genes in *Rad21* KO macrophages. A) Top known-motif enriched at the promoter of down-regulated genes in *Rad21* KO macrophages compared to non-DE genes (DEseq2, untreated *Rad21* KO macrophages vs WT). p-value, over-represented p-value. B) IFN signalling during macrophage activation. C) Log2fold changes of different classes of inducible genes vs baseline expression for WT and *Rad21* KO. The later the class, the later LPS-induction occurs, as described in Bhatt et al. 2012. Statistical test: Wilcoxon Rank test as implemented in HypothesisTests.jl Julia package. Only statistically significant comparisons are displayed. D) Inducible genes classes D, E and F DE in *Rad21* KO vs WT after 8h of LPS and 8h of LPS + IFN γ pre-treatment, DEseq2 comparison.

Redistribution of cellular resources upon LPS and IFN γ

I then studied overall changes in gene expression in response to antimicrobial signals. To do so, we decided to use Voronto (Santamaría and Pierre 2012). Voronto employs Voronoi diagrams to visualise hierarchical gene expression ontologies. Terms locate in the map by similarity and their size correspond to the number of genes they contain. In that sense, they can help to visualise redistribution of cellular resources in response to a signal.

The distribution of transcription is very similar between WT and *Rad21* KO untreated macrophages. However, upon LPS activation, WT cells reorganise their gene expression drastically, switching down most metabolic pathways, and terms related to DNA replication and repair. This reorganisation is accentuated when WT macrophages are pre-treated with IFN γ (Figure 2.6).

Rad21 KO macrophages fail to redistribute their resources after 8h of LPS activation. Only if *Rad21* KO macrophages are primed with IFN- γ some notions of redistribution of cellular

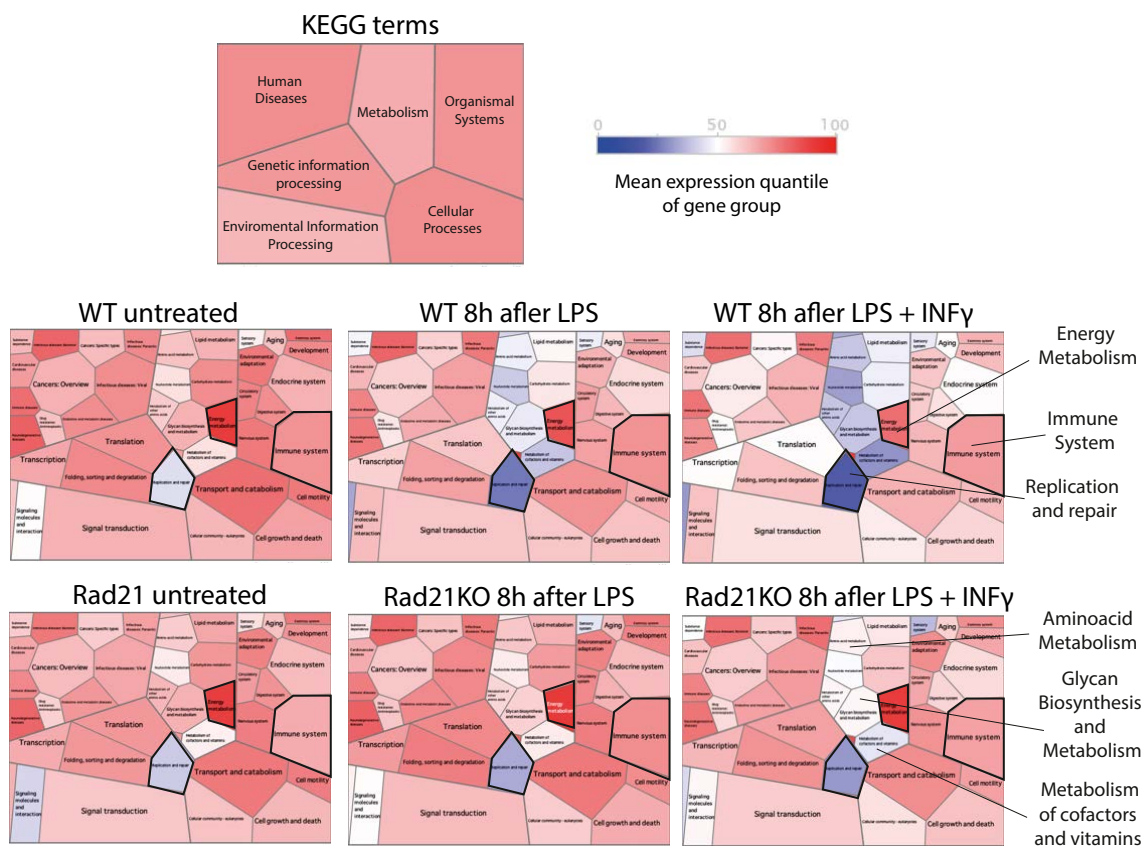


Figure 2.6: **Redistribution of cellular resources upon LPS and IFN γ .** Voronto maps of KEGG gene expression ontologies. Colour correspond to the mean expression quantile of term's genes (Santamaría and Pierre 2012). Input data: mean FPKM across replicates.

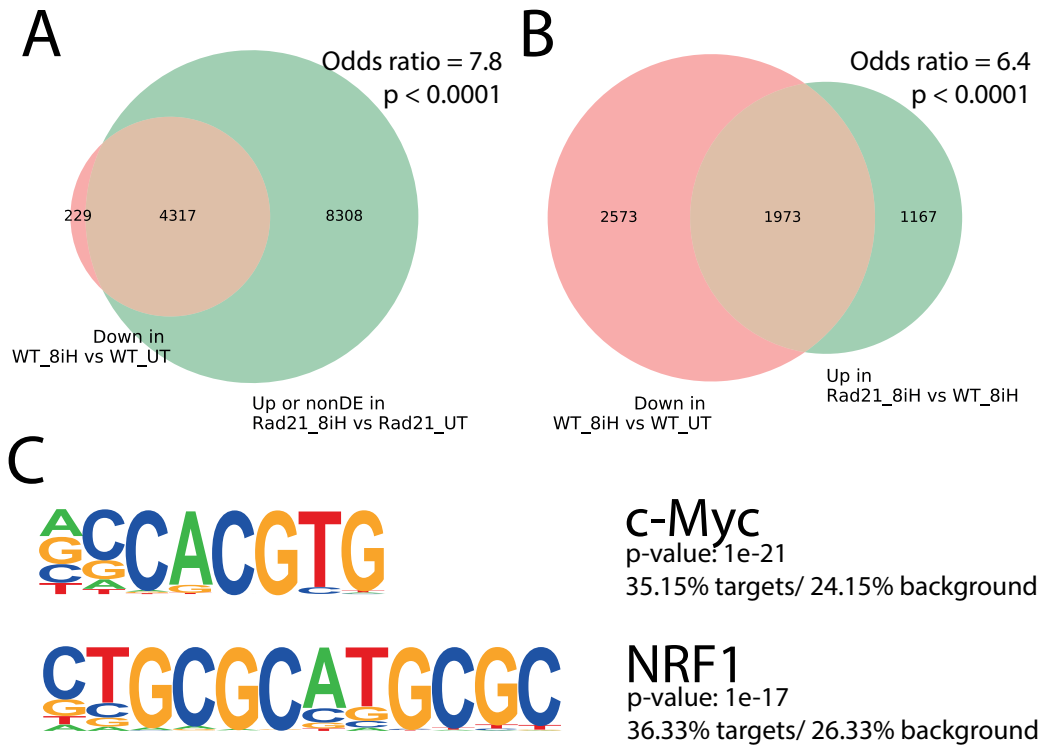


Figure 2.7: ***Rad21* KO macrophages failed to repress genes with c-Myc and NRF1 motifs.** A) Overlap between LPS-repressed genes and genes not down-regulated genes in *Rad21* KO upon $\text{INF}\gamma$ + 8h LPS. p-value and odds ratio calculated using the R package GeneOverlap (Shen 2014). B) Overlap between repressed genes during $\text{INF}\gamma$ + 8h LPS in WT and genes up-regulated in *Rad21* KO upon 8h LPS compared to WT. C) Promoter enriched motifs in 2.7A intersection as gene target group compared the background of all expressed genes. p-value, over-represented p-value.

resources occur. This partial rescue suggests that the activation of the IFN pathway helps macrophages to reallocate their gene expression in response to antimicrobial signals (Figure 2.6).

Repression failure: OXPHOS and Myc

After studying defective LPS-activation, we decided to study LPS-repression in cohesin deficient macrophages. The majority of LPS-repressed genes were not repressed in LPS-activated *Rad21* KO macrophages, even with $\text{INF}\gamma$ priming (95.0%). As a result, 43.4 % of them are up-regulated in *Rad21* KO activated macrophages compared to WT .

Next, to uncover patterns associated with failed LPS-repression, we performed known motif enrichment on LPS-repressed genes in the WT but not in the *Rad21* KO (Heinz et al. 2010).

The two top over-represented motifs were c-Myc and NRF1 (Figure 2.7). NRF1 is one of the master regulators of mitochondria biogenesis and OXPHOS. Myc targets and OXPHOS hallmarks were repressed during macrophage activation, and their expression was consistently increased in cohesin deficient systems (Figure 2.4). Furthermore, these motifs were also enriched among the up-regulated genes in cohesin mutated human AMLs and cohesin deficient mouse LSKs (Table 2.2 and 2.3).

Target genes	Cells	Targets (%)	Background (%)	% Change over background
WTi8 vs WT untreated (Down)				
<i>Rad21</i> KOi8 vs <i>Rad21</i> KOUT (Up)	mouse MF	35.31	26.24	34.57
shSmc1 vs Ctrl (Up)	mouse LSK	28.84	23.79	21.23
shStag2 vs Ctrl (Up)	mouse LSK	31.45	27.45	14.57
CohMut vs CohWT (Up)	human AML	29.05	26.88	8.07

Table 2.2: **NRF1 motif enrichment.** NRF1(NRF)/MCF7-NRF1-ChIP-Seq(Unpublished) motif enrichment in the promoter of target genes against all other expressed genes as background. 8i, IFN γ pre-treatment + 8h LPS; mouse MF data (Cuartero et al. 2018) mouse LSK data (Mullenders et al. 2015), human AML data (Tyner et al. 2018). MF, BMD macrophages. Up, up-regulated set of genes in DESeq2. Down, down-regulated set of genes in DESeq2.

Target genes	Cells	Targets (%)	Background (%)	% Change over background
WTi8 vs WT untreated (Down)				
<i>Rad21</i> KOi8 vs <i>Rad21</i> KOUT (Up)	mouse MF	59.08	48.7	21.31
shSmc1 vs Ctrl (Up)	mouse LSK	53.95	47.78	12.91
shStag2 vs Ctrl (Up)	mouse LSK	51.1	51.09	0.02
CohMut vs CohWT (Up)	human AML	51.43	45.75	12.42

Table 2.3: **Myc motif enrichment.** c-Myc(bHLH)/LNCAP-cMyc-ChIP-Seq(Unpublished) motif enrichment in the promoter of target genes against all other expressed genes as background. 8i, IFN γ pre-treatment + 8h LPS; mouse MF data (Cuartero et al. 2018) mouse LSK data (Mullenders et al. 2015), human AML data (Tyner et al. 2018). MF, BMD macrophages. Up, up-regulated set of genes in DESeq2. Down, down-regulated set of genes in DESeq2

Increased mitochondrial content in *Rad21*KO macrophages

Mitochondrial genes are up-regulated in *Rad21* KO macrophages

Myc and NRF1 are regulators of mitochondria biogenesis and OXPHOS, and their motifs associate to LPS-repression failure in *Rad21* KO macrophages (Figure 2.7). Moreover, an increase in OXPHOS is a common feature of cohesin deficient systems (Figure 2.4). OXPHOS occurs

in the mitochondria, although most OXPHOS genes are encoded in the nucleus. We decided to interrogate whether the OXPHOS increase came with an increase in mitochondrial content, which is often associated with an up-regulation of mitochondrial encoded genes. In *Rad21* KO macrophages, there was an up-regulation of mitochondria-encoded genes (34.78% of 23 genes up-regulated). In the mouse LSK data from Mullenders et al. 2015, from 27 mitochondria-encoded genes, only one was significantly up-regulated. However, a GSEA analysis revealed a positive enrichment of mitochondrial encoded genes in cohesin deficient LSK (shStag2 vs control, NES: 1.96, p-value: 0.00187; shSmc1 vs control, NES: 2.57, p-value: 0.00036).

***Rad21* KO macrophages have increased mitochondria content**

One way to calculate the average number of mitochondria per cell is to calculate the relative mitochondrial DNA (mtDNA) content. Relative mtDNA content is the number of mitochondrial DNA copies over 2 x the number of copies of a nuclear DNA (nuDNA) region by qPCR.

With no 4-OTH in the medium, creERT2 is unable to knock-out floxed *Rad21* alleles. In this case, there are no significant differences between the relative mtDNA of WT and *Rad21* floxed macrophages (Figure 2.8A). When 4-OHT is added to the media at day 4, mitochondrial content increases by ~25-50% in *Rad21* KO cells (Figure 2.8B). These results were confirmed by imaging. MitoTracker Red CMH2XRos is a reduced dye that fluoresces when it is oxidised. It is used to visualise mitochondria in live cells because it accumulates depending on the potential of mitochondrial membranes. We observed an increased in the mitochondrial mass in terms of the fraction of cell cover by mitochondria, MitoTracker intensity by cell, and by mitochondria in *Rad21* KO macrophages compared to WT (Figure 2.9B). No differences between cell sizes were found between WT and *Rad21* KO macrophages.

Mitochondria during macrophage differentiation

Knocking out of *Rad21* alleles started 3 days before we usually performed our experiments. At this point in the project, we wanted to identify at which point mitochondria dynamics were affected by cohesin loss. Therefore, we decided to study mitochondrial dynamics over macrophage differentiation by relative mtDNA content (Figure 2.10A). Macrophages increase

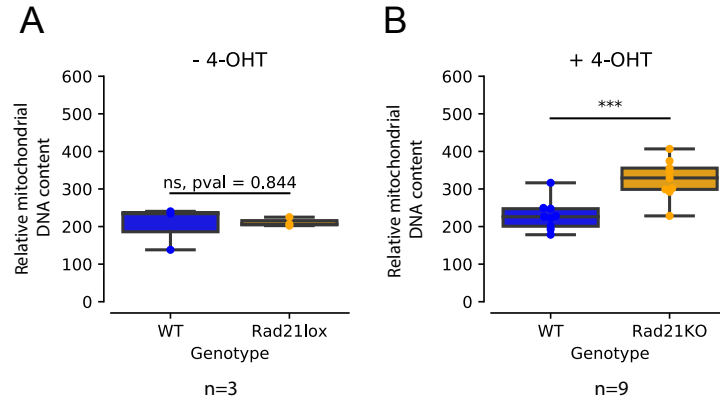


Figure 2.8: **Increased relative mtDNA content in *Rad21* KO macrophages.** Relative mtDNA content in cells where 4-OHT had not been added (A) or had been added (B). Cells resulting from differentiate bone marrow cells into macrophages for 7 days in conditional L929 media. n, biological replicates. Data obtained by qPCR. Statistical test: equal variance t-test as implemented in HypothesisTests.jl.

their relative mtDNA as they differentiate in two jumps. The first jump was a drastic increase between day 3 and 4, right before the addition of tamoxifen, and the second jump spanned from day 6 onwards. Differences between *Rad21* KO cells and WT are only significant after day 7. Increasing the number of biological replicates (from 3 to 6) did not reach statistical significance at day 6, for significance at day 7 take a look at Figure 2.8B.

Mitochondrial regulators during macrophage differentiation

Then, we decided to look at the expression of different mitochondrial regulators (Figure 2.10B). *Polg2* is the mitochondrial DNA polymerase. *Polrmt* is the mitochondrial RNA polymerase. *Tfam*, *Tfb1m* and *Tfb2m* are the transcription factors that initiate mitochondrial DNA replication. *Nrf1* controls mitochondrial biogenesis and the expression of several oxidative phosphorylation gene. *Gabpa* and *Gabpb2* are a subunits of NRF2 with similar functions to *Nrf1*. Other mayor regulators are the peroxisome proliferator-activated receptors (PPARs) such as *Pparg* and *Ppard*. *Ppargc1a* and *Ppargc1b* phosphorylation regulate mitochondria biogenesis by activating *Pparg* and NRF1, respectively. Another coactivator of NRF1 is *Pprc1*. *Ptpmt1* is mitochondrial phosphatase required for the diferentiation of HPSCs, promoting OXPHOS (Yu et al. 2013). None of these regulators was DE by RNAseq between untreated WT and *Rad21* KO macrophages.

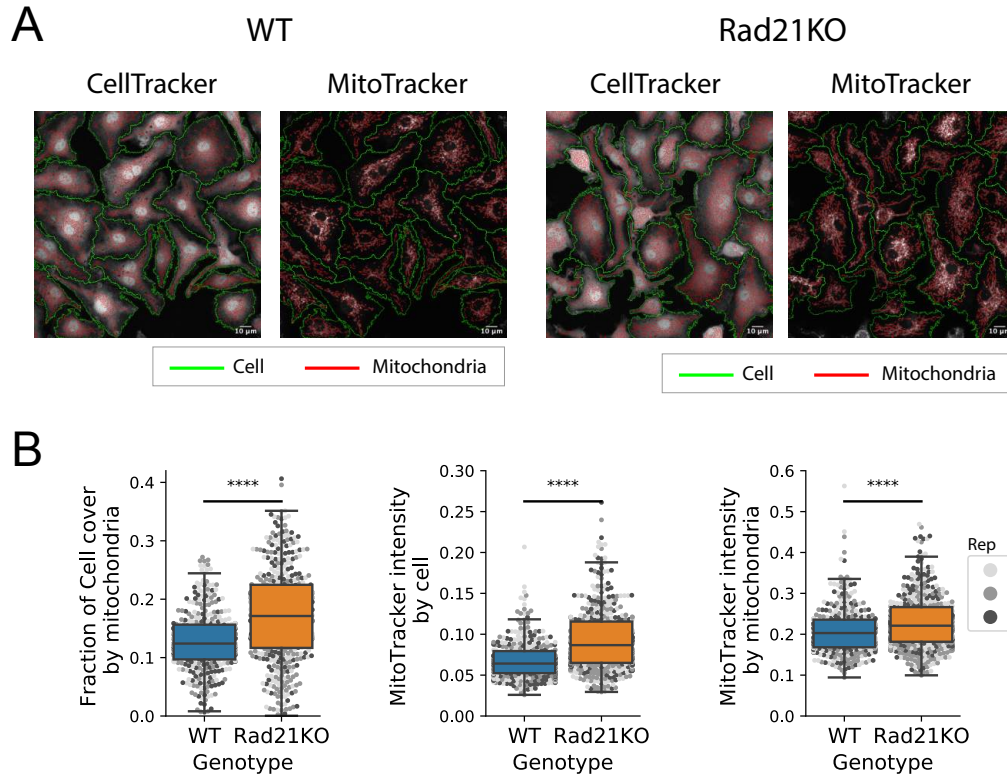


Figure 2.9: **Increased mitochondrial mass in *Rad21* KO cells.** A) Example mitochondria and cell outlines of WT and *Rad21* KO CellTracker and MitoTracker channels. B) Image quantification of mitochondrial content. Cells resulting from differentiate bone marrow cells into macrophages for 7 days in conditional L929 media. Three biological replicates. Between 115 to 167 cells imaged per sample. Statistical test: equal variance t-test as implemented in HypothesisTests.jl.

In general, from day 3 to 5, most mitochondrial regulators drop their expression. From day 5 to 6, their trend is to increase their expression. In particular, the NRF2 subunit *Gabpa* (Equal Variance t-Test p-value: 0.0032). When comparing WT and *Rad21* KO cells at day 6, there is an apparent up-regulation of *Ppard* (Equal Variance t-test p-value: 0.0252), but failed to reach statistical significance after multiple testing correction Figure 2.10B). Sarah Wells performed the RT-PCR for Figure 2.10B and Figure 2.10C.

MYC-MAX-MXD1 axis during macrophage differentiation

Myc, *Max* and *Mxd1* are bHLH domains transcription factors. *Myc* and *Mxd1* can form heterodimers with *Max* with opposite functions. MYC:MAX dimers activate gene expression

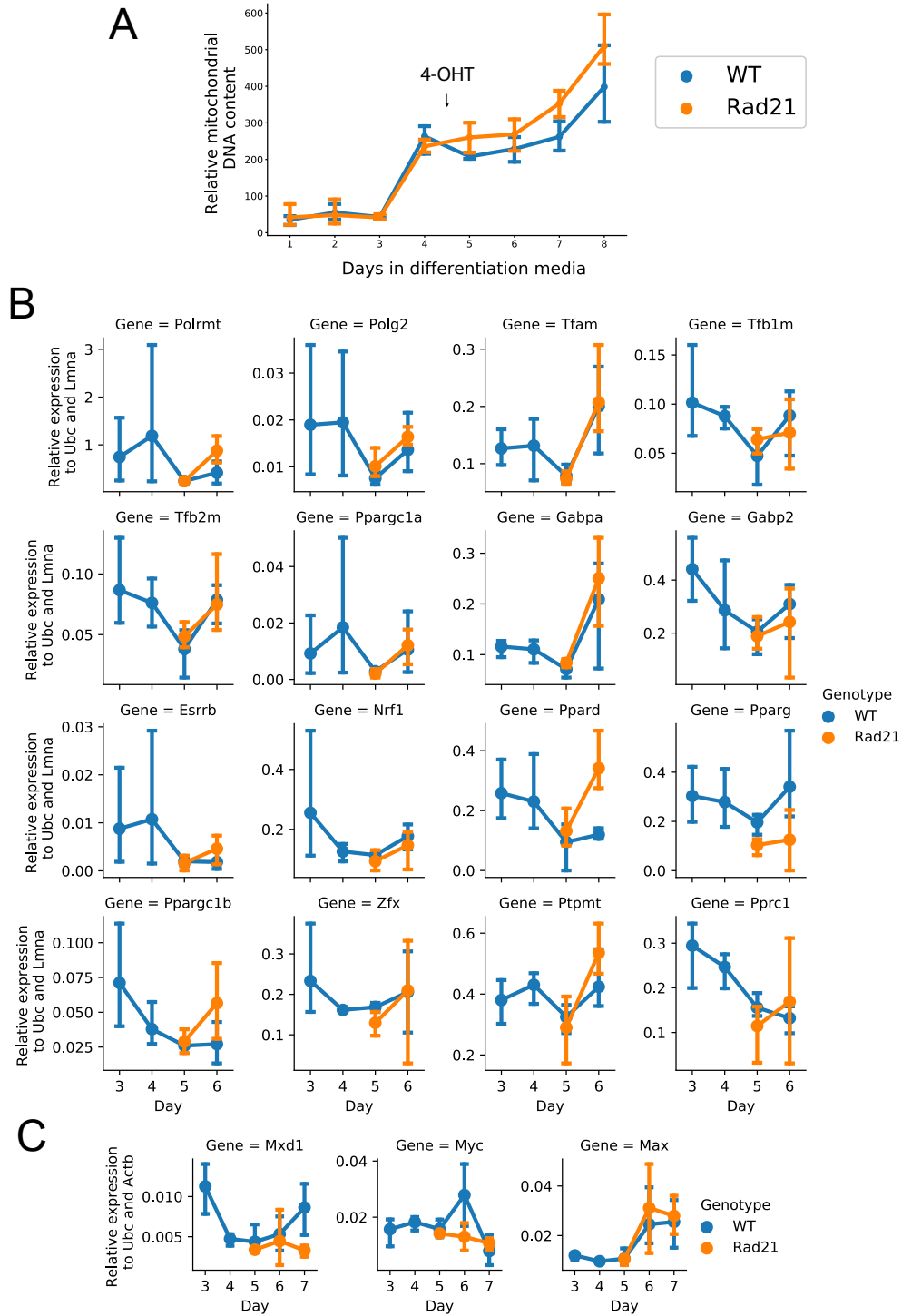


Figure 2.10: Macrophage increase on mitochondrial content during differentiation and expression of potential regulators. A) Relative mitochondria content along macrophage differentiation. At day 4, 4-OHT is added to the media to promote creERT2 translocation and knockout of Rad21 allele. B) qPCR of main mitochondrial regulators. Expression normalised to *Ubc* and *Lmna*. C) qPCR of the MYC-MAX-MXD transcriptional network. Expression normalised to *Ubc* and *Lmna*. x axis, days in L929 conditional media. Each sample has at least 3 biological replicates. Whiskers correspond to 95% confidence interval.

whereas MAX:MXD dimers do not. An increase in MYC activity has been link to an increase in mitochondrial mass and several MYC targets encode for mitochondrial components (Morrish and Hockenbery 2014).

Myc levels were similar across macrophage differentiation (Figure 2.10C). Between day 3 and 4, while relative mtDNA content duplicates (Equal Variance t-test pvalue: 0.0008), *Myc* and *Max* levels were constant, but *Mxd1* expression dropped (Equal Variance t-test pvalue: 0.0253). This could lead indirectly to an increase on the activity of transcriptional activator MYC:MAX dimer, since MAX will not be interacting with MXD.

Between day 4 and 6, relative mtDNA stabilises. *Myc* and *Mxd1* levels were constant. However, *Max* levels increased from day 5 to 6 (taking together data from WT and *Rad21* KO, Equal Variance t-test pvalue: 0.0165, no significant differences between WT and *Rad21* KO). This preceded the increase of mitochondrial numbers between day 6 and 8. Then, *Mxd1* increases in WT, but not in *Rad21* KO (Equal Variance t-test pvalue: 0.0110 by qPCR; log2FoldChange: -1.86, padj: 0.0033 by RNAseq DEseq2 analysis). Suggesting that the presence of MYC:MAX dimers is greater in *Rad21* KO cells than in WT, as MXD is not sequestering MAX.

Discussion

Partial loss of cohesin increases HPSCs proliferation and alters differentiation, presumably by altering the control of gene expression (Chen et al. 2019c; Fisher et al. 2017; Galeev et al. 2016; Mullenders et al. 2015; Viny et al. 2015). The balance between HPSC proliferation versus differentiation integrates internal and external signals, including inflammation (Cuartero et al. 2019; Pietras 2017). In this context, mouse BMD macrophages are an excellent system to study cohesin-dependent gene regulation with potential implications for AML (Cuartero et al. 2018). First, they do not undergo cell-division, so aberrant chromosome segregation is not responsible for gene expression changes. Second, cohesin depletion occurs in primary cells, not cancer cell lines that contain several mutations, which make the behaviour of these cells aberrant and unfit for mechanistic studies. Third, they have a well-characterised antimicrobial response (Bhatt et al. 2012; Kaikkonen et al. 2013; Ostuni et al. 2013) and are close relatives of the progenitor cells in which AML arises (Cuartero et al. 2019). Fourth, expression changes due to cohesin depletion in BMD macrophages resembled gene expression changes in cohesin-mutated AMLs

(Figure 2.3A).

In our analysis, cohesin-related mutations showed a distinctive gene expression pattern that cluster with mutations in FLT3 and WT1, promoting a transcriptional landscape with an increased OXPHOS and c-Myc target expression and reduced inflammation (Figure 2.1). Therefore, FLT3, cohesin and WT1 mutations seem to promote similar expression phenotypes. WT1 is a TF with an essential role in development which is frequently mutated in diverse cancer types (Hastie 2017). FTL3 is a tyrosine kinase with gain-of-function mutations in AML (Sutamtwagul and Vigil 2018). Cohesin and FTL3 mutations co-occur in AML patients, and SMC3 heterozygosity and the FTL3-IDH mutation synergistically promote a leukaemic phenotype in mice (Viny et al. 2015).

Our cohesin knockout system disrupts the cohesin α -kleisin RAD21. By the time we performed our experiments, there was a 99.39 % of *Rad21* genetic depletion and around 85% depletion of RAD21 protein levels (Figure 2.2). As reported by other studies, cohesin subunit depletion had little impact on constitutive gene expression (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Seitan et al. 2013). However, *Rad21* KO macrophages were unable to establish an adequate response to a microbial stimulus, LPS (Figure 2.3B), not because they were not able to detect LPS, but because they were not able to establish an appropriate transcriptional response (Figure 2.3D).

ISRE motifs were enriched at the promoters of down-regulated genes in *Rad21* KO macrophages, suggesting a defective type 1 IFN signalling at baseline. IFN γ priming partially rescues the expression of late inducible genes in *Rad21* KO (Figure 2.5C), and the reallocation of cellular resources (Figure 2.6). Up-regulated genes in *Rad21* KO macrophages overlap with genes repressed during macrophage activation. The intersection between these two gene groups was enriched with c-Myc and NRF1 motifs (Figure 2.7). c-Myc motif associates with the characteristically high "Myc targets" in the GSEA analysis, whereas NRF1 is one of the master regulators of mitochondrial function, and associates to the increase in the "OXPHOS" gene group (2.4). Neither of them was up-regulated in cohesin deficient macrophages. However, increased *Myc* function could be ascribed to a decreased expression of its negative regulator *Mxd1* (Figure 2.10B an C).

While most up-regulated OXPHOS genes were not master regulators of mitochondria biogenesis, a majority of them were mitochondrial components, such as *Ndufa4*, involved in mitochondrial

respiration and *Fhl1*, involved in the Krebs cycle. *Rad21* KO macrophages showed an increase of ~25-50% of mtDNA content (Figure 2.8) which was accompanied by an increase in mitochondrial mass (Figure 2.9). mtDNA content increased during macrophage differentiation in two steps, first between day 3 and 4 and then again from day 6 onwards (Figure 2.10A). We tested several mitochondrial gene regulators in three replicates alongside this time window, but we could not find any significant differences between *Rad21* KO and WT after multiple testing. In order to identify potentially significant and more subtle differences, further research is still required. Most mitochondrial regulators drop between day 3 to 5, and their expression increase after day 6 (Figure 2.10B). One possibility is that the second wave of mtDNA increase was linked to an increase in MYC-MAX heterodimers. Hypothetically, MYC-MAX heterodimers would be increased in *Rad21* KO macrophages, as the expression of the negative regulator *Mxd1* was down-regulated.

Multicellular organisms depend on the collaboration between cells across multiple tissues. Cancer occurs when a cell or a group of cells disengage from the community and start to behave independently. Different mutations mosaics can achieve this phenotype, but they do not respond equally to the same treatment. In this case, we want to understand how cohesin defects modify expression patterns in the haematopoietic lineage, as they are common in myeloid neoplasms (Cancer Genome Atlas Research Network 2013; Corces-Zimmerman et al. 2014; Kon et al. 2013). As inflammatory signals prone HPSCs towards differentiation, it is tempting to conclude that cohesin defects promote a cell-state that is resistant to differentiation stimuli. This resistance is be partially rescued in *Rad21* KO macrophages by IFN γ pre-treatment.

Chapter 3

The quantitative impact of cohesin on transcription

Transcriptional control and cohesin

The expression of protein-coding genes involves the transcription of DNA into mRNA by RNA pol II. Transcription is controlled at different levels, including chromatin accessibility, transcription initiation and pause release (Core and Adelman 2019; Sainsbury et al. 2015). Transcription occurs in bursts (Raj et al. 2008). Imaging experiments define these bursts as the accumulation of nascent mRNA (Mueller et al. 2013). Burst frequency and size (number of mRNA molecules produced per burst) shape gene expression distributions in populations of outwardly identical cells. High burst frequencies and small burst sizes result in low cell-to-cell variability, whereas low frequencies and large sizes result in an increased variability (Cook et al. 1998; Raj et al. 2006b; Singh et al. 2010).

In general, burst sizes are controlled by promoter elements, whereas enhancer elements adjust burst frequencies (Larsson et al. 2019). Promoter elements are responsible for assembling the transcription machinery, whereas enhancers tune gene expression depending on the context. Cohesin depletion disrupts promoter-enhancer interactions (Jaeger et al. 2020) and causes deregulation of genes nearby enhancers (Cuartero et al. 2018; Ing-Simmons et al. 2015; Rao et al. 2017). Enhancers lead gene expression changes during macrophage activation (Kaikkonen et al. 2013; Ostuni et al. 2013). In this chapter, we study how different transcription parameters

are affected by the loss of cohesin. In particular, how transcriptional bursts behave during the failed LPS-response of *Rad21* KO macrophages.

Global transcriptional changes in *Rad21*KO macrophages

Transcription is a major cause of gene expression changes in *Rad21* KO macrophages

Cohesin is a nuclear protein complex (Figure 2.2C) that moves along chromatin fibers (Davidson et al. 2019). RNAseq measures total RNA levels, whereas GROseq measures only at newly made transcripts. RNAseq and GROseq log2Fold changes correlate at baseline (0.642 Pearson correlation, $p\text{-value} < 2.2\text{e-}26$) and especially after 6h of activation (0.722 Pearson correlation, $p\text{-value} < 2.2\text{e-}26$). (Figure 3.1A). This high correlation suggests that gene expression changes in *Rad21* KO macrophages are caused mainly by processes that alter transcription, not by processes that alter RNA stability.

Increase in total RNA content in *Rad21* KO macrophages

We were able to estimate average RNA content per cell because our RNAseq data contained spike-ins and the RNA was extracted from the same number of cells in each sample. At baseline, *Rad21* KO macrophages have around 50% more RNA per cell than WT. LPS activation significantly increases the RNA content of WT cells, but not of *Rad21* KO (Figure 3.1B and Table 3).

Comparison	Transcriptome fold-change	p-value
<i>Rad21</i> KO UT vs WT UT (Untreated)	1.49	5.55e-9
WT 8h LPS vs WT UT	1.38	2.62e-7
<i>Rad21</i> KO 8h LPS vs <i>Rad21</i> KO UT	0.99	0.85

Table 3.1: RNA content comparisons from spike-in RNAseq data

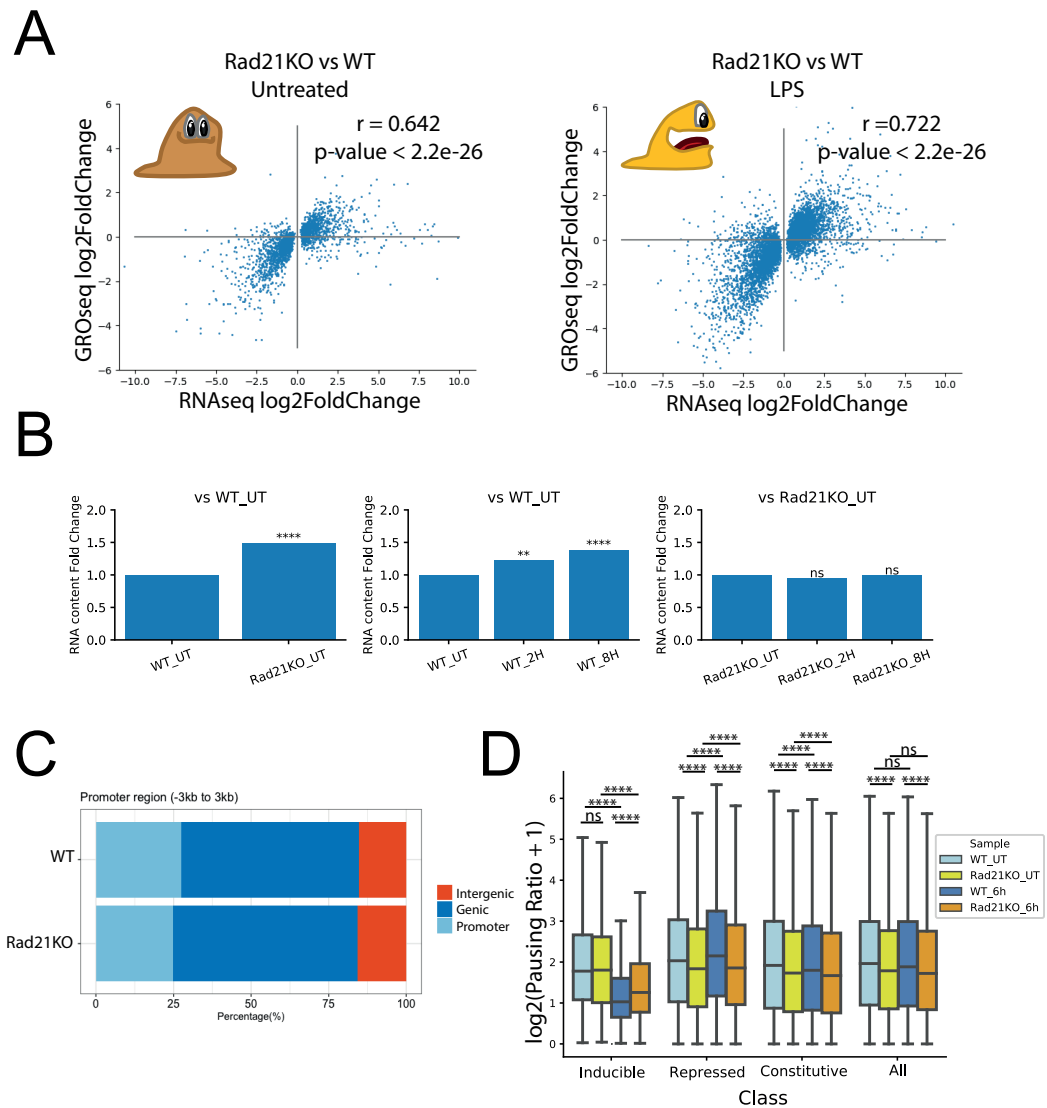


Figure 3.1: Global transcription changes in *Rad21* KO macrophages. A) GROseq fold change vs RNAseq fold change before and after LPS (8h RNAseq, 6h GROseq). DE genes by DEseq2 of RNAseq. r , Pearson correlation. Statistical test: cor.test function in R. B) RNA content fold change compared to WT untreated (vs WT_UT) and *Rad21* KO untreated (vs *Rad21* KO_UT). RNA content calculated from RNAseq data based on spike-ins. C) Proportion of GROseq reads that map in the promoter region (± 3 kb), genic and intergenic regions. D) $\log_2(\text{RNA pol2 pausing ratio} + 1)$ calculated from GROseq data from different gene classes. Statistical test: Wilcoxon Rank test. Multiple testing correction by the Bonferroni method.

Reduction in RNA pol II pause ratios in *Rad21* KO macrophages

Compared to WT, the proportion of GROseq reads that map to promoter regions is reduced in *Rad21* KO macrophages, (Figure 3.1C). RNA pol II release rates affect the number of reads that map to the promoter region compared to the rest of the gene body. Their ratio indicates how paused RNA pol II is on a particular gene (Day et al. 2016). In general, there were lower pausing ratios in *Rad21* KO, except for LPS-inducible genes (Bhatt et al. 2012) which in untreated samples showed no differences. Activation by LPS reduces inducible genes pausing. This reduction is more significant in WT than in *Rad21* KO macrophages. In contrast, LPS-repressed genes displayed an small increase in RNA pol II pause ratios upon LPS, which was more evident in WT than in *Rad21* KO.

Single-molecule perspective of gene expression

Up to now, we focused on cell population data. However, transcription is a single-cell process with low numbers of molecules involved, leading to intrinsic noise, that occurs in bursts which amplifies the resulting cell-to-cell differences. By analysing those differences and the shape of cell distributions, we gain insights into not only on whether mean expression increases or decreases, but also on the transcription parameters and mechanisms. We used smFISH to analyse single-cell differences because it is the gold-standard to measure gene expression in individual cells. In smFISH, fluorescently labelled probes hybridise by sequence-complementarity to target RNA sequences in fixed cells. Specifically, we use a protocol for branched DNA smFISH (bDNA smFISH), developed by Battich et al. 2013. bDNA smFISH uses sequential hybridisations to build a tree-like structure that can increase the number of labels per mRNA, improving signal-to-noise ratio (Figure 3.2A). For image acquisition, we used a confocal microscope to acquire images coupled with high throughput image analysis. Computer software identified nuclei, cells and dot-like signals that represented individual mRNA molecules (Figure 3.2B). (Battich et al. 2013; Carpenter et al. 2006; Mueller et al. 2013; Stoeger et al. 2015).

Validation of bDNA smFISH approach

In order to validate the smFISH approach, we compared smFISH data to RNAseq data, analysed smFISH reproducibility, and compare different features of our data to published data. Our smFISH experiments started with the expression of 6 genes in 2 independent experiments with 3 replicates each. In the first set, we studied *Actb*, a housekeeping gene, and 2 inducible genes

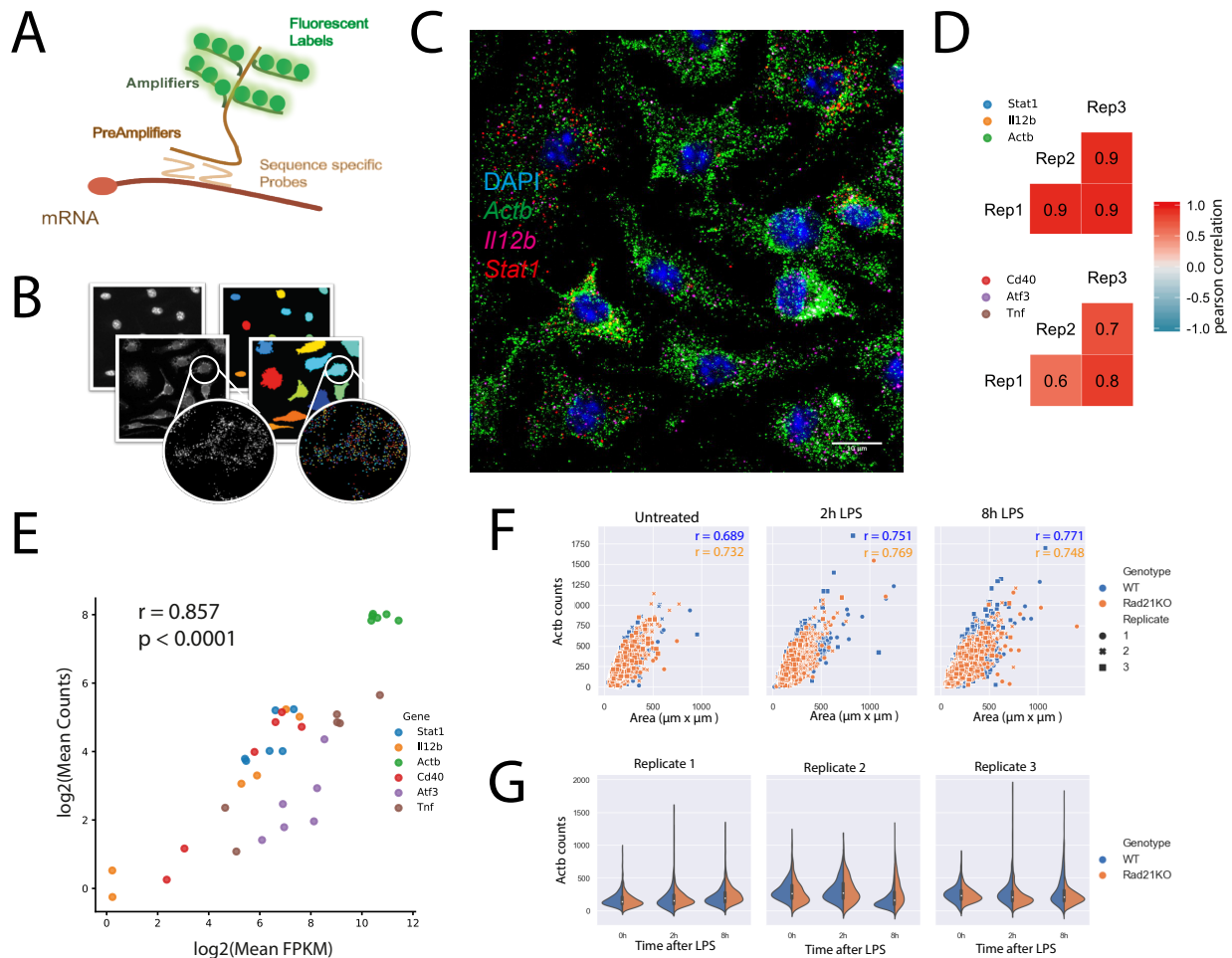


Figure 3.2: **Establishing bDNA smFISH in the lab.** A) bDNA smFISH hybridisation scheme. Sequential hybridisation of DNA probes to amplify the fluorescent signal from a single mRNA molecule (Adapted from Battich et al. 2013). Scale bar represents 10 μm . B) Image analysis pipeline scheme. First, nuclei were identified in the DAPI channel. Then, CellTracker staining was used to identify their associated cytoplasm. mRNA were distinguished as dot-like signals. C) Example of bDNA smFISH image using probes against three different genes in activated macrophages (8h LPS). D) Experiment reproducibility, Pearson correlation among replicates. E) log2 mean counts of smFISH vs log2 mean FPKM by RNAseq. r , Pearson correlation; p , p-value F) Pearson correlation between cell size and *Actb* counts G) *Actb* replicate distributions.

down-regulated in cohesin deficient macrophages: *Il12b* and *Stat1*, a cytokine subunit and a signal transducer molecule (Figure 3.2C as an example image). In the second set, *Tnf*, *Cd40* and *Atf3*.

There was a high correlation between smFISH replicates (Figure 3.2D) and between smFISH data and RNAseq (Figure 3.2E). This indicates that the technique is consistent with RNAseq, and reproducible. As reported (Battich et al. 2013), there is a high degree of correlation between cell size and *Actb* expression, a cytoskeleton component (Figure 3.2F).

Increasing the throughput of image acquisition

At this point in the project, we realised that our imaging had excellent resolution, but it was too slow. The acquisition of one image took about 30 min, and I had to select each field manually. Furthermore, each sample had to be manipulated individually at the microscope. With the help of the microscopy facility, we sped up image acquisition by using confocal resonant microscopy coupled with a bidirectional scan. The resonant-mode led to an increase in noise, which we addressed by increasing the line acquisition average. We were able to image several samples per session by switching to a dry objective and chamber slides. To automatically select the image acquisition sites, we needed an even distribution of cells on the slide. To do so, we changed to a high-throughput protocol that minimised cell loss, which allowed the use of a matrix screener software. In this way, we were able to image around 10000 cells in 8 different samples per night. A substantial improvement compared our initial acquisitions of 500 cells in just one sample per night (Table 3).

	First smFISHs	Optimisation
Image acquisition time	30 min	2 min
Replicate acquisition time (6 samples)	6 nights	1 night
Cells imaged per night	500	10000
Cell selection	Manual	Matrix screener
Slide	13 mm	Chamber slides
Sample change	Manual	Matrix screener
Objective	x40 Oil	x20 Dry
Resolution	101 nm (xy), 300 nm (zx)	189 nm (xy), 500 nm (zx)

Table 3.2: Optimisation of smFISH image acquisition.

Quantification of transcription bursts

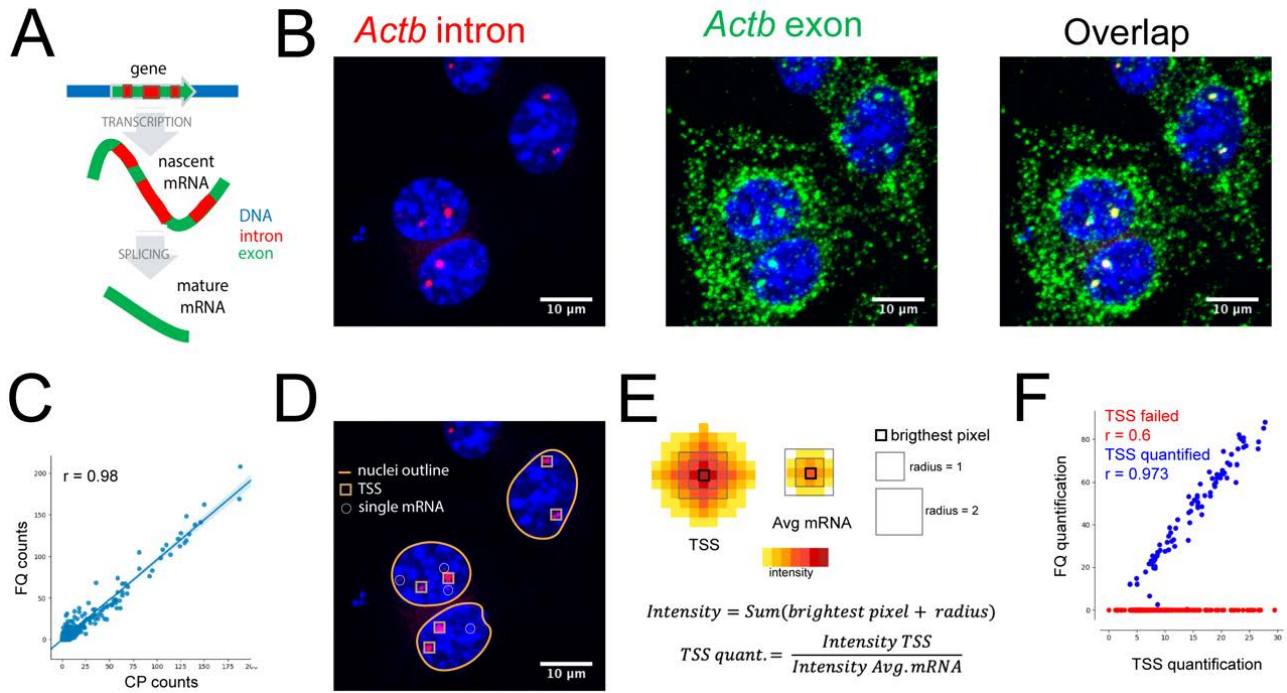


Figure 3.3: Imaging and quantifying transcriptional bursts. A) Nascent transcripts contain introns which are then removed by splicing. B) Intron/Exon bDNA smFISH example, DAPI in blue. C) Comparison between single-molecule quantification of exon dots by a CellProfiler pipeline (CP, Battich et al. 2013) and FISH-Quant (FQ, Mueller et al. 2013) on a set of ~ 1812 WT cells expressing *Il12b* at 0, 30, 60, 90 and 120 min after LPS. D and E) Identification and quantification of transcriptional bursts: D) Nuclear outlines were identified by CP and up to 2 TSS were identified by FQ. Later, FQ detected individual dots and generated an average image of a single mRNA. E) For each TSS, burst quantification was the integrated intensity of a TSS brightest pixel plus a radius of 1 or 2 pixels divided by the integrated intensity of the same area of the average single mRNA. F) Comparison between FQ measurements and ours (using radius = 1) on 717 detected *Il12b* bursts.

Next, we studied how cohesin depletion affected transcriptional bursts of macrophages. Newly-made mRNA contain introns which are removed during their maturation by the splicing machinery (Figure 3.3A). Therefore, nascent transcripts accumulate at their production site, the gene TSS. For imaging experiments, a transcriptional burst and a detected TSS are therefore synonymous. By combining intronic and exonic probes, we can simultaneously quantify nascent and mature transcripts in single cells (Figure 3.3B).

To quantify TSSs, we switched our dot quantification from CellProfiler (ibid.) to FISH-Quant (FQ, Mueller et al. 2013), a specialised program for burst quantification. FQ identifies dot-like signals by fitting a Gaussian function in 3D image stacks; an algorithm initially developed

to find stars. Despite their differences, they produce very similar results on a test image set (Figure 3.3C).

FQ identifies two types of dots, individual mRNA molecules and TSSs. Active TSSs (or transcriptional bursts) are only located in the nucleus, and they are brighter and bigger than the signal coming from individual mRNAs due to the accumulation of nascent transcripts (Figure 3.3D). We tried to quantify burst with FQ built-in algorithm, but it failed in most cases. We worked around this by writing software that compares each TSS to the average image of a dot (Figure 3.3E) with help from Dr Mohammad Mahdi Karimi. There was a good correlation between TSSs quantified with our software, and the subset of TSSs that FQ was able to quantify (Figure 3.3F).

Housekeeping gene expression: *Actb*

There are global changes in transcription caused by cohesin depletion in macrophages (Figure 3.1), despite having a minor impact in constitutive gene expression (Figure 2.3B and C) which has been reported in other systems (Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Seitan et al. 2013).

To test whether bursting differences could occur in non-DE genes, we study the expression from a housekeeping gene, *Actb*, at the single-cell level (Figure 3.4A). RNAseq data did not show any significant differences between untreated WT and *Rad21* KO macrophages (Figure 3.4B), but the higher level of detail of smFISH was able to detect minor differences (Figure 3.4C).

Unexpectedly, we found an small increase in nascent *Actb* transcription which escaped detection by GROseq. Burst frequency was $\sim 5\%$ higher in *Rad21* KO, and each transcriptional burst produced ~ 0.5 more molecules in cohesin deficient macrophages (Figures 3.4D and E, respectively). For a summary of *Actb* transcription measures, see Table 3.3.

Sample	Cells	Pause Ratio	Burst Fraction	TSS intensity	Mature mRNA
<i>Rad21</i> KO	2366	0.83	0.861	6.62	179.8
WT	3201	0.79	0.808	6.15	197.5

Table 3.3: ***Actb* transcription measures in WT and *Rad21* KO macrophages.** Cells, number of cells analysed by FISH in three biological replicates. Pause ratio calculated from GROseq data. Burst fraction, mean burst fraction across smFISH replicates. TSS intensity, mean TSS intensity across smFISH replicates. mature mRNA, mean counts

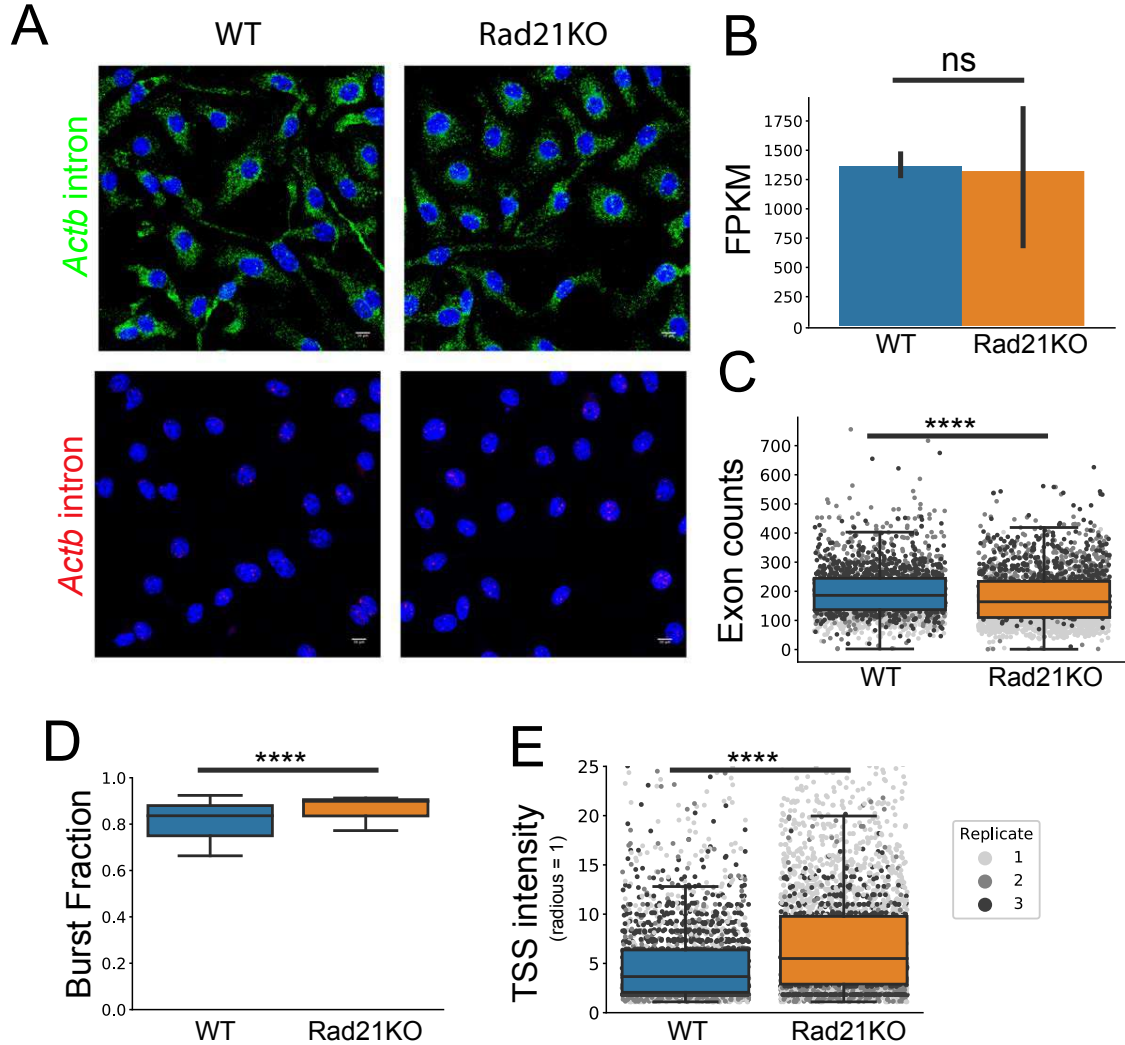


Figure 3.4: **smFISH analysis revealed *Actb* expression differences between WT and *Rad21* KO macrophages.** A) Example images of bDNA smFISH *Actb* intron/exon. Scale bar represents 10 μ m. B) *Actb* expression by RNAseq showed no significant differences. C) Exon counts per cell. Statistical test: Tukey HSD Test (ANOVA design formula $\sim Genotype + Replicate$). D) *Actb* burst fraction. Statistical test: Cochran-Mantel-Haenszel test with replicate as strata variable. E) TSS intensities. Statistical test: Tukey HSD test (ANOVA design formula $\sim Genotype + Replicate$). 4453 cells analysed in total from 3 biological replicates.

Inducible gene expression: *Il12b* and *Peli1*

Detailed timecourse expression data

LPS-inducible response is particularly sensible to cohesin depletion in BMD macrophages (Cuartero et al. 2018, Figure 2.3B and C). The LPS-response is driven by enhancers (Kaikkonen et al. 2013; Ostuni et al. 2013) which are supposed to modulate transcriptional burst frequency (Fukaya et al. 2016; Larsson et al. 2019). through possibly looping (Bartman et al. 2016). Cohesin promotes contacts through loop extrusion between promoters and enhancers (Jaeger et al. 2020). Therefore, we hypothesized that the weakened LPS-response could be a consequence of a weakened control of burst frequencies.

Using intronic and exonic probes, we generated a detailed timecourse of nascent and mature RNA expression for two inducible genes up-regulated in response to LPS but that show lower expression levels in *Rad21* KO macrophages after 2h of LPS-activation, *Il12b* (DEseq2 2h, padj:3.30e-0, log2FoldChange: -2.53) and *Peli1* (DEseq2, padj:1.68e-85, log2FoldChange: -2.4). To capture population distribution dynamics we imaged each gene at 4 consecutive time points. A total of 31,977 of cells were analysed for the *Il12b* timecourse and 27,444 for the *Peli1* from three biological replicates, generating a valuable resource for future mathematical modelling approaches. For a summary of smFISH measures, see Table 3.4.

Mature mRNA analysis: Different gene activation distributions

Il12b and *Peli1* displayed very different activation distributions. Whereas the *Il12b* increase occurred only in a fraction of cells, *Peli1* increase occurred in all cells. *Rad21* KO macrophages managed to increase the expression of both genes upon LPS-activation, but never reached WT induction levels (Figure 3.5).

Nascent mRNA analysis: Reduced burst frequency in *Rad21* KO.

The signal from intronic probes was nuclear and overlapped with exonic signal (Figure 3.6), a good indication of probe specificity. *Il12b* is expressed in big but infrequent bursts, whereas *Peli1* is transcribed more frequently but in smaller burst sizes.

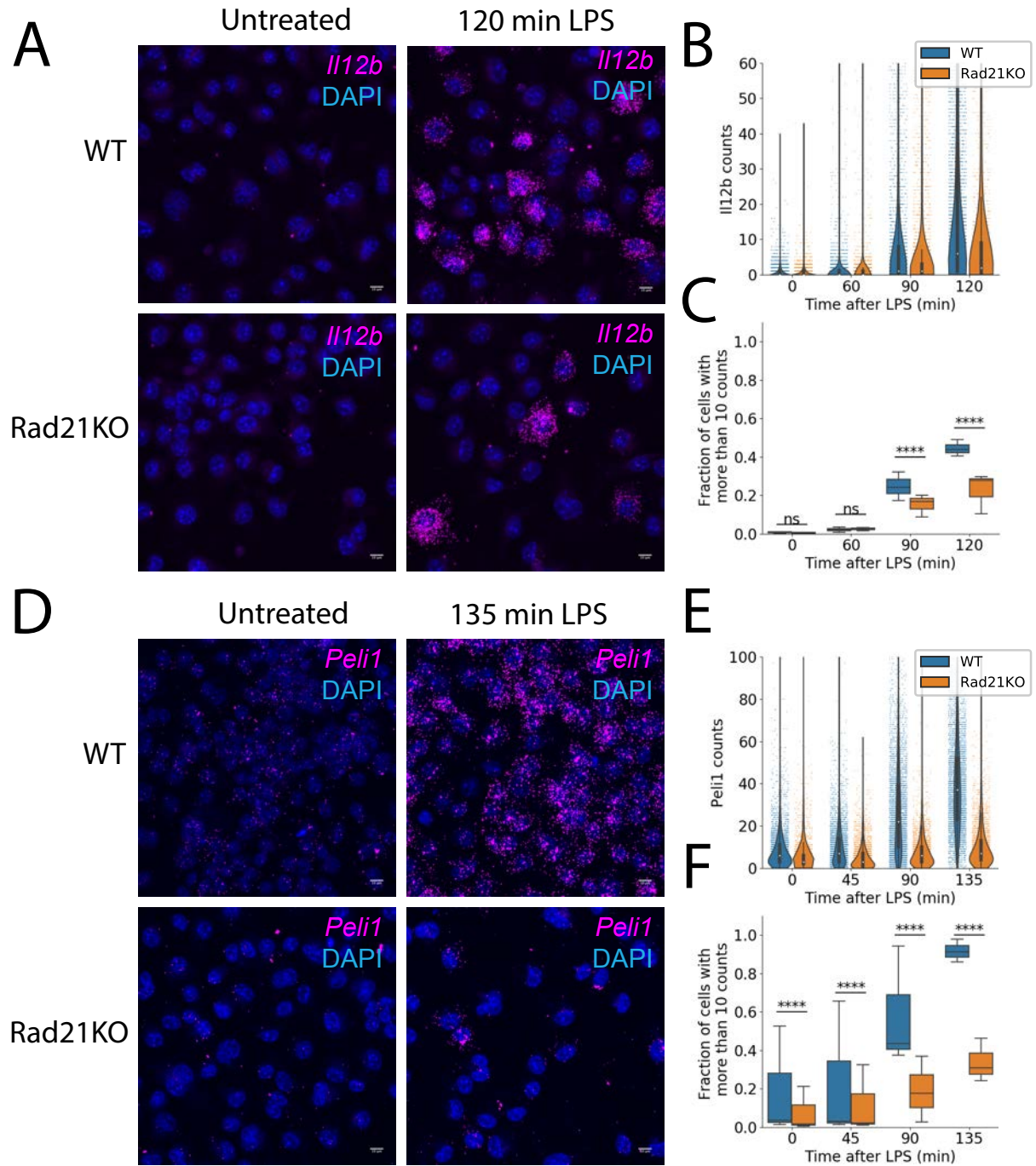


Figure 3.5: **Single molecule perspective of inducible gene expression in WT and *Rad21* KO macrophages.** Example images of *Il12b* (A) and *Peli1* (B) bDNA smFISH. Scale bars represent 10 μ m. Representative mRNA count distribution from one *Il12b* (B) and *Peli1* (E) replicate. Fraction of cells with more than 10 counts in each replicate of *Il12b* (C) and *Peli1* (F). Statistical test: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni correction for multiple testing.

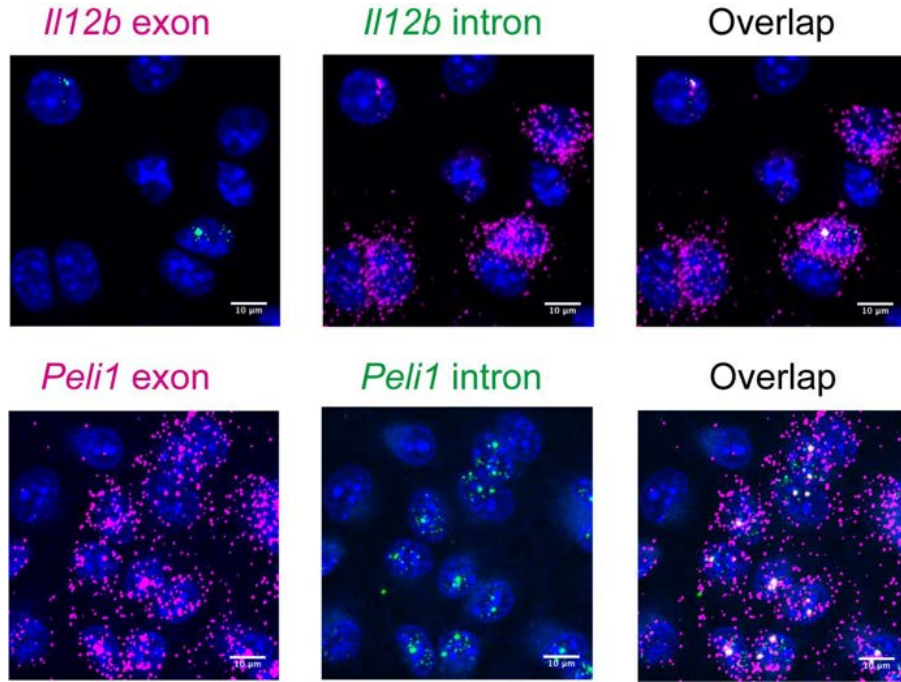


Figure 3.6: *Il12b* and *Peli1* transcriptional burts. Example images of cells bursting *Il12b* and *Peli1* by combining intronic and exonic bDNA smFISH probes.

Genotype	Time after LPS (min)	Gene	Cells	Burst Fraction	TSS intensity	mature mRNA	Fraction >10
WT	0	<i>Il12b</i>	4580	0.002	6.23	0.78	0.006
WT	60	<i>Il12b</i>	4882	0.007	8.23	1.33	0.020
WT	90	<i>Il12b</i>	4199	0.051	7.68	16.59	0.235
WT	120	<i>Il12b</i>	5307	0.039	6.56	31.50	0.431
<i>Rad21</i> KO	0	<i>Il12b</i>	3743	0.001	9.13	0.59	0.003
<i>Rad21</i> KO	60	<i>Il12b</i>	3219	0.009	6.06	1.47	0.024
<i>Rad21</i> KO	90	<i>Il12b</i>	4210	0.018	6.91	9.35	0.146
<i>Rad21</i> KO	120	<i>Il12b</i>	1837	0.006	6.21	16.10	0.219
WT	0	<i>Peli1</i>	4313	0.170	3.23	6.43	0.004
WT	45	<i>Peli1</i>	3594	0.196	3.12	7.15	0.025
WT	90	<i>Peli1</i>	4290	0.518	4.19	18.75	0.221
WT	135	<i>Peli1</i>	4555	0.675	5.18	35.57	0.351
<i>Rad21</i> KO	0	<i>Peli1</i>	2240	0.068	2.95	4.83	0.001
<i>Rad21</i> KO	45	<i>Peli1</i>	3053	0.070	2.96	4.46	0.002
<i>Rad21</i> KO	90	<i>Peli1</i>	2596	0.114	2.96	6.09	0.013
<i>Rad21</i> KO	135	<i>Peli1</i>	2803	0.154	2.98	9.47	0.016

Table 3.4: **Summary of *Il12b* and *Peli1* transcription measures in WT and *Rad21* KO macrophages.** Cells. number of cells analysed by FISH in three biological replicates. Burst fraction, mean burst fraction across smFISH replicates. TSS intensity, mean TSS intensity across smFISH replicates. mature mRNA, mean counts. Fraction >10, fraction of cells with more than 10 counts,

During macrophage activation by LPS, inducible expression of both genes occurred via an increase in burst frequencies. *Rad21* KO macrophages also showed increased *Peli1* and *Il12b* burst frequencies, but to a lesser extent than WT macrophages.

Il12b showed no differences in burst size between WT and *Rad21* KO macrophages (Figure 3.7D). Regarding activation, WT burst sizes of *Il12b* at 90 min were more prominent than 30 min later (Figure 3.7E). On the other hand, *Peli1* burst sizes were smaller in *Rad21* KO cells (Figure 3.7I). After an initial increase in burst frequency *Peli1* burst sizes gradually increased at 90 and 135 min post LPS-activation in WT macrophages but not in *Rad21* KO (Figure 3.7J).

For *Peli1* and *Il12b*, the probability of finding a cell with two bursts was higher expected, revealing coordination between bursts from the two chromosomes. This indicates that the expression of both loci is coordinated and suggests the presence of hidden states that predispose to *Il12b* or *Peli1* expression (Figure 3.7 C and H).

Repressed genes: *Fh1* and *Hprt*

Gene repression by LPS and IFN- γ

Next, we studied the repression mechanisms by LPS and IFN- γ . As *Myc* targets and OXPHOS failed to be repressed by LPS and IFN- γ in *Rad21* KO macrophages, we chose one of each group as representatives. As a *Myc* target we selected *Hprt* and as OXPHOS gene we selected *Fh1*. This pair was interesting because by IFN- γ priming partially rescued *Fh1* repression in *Rad21* KO macrophages, whereas *Hprt* was not rescued (Figure 3.8A). We imaged 26858 cells in 4 biological replicates. Due to technical issues, three samples were removed from analyses. See Table 3.5 for a summary of the transcriptional burst measures of both genes.

Hprt burst dynamics

Hprt encodes for a hypoxanthine phosphoribosyltransferase. According to the MSigDB database, it is among *Myc* targets. *Hprt* is a constitutive gene involved in the purine salvage pathway with high levels of expression. Baseline expression of *Hprt* was characterised by a relatively high

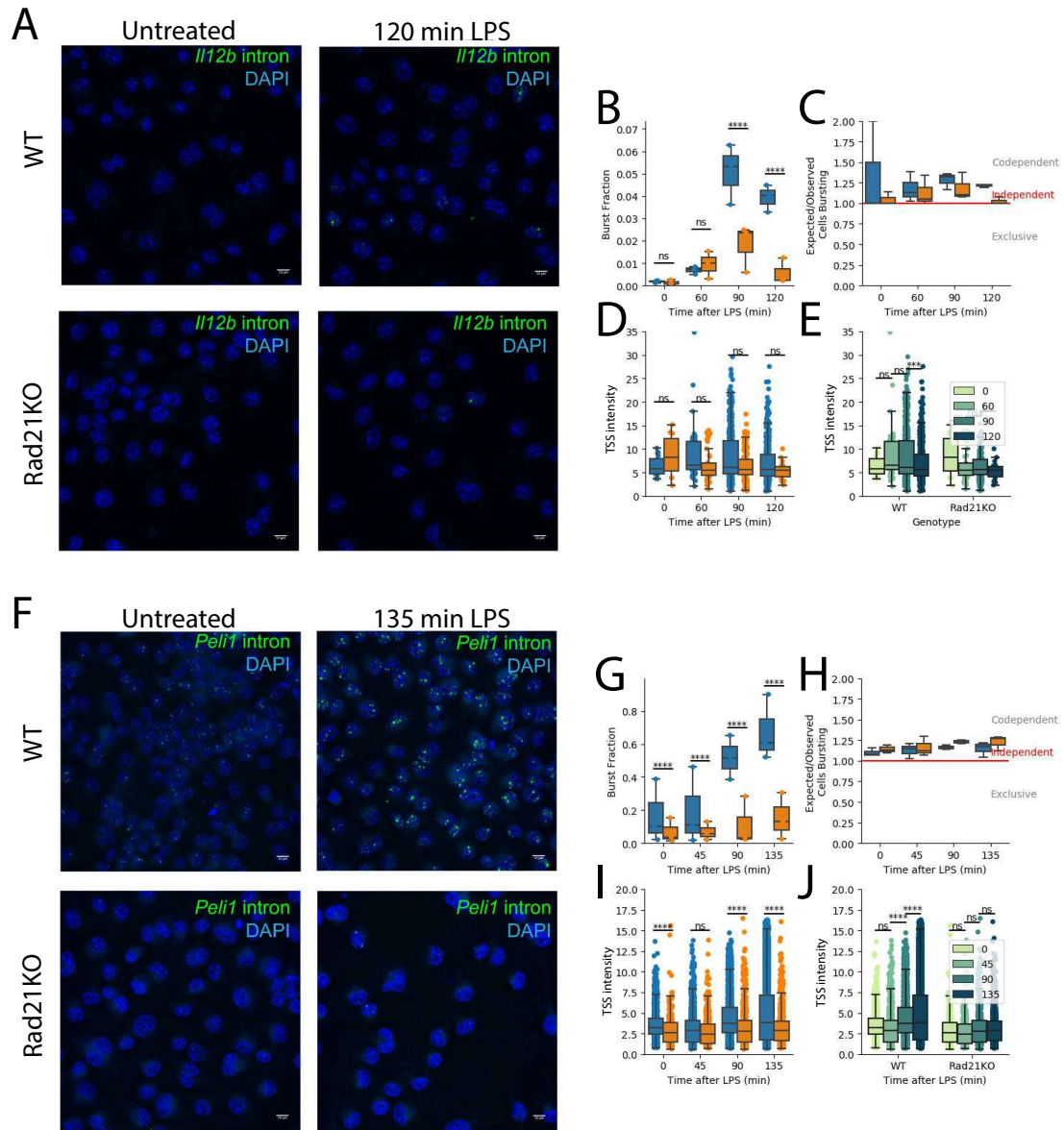


Figure 3.7: ***Rad21* KO cells fail to increase burst frequencies of inducible genes.** A) Examples of *Il12b* intron bDNA smFISH images. B) *Il12b* burst frequencies. C) Number of cells with 2 *Il12b* bursts detected over the expected number if the two alleles burst independently. D and E) *Il12b* burst sizes. F) Examples of *Peli1* intron bDNA smFISH images. G) *Peli1* burst frequencies. H) Number of cells with 2 *Peli1* burst detected over the expected number of cells with two burst detected if the two alleles burst independently. I and J) *Peli1* TSS intensities. Due to background noise, TSS quantified as less than 1 are excluded. Statistical test for burst frequency: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni correction for multiple testing. Statistical test for TSS intensity: Tukey HSD test (ANOVA design formula $\sim Sample + Replicate$). Scale bars in A and F represent 10 μ m.

burst frequency and size (burst frequency of 0.29 and mean TSS intensity of 29 molecules in WT macrophages). *Rad21*KO macrophages showed a mild increase in burst frequency (+8%, p-value < 0.05, Figure 3.8C) and a major increase in TSS intensity (+42.6%, p-value < 0.001, (Figure 3.8F).

Pre-treatment with IFN- γ increases *Hprt* burst frequency, and activation by LPS decreases burst frequency in both, WT and *Rad21* KO macrophages. The difference between WT and *Rad21*KO macrophages in *Hprt* burst frequencies was significant in response to LPS alone but no longer significant after priming with IFN- γ . This was due to an increase in burst frequencies in response to IFN- γ in WT macrophages, not in *Rad21* KO macrophages (Figure 3.8E) .

WT cells experienced a reduction in *Hprt* TSS intensities after activation with LPS or IFN- γ priming. However, if cells were pre-treated with IFN- γ and then activated by LPS, TSS intensities of *Hprt* increased in WT. Interestingly, although a reduction in burst frequency also occurred in *Rad21* KO (Figure 3.8E), LPS-activation and IFN- γ -pre-treatment synergistically increased *Hprt* TSS intensities in *Rad21* KO (Figure 3.8F).

The analysis of *Hprt* burst parameters suggests three conclusions: 1) Gene expression differences between *Rad21* KO and WT in LPS-repressed genes were due changes in TSS intensity and burst frequencies; 2) *Hprt* repression occurred mainly through a reduction in burst frequency; and 3) Activation via LPS or pre-treatment with IFN- γ increased *Hprt* TSS intensities only in *Rad21* KO macrophages.

Pre-treatment with IFN- γ restores the ability of *Rad21* KO macrophages to repress the burst frequency of *Fh1* upon LPS

Fh1 encodes for a fumarate dehydrogenase. This mitochondrial enzyme is involved in OXPHOS through the catalysis of fumarate into L-malate during the Krebs cycle. *Fh1* expression was around 10 times lower than *Hprt* by RNAseq data. Accordingly, it displayed lower burst frequencies, and sizes than *Hprt* (burst frequency of 0.11 vs 0.28 and TSS intensity of 2.14 vs 29 molecules in WT cells).

Despite displaying higher levels of expression in untreated cohesin-deficient macrophages by RNAseq (padj < 0.0001, log2FoldChange: 0.81), GROseq data, which measures nascent transcription, did not show significant differences between untreated *Rad21* KO and WT macrophages

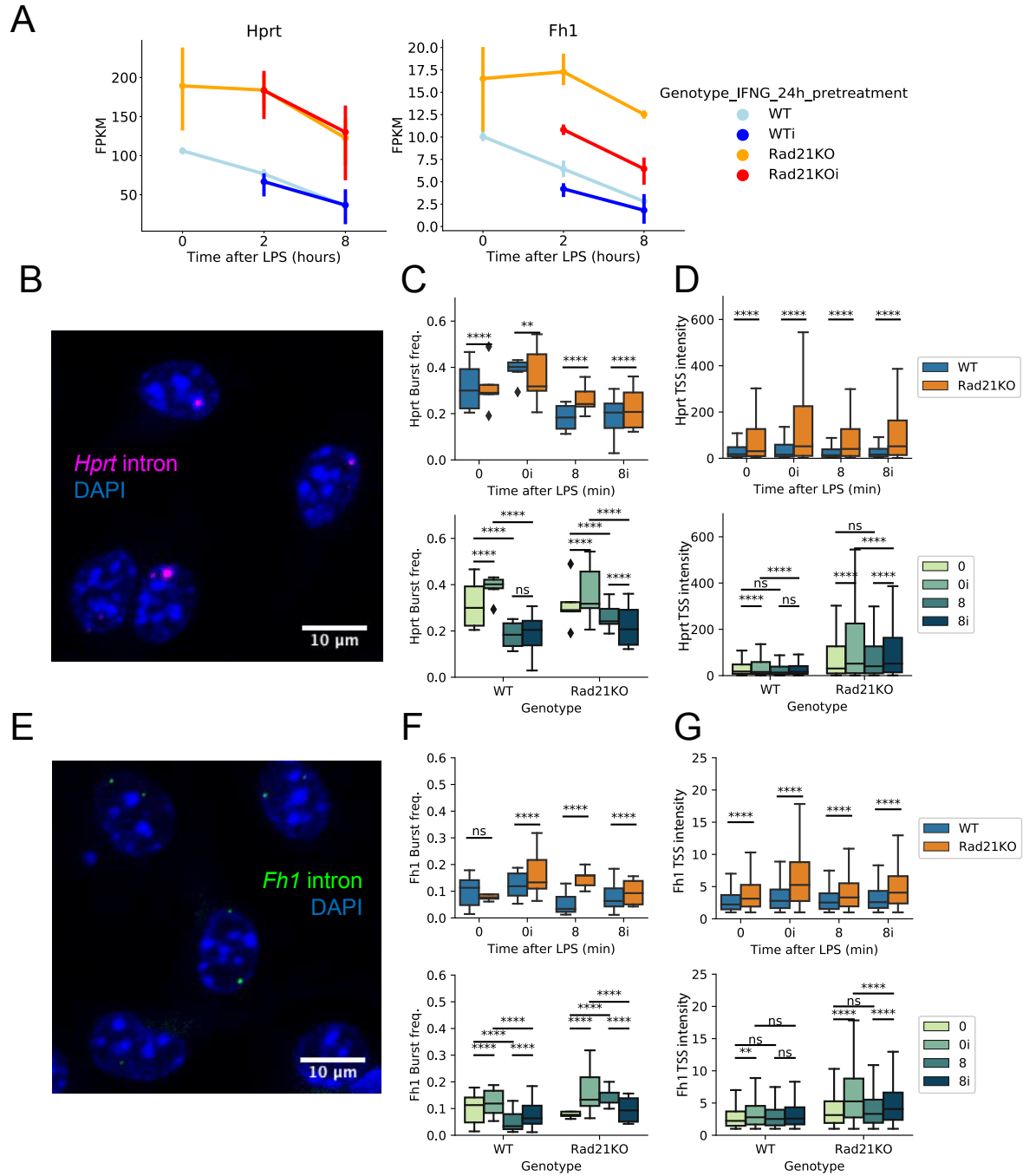


Figure 3.8: **Transcriptional burst of a *Myc* target (*Hprt*) and an OXPHOS gene (*Fh1*).** A) RNAseq FPKM from WT and *Rad21* KO macrophages. (Cuartero et al. 2018). B) Transcriptional burst examples of *Hprt* by smFISH. C and E) *Hprt* burst frequencies. D and F) *Hprt* burst sizes. G) Transcriptional burst examples of *Fh1* by smFISH. H and J) *Fh1* burst frequencies. I and K) *Fh1* burst sizes. An i at the end represents that macrophages were pre-treated with IFN- γ for 24h. Burst frequencies p-values from Mantel-Haenszel test with the Bonferroni method for multiple testing correction. TSS intensity p-values from Tukey HSD test (comparing all burst detected using the design formula $\sim Sample + Replicate$). Each sample has between 3-4 biological replicates.

Genotype	Treatment	Replicates	Cells	<i>Hprt</i>	<i>Hprt</i>	<i>Fh1</i>	<i>Fh1</i>
				Burst Fraction	TSS intensity	Burst Fraction	TSS intensity
WT	Untreated	3	3021	0.29	29.05	0.11	2.15
<i>Rad21</i> KO	Untreated	3	3060	0.31	36.83	0.08	2.35
WT	8h LPS	4	3016	0.17	26.57	0.06	1.75
<i>Rad21</i> KO	8h LPS	4	5028	0.26	49.32	0.14	2.55
WT	24h IFN	4	2549	0.36	30.47	0.14	2.02
<i>Rad21</i> KO	24h IFN	4	4153	0.37	45.69	0.12	2.59
WT	24h IFN + 8h LPS	4	3242	0.23	36.87	0.10	2.08
<i>Rad21</i> KO	24h IFN + 8h LPS	3	2782	0.25	52.07	0.07	2.72

Table 3.5: ***Hprt* and *Fh1* transcriptional burst measures in WT and *Rad21* KO macrophages.** Replicates, number of replicates. Cells, number of cells analysed by FISH in three biological replicates. Burst fraction, mean burst fraction across smFISH replicates. TSS intensity, mean normalised TSS intensity across smFISH replicates.

(p_{adj} : 0.88, $\log_2\text{FoldChange}$: 0.33). This indicates that processes other than transcription were responsible for the up-regulation of total *Fh1* (possibly, through stabilisation of its mRNA). In fact, WT showed higher burst frequencies than *Rad21* KO macrophages and no significant differences in TSS intensities were found at baseline (Figure 3.8H and I).

LPS-activation or IFN- γ pre-treatment had no effect in the *Fh1* TSS intensities of WT macrophages. However, as with *Hprt*, TSS intensities of *Fh1* increased in *Rad21* KO upon LPS-activation or IFN- γ priming (Figure 3.8K). LPS-activation decreases *Fh1* burst frequencies of WT macrophages; in contrast, *Rad21* KO *Fh1* TSS intensities increase. It was only if *Rad21* KO cells had been pre-treated with IFN- γ when the reduction in burst frequency of *Fh1* occurred in *Rad21* KO macrophages (Figure 3.8J).

In summary, the conclusions about *Fh1* were: 1) *Fh1* LPS-repression occurred through a reduction in its burst frequency; 2) *Rad21* KO macrophages failed to decrease *Fh1* burst frequency upon LPS-activation unless they were pre-treated with IFN- γ ; 3) LPS-activation or IFN- γ increased TSS intensities only in *Rad21* KO macrophages.

The *Spred2-Sertad2-Peli1* locus

The *Spred2-Sertad2-Peli1* locus is a 2Mb A compartment region composed of several contact domains hierarchically interconnected. Within this region, there are 3 DE genes, 2 inducible genes (*Sertad2* and *Peli1*), and *Spred2* (Figure 3.9). According to our RNAseq data, both inducible genes have lower levels of expression upon LPS-activation in *Rad21* KO macrophages, whereas *Spred2* had higher levels in *Rad21* KO macrophages compared to WT in response to LPS. We hypothesised that LPS-activated regulatory elements that interact with the promoters of the inducible genes might establish aberrant interactions with *Spred2* in the absence of cohesin.

Spred2 and *Peli1* followed the same DE pattern in another cohesin deficient macrophage system where the RAD21 subunit is degraded proteolytically (Cuartero et al. 2018). Therefore, it is likely that both genes are directly regulated by cohesin during macrophages activation by microbial signals.

The order of genes in the *Spred2-Sertad2-Peli1* genomic region is conserved from higher vertebrates to coelacanth, but not in *Drosophila* or Zebrafish (UCSC reference genomes). Coelacanth is a fish, but very closely related to tetrapods. Therefore, tetrapods have preserved the synteny of this region for at least 360 million years.

smFISH detection of eRNA

Whereas gene expression is unidirectional, active enhancers are transcribed bidirectionally. Enhancer transcription reflects enhancer activity (Kaikkonen et al. 2013; Kim et al. 2010), allowing us to use enhancer transcription as a proxy of enhancer activation.

The genomic region between *Peli1* and *Sertad2*, harbours an inducible SE (Enh), which is induced within 1h of LPS treatment. GROseq showed three foci of bidirectional Enh transcription upon macrophage activation, and we designed smFISH probes specific for the strongest Enh TSS. As expected, we detected an increase in Enh expression upon LPS-activation in both WT and *Rad21* KO macrophages (Figure 3.10).

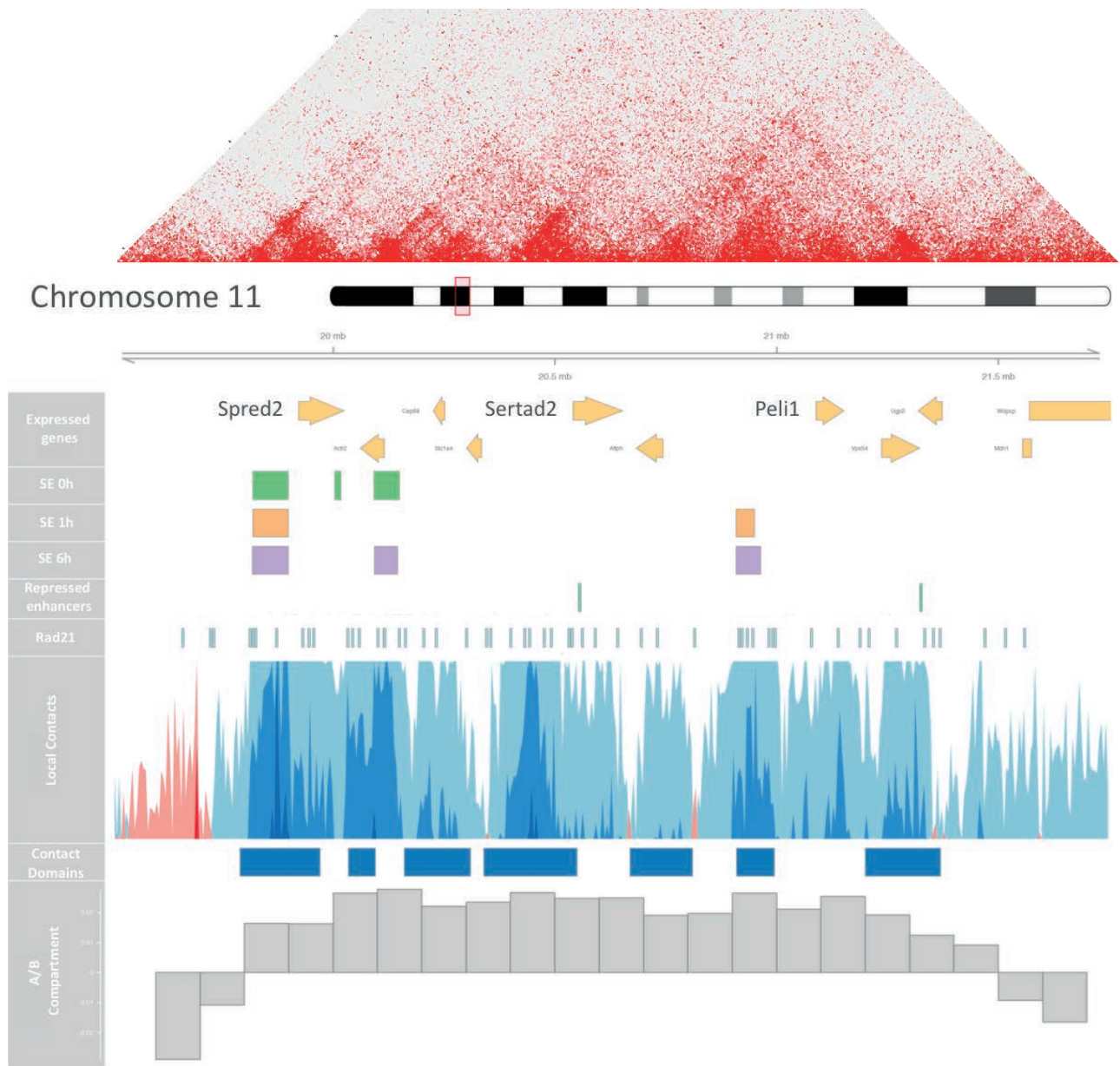


Figure 3.9: **The *Spred2-Sertad2-Peli1* genomic region.** Expressed genes: Expressed genes in macrophages. SE, super-enhancers. 0h, 1h, 6h, hours after LPS induction. Rad21, Rad21 ChIP peaks in WT untreated macrophages. Expressed genes, SE and Rad21 peak data from Cuartero et al. 2018. Repressed macrophage enhancers from Ostuni et al. 2013. A and B compartments were distinguished by gene density (UCSC refGene). HiC and contact domains were called as described (Rao et al. 2017) by Ya Guo from mouse BMD macrophages Hi-C data (GSE115524, Mumbach et al. 2019) processed using HiC-Pro (v2.7.8), and visualised with JuiceBox. Normalisation set as "balanced". Genomic features were visualised using the R package GViz.

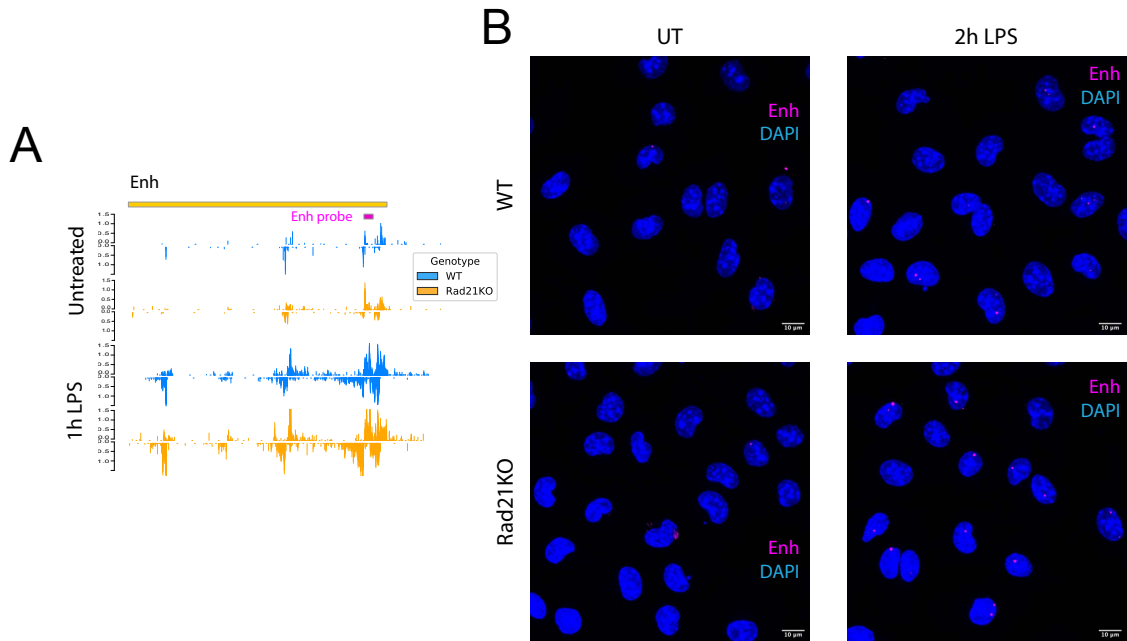


Figure 3.10: **Enh eRNA detection.** A) Macrophage GROseq alignments over the enhancer Enh region and Enh probe mapping. B) Example images of WT and *Rad21* KO macrophages bursting from Enh (SP5 Leica microscope x63 oil objective, zoom x2). Scale bar represents 10 μm .

Enh eRNA bursting

Enh increases burst frequency, not TSS intensity upon LPS-activation.

Enh eRNA increased its expression through burst frequency in activated macrophages. Resting macrophages have a relatively low Enh eRNA burst frequency (0.024 in untreated WT macrophages). Burst frequency of Enh increases by almost a factor of 10 when LPS activates macrophages (0.227 in WT 2h after LPS) (Figure 3.11B). In contrast, LPS-activation had no impact in Enh TSS intensities (Figure 3.11C). See Table 3.6 for a summary of the Enh bursting measures.

Enh showed a slight increase in burst frequency in *Rad21* KO macrophages

GRO-seq showed no statistically significant differences between Enh transcription in WT and *Rad21* KO macrophages upon LPS (enhancer mm9 chr11:20,946,477-20,950,15, log2FoldChange: -0.056, padj: 0.978). However, the high number of cells analysed revealed that *Rad21* KO have a slight increase in burst frequencies in untreated and LPS-activated macrophages compared to

WT (Figure 3.11B). Only a minor difference in TSS intensities was detected after LPS (Figure 3.11C).

Inducible genes are mainly regulated by burst frequency upon LPS activation

As reported previously in chapter 2, the inducible genes *Peli1* and *Sertad2* increased their expression in activated macrophages through changes in burst frequencies. In WT macrophages, *Peli1* went from a burst frequency of 0.186 to 0.664. *Sertad2* went from a burst frequency of 0.122 to 0.384 (Figure 3.11B). Furthermore, *Peli1* and *Sertad2* burst were slightly brighter upon LPS-activation. *Peli1* displayed a fold-change of 1.19 and *Sertad2* displayed a fold-change of

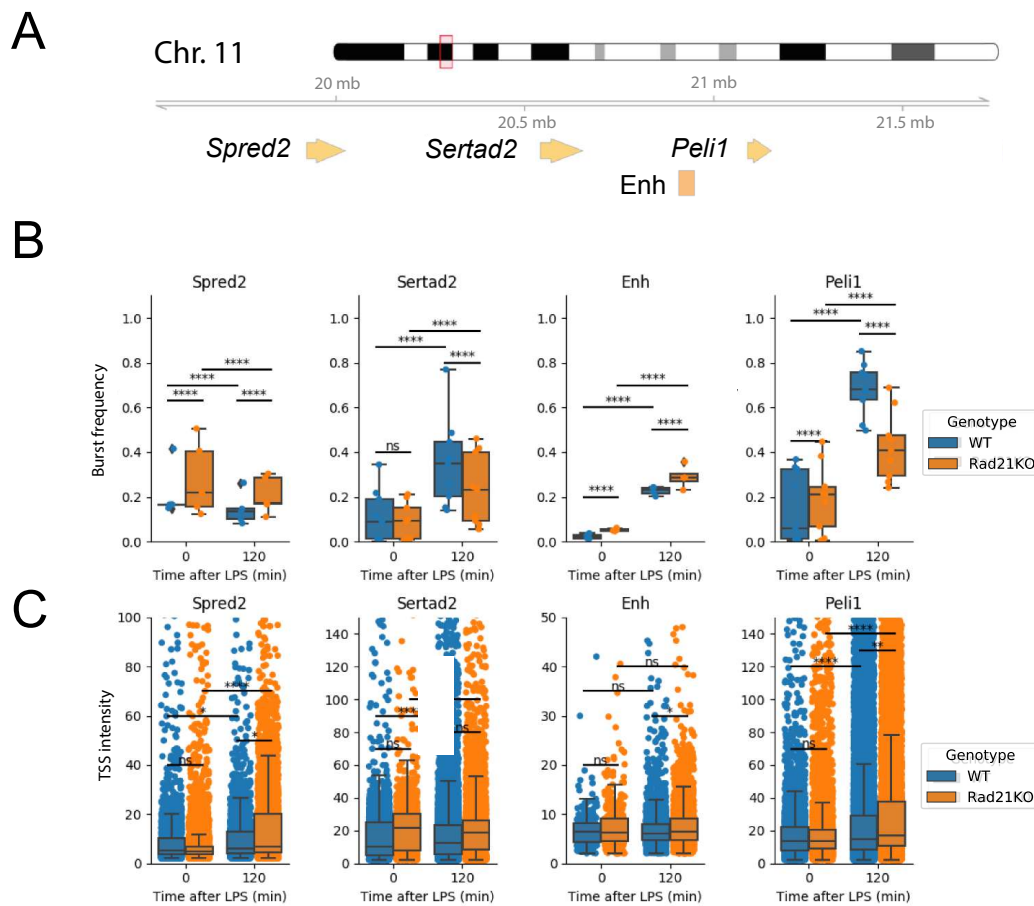


Figure 3.11: **Burst fraction and burst frequency of enhancer Enh and locus genes in WT and *Rad21* KO macrophages.** A) Localisation of genes in the locus. B) Burst frequencies. Statistical test: Cochran-Mantel-Haenszel Test (replicate as strata) with the Bonferroni method for multiple testing correction. C) TSS intensity statistical test: Tukey HSD test (ANOVA design formula $\sim Sample + Replicate$).

Genotype	Target	Treatment	Replicates	Cells	Burst Fraction	TSS intensity
WT	Enh	Untreated	4	3897	0.02	10.15
<i>Rad21</i> KO	Enh	Untreated	4	3611	0.05	9.96
WT	Enh	2h LPS	4	2914	0.23	9.07
<i>Rad21</i> KO	Enh	2h LPS	4	4650	0.29	12.66
WT	<i>Peli1</i>	Untreated	9	13281	0.16	20.05
<i>Rad21</i> KO	<i>Peli1</i>	Untreated	9	10189	0.16	23.68
WT	<i>Peli1</i>	2h LPS	9	15236	0.72	24.96
<i>Rad21</i> KO	<i>Peli1</i>	2h LPS	9	14290	0.32	29.68
WT	<i>Sertad2</i>	Untreated	9	13281	0.11	13.66
<i>Rad21</i> KO	<i>Sertad2</i>	Untreated	9	10189	0.09	19.56
WT	<i>Sertad2</i>	2h LPS	9	15236	0.35	26.43
<i>Rad21</i> KO	<i>Sertad2</i>	2h LPS	9	14290	0.23	25.39
WT	<i>Spred2</i>	Untreated	5	9384	0.21	15.58
<i>Rad21</i> KO	<i>Spred2</i>	Untreated	5	6578	0.28	16.43
WT	<i>Spred2</i>	2h LPS	5	10270	0.15	20.16
<i>Rad21</i> KO	<i>Spred2</i>	2h LPS	5	9640	0.21	20.46

Table 3.6: **TSS intensities and burst frequencies measured at the *Spred2-Sertad2-Peli1* locus in WT and *Rad21* KO macrophages.** Replicates, number of replicates. Cells, number of cells analysed. Burst fraction, mean burst fraction across smFISH replicates. TSS intensity, mean TSS intensity across smFISH replicates.

1.05. These differences are minor compared to the differences in burst frequencies but they are statistically significant (Figure 3.11C). See Table 3.6 for a summary of the individual bursting measures.

***Rad21* KO macrophages fail to increase the burst frequency of inducible genes upon LPS-activation**

Despite Enh showing an increased bursting frequency in *Rad21* KO macrophages, surrounding inducible genes displayed reduced bursting frequencies compared to WT. In response to LPS, WT macrophages increased burst frequencies for *Peli1* by a fold-change of 3.58, and for *Sertad2* by 3.18. In contrast, *Rad21* KO macrophages increased burst frequencies for *Peli1* by only 1.81-fold and for *Sertad2* by 1.97 (Figure 3.11B, Table 3.6). *Sertad2* showed no difference in TSS intensity between *Rad21* KO and WT macrophages. *Rad21* KO macrophages presented slightly brighter *Peli1* bursts than WT in LPS-activated macrophages. Although statistically significant, the fold-change is minimal, 1.06.

***Spred2* burst frequency is reduced upon LPS-activation.**

In WT macrophages, the burst frequency of *Spred2* was repressed upon LPS-activation and dropped from (Figure 3.11B). In contrast, TSS intensities increase slightly but significantly (Figure 3.11C). See Table 3.6 for a summary of the the burst measures.

***Spred2* burst frequency is higher in *Rad21* KO macrophages**

Spred2 burst frequency is always higher in *Rad21* KO macrophages. As in WT, *Spred2* repression upon LPS-activation occurs in *Rad21* KO macrophages, and the fold-change between untreated and treated samples is similar to WT (0.75 in WT and 0.63 in *Rad21* KO, Figure 3.11B). TSS intensity increase are around $\sim 25\%$ brighter upon LPS in WT and *Rad21* KO macrophages.

Burst co-dependence

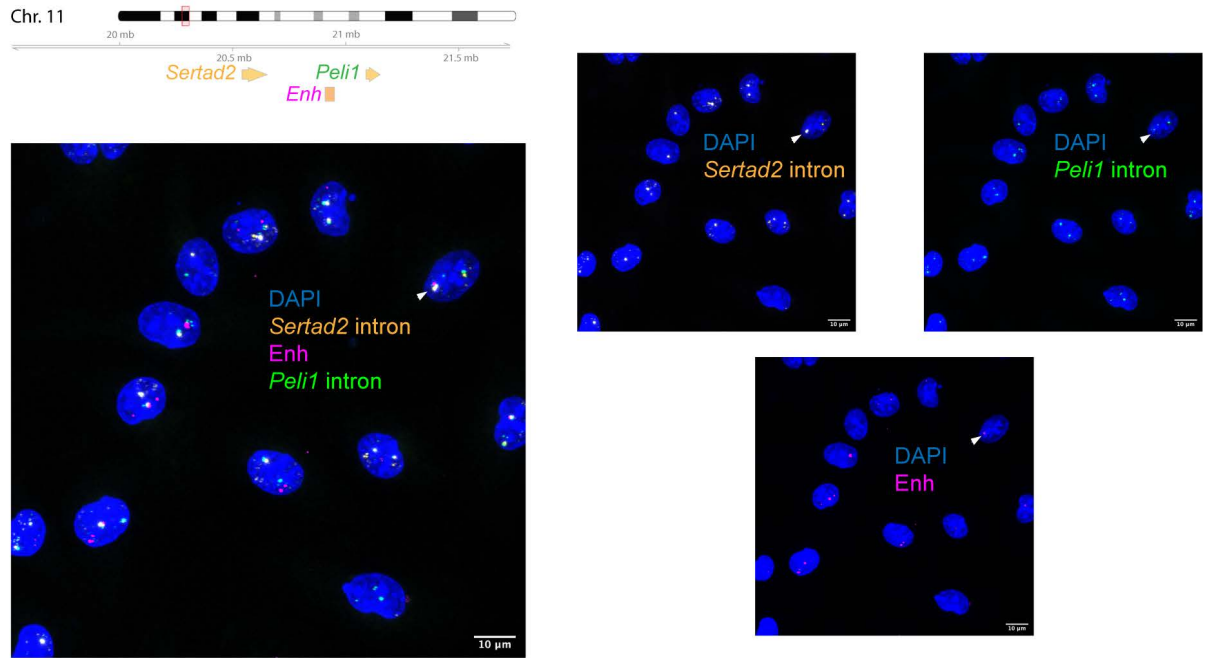
To evaluate burst co-dependencies, we combined the intronic probes of *Peli1* and *Sertad2* to either the Enh probe or the *Spred2* intronic probe. In both cases, we were able to detect co-localization of bursts (Figure 3.12).

Assigning multiple TSSs to a single locus

To assign TSSs to different chromosomes in individual cells, I wrote an algorithm in Julia with the help of Dr Dirk Dormann that starts with 3D stacks and the FQ TSSs outlines. The algorithm measures the distance between the brightest pixel of each gene TSS and all the other gene TSSs in a given cell. Because macrophages are quiescent, we assume that cells are diploid and call the first locus with bursts locus 1, and the second locus 2. TSSs of different genes are assigned to the same locus unless the distance between them is larger than $5\text{ }\mu\text{m}$, which corresponds to the typical distance between two TSSs from the same gene in our data (around $5\text{-}6\mu\text{m}$).

Once the first pair is assigned, the second closest TSSs pair is assigned. If one of the loci has already one of the TSSs, then the other TSS is ascribed to the same locus, unless the distance

A



B

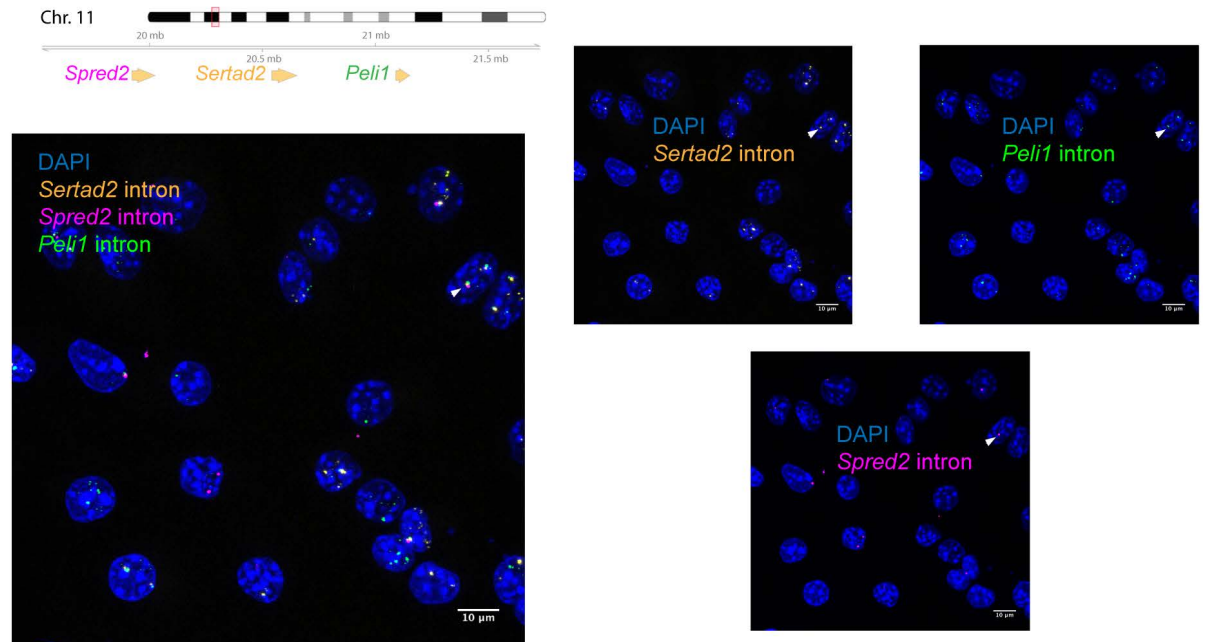


Figure 3.12: **Locus burst co-localisation.** A) Co-localisation of the inducible enhancer Enh and the inducible genes, *Sertad2* and *Peli1*. A total of 17,124 cells ($\sim 1,000$ cells per sample) were imaged from 2 technical and 4 biological replicates. B) Co-localisation of the inducible genes *Sertad2* and *Peli1*, and *Spred2*. A total of 35,872 cells ($\sim 1,500$ cells per sample) were imaged from 3 technical and 5 biological replicates. White arrow indicates a locus where the three burst were detected at once. Scale bars represent $10 \mu\text{m}$.

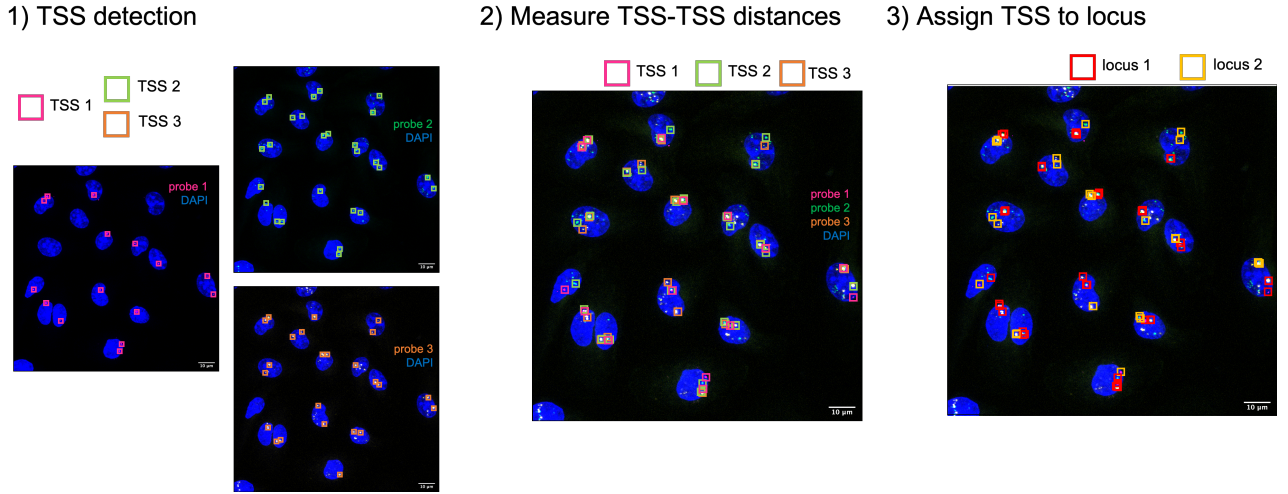


Figure 3.13: **Assigning TSSs to locus 1 or 2.** First, TSSs are detected by FQ. Second, for each cell, the distances between a gene TSS and all the other gene TSSs detected are measured. Third, TSSs are assigned to different locus based on their proximity. A locus can only have one burst from each gene, and if two TSSs are further than 5 μm , they are assigned to different locus. Our algorithm assumes that cells are diploid. Scale bars represent 10 μm .

between them is larger than the threshold. However, there are some exceptions. A locus can only have one burst of each gene. So, if a cell has two TSSs of a gene, each one of them will automatically belong to different loci (Figure 3.13).

TSS-TSS distances

Active TSS-TSS distances remain unchanged upon LPS-activation

In the DNA sequence, Enh is closer to *Peli1* (~ 40 kb) than to *Sertad2* (~ 550 kb). As expected, we detected that Enh burst distance to *Peli1* was on average 0.31 μm shorter than to *Sertad2* (t-test, p-value: $<e-7$). Interestingly, the distance between *Peli1-Sertad2* pairs was also shorter than *Sertad2-Enh*, 0.25 μm shorter on average (t-test, p-value: $<e-5$). However, we detected no differences in TSS burst distances between WT untreated and LPS-activated macrophages (Figure 3.14).

Gene-enhancer burst distances are similar between WT and Rad2KO, but gene-gene distances are generally longer in *Rad21* KO

We detected no statistically significant differences between bursts from the enhancer Enh and inducible genes between WT and *Rad21* KO. However, we detected some distance changes upon LPS-activation in *Rad21* KO that were not detected in WT. Specifically, the distances between *Peli1* burst and the other two genes (*Spred2* and *Sertad2*) were slightly shorter (Figure 3.14). In general, distances between *Peli1* burst, *Spred2* and *Sertad2* TSS burst are larger in *Rad21* KO macrophages than in WT cells (Figure 3.14). It is also important to mention the limitations of our TSS-TSS distances. They depend on our locus association algorithm, which is automatic, but not perfect. Sometimes manual TSS identification and association do not match the algorithm results. Distance measures are dependent on our voxel size (189x189x500) and on the assumption that the highest RNA accumulation (brightest pixel) correspond exactly

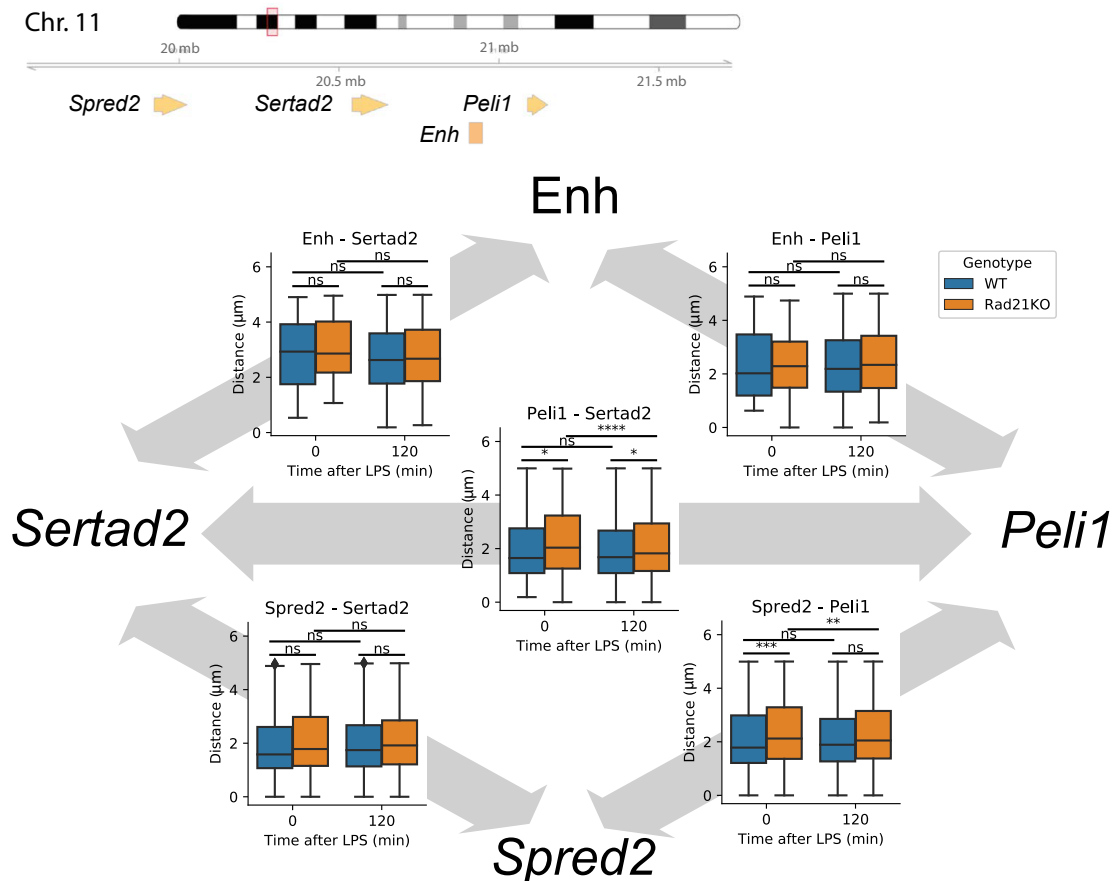


Figure 3.14: **Distances between detected active TSSs.** TSS-TSS distance measured when one cell locus is bursting more than one gene or enhancer at the same time (*Spred2*, *Sertad2*, *Enh* or *Peli1*). Statistical test: Tukey HSD Test (ANOVA design formula $\sim Sample + Replicate$).

to the DNA TSS location. As it is not biased, the algorithm can be used to make comparisons, but TSS-TSS distances are often overestimated.

Co-dependence between burst frequencies

To test whether the burst frequencies are co-dependent or mutually exclusive, we assumed that for a given experiment, the probability of sampling a burst follows a Bernoulli distribution, with $p = \text{Burst Fraction}$. If we have 2 alleles, we have p_1 and p_2 for each. If alleles burst independently, the probability of sampling both at the same locus is:

$$p_{1and2} = p_1 * p_2 \quad (3.1)$$

And it follows a Bernoulli probability distribution with $p = p_{1and2}$ in our sample.

For a statistical test, our null hypothesis is that genes burst independently, and our alternative hypothesis is that they do not. To calculate a pvalue, we sampled co-bursting frequencies given the number of cells sampled. We assume that p_1 and p_2 were equal. To do so, we simulate 10,000 experiments where bursts are independent and calculate burst co-occurrence in each of them. These co-occurrences construct the probability distribution of the null hypothesis (Figure 3.15A).

We then examined burst co-occurrences in our two set of experiments. In total, we have 4 replicates for *Enh-Sertad2-Peli1* and 5 for *Spred2-Sertad2-Peli1*. For consistency, we calculate and tested co-occurrences in each replicate independently.

Enh and the two inducible gene burst frequencies are co-dependent

Peli1 and *Enh* tend to burst together. In untreated WT macrophages, the probability of sampling them together is twice the probability of sampling them by chance. This is statistically significant overall (p-value of 0.0178, considering all *Peli1* and *Enh* bursts in WT untreated macrophages) and in 2 out of our 4 replicates. In the other two where it is not significant, it is due to the lack of *Enh* burst in untreated samples (Figure 3.15B). In LPS-activated macrophages, *Enh* and *Peli1* are also codependent (p-value of > 0.001 , considering

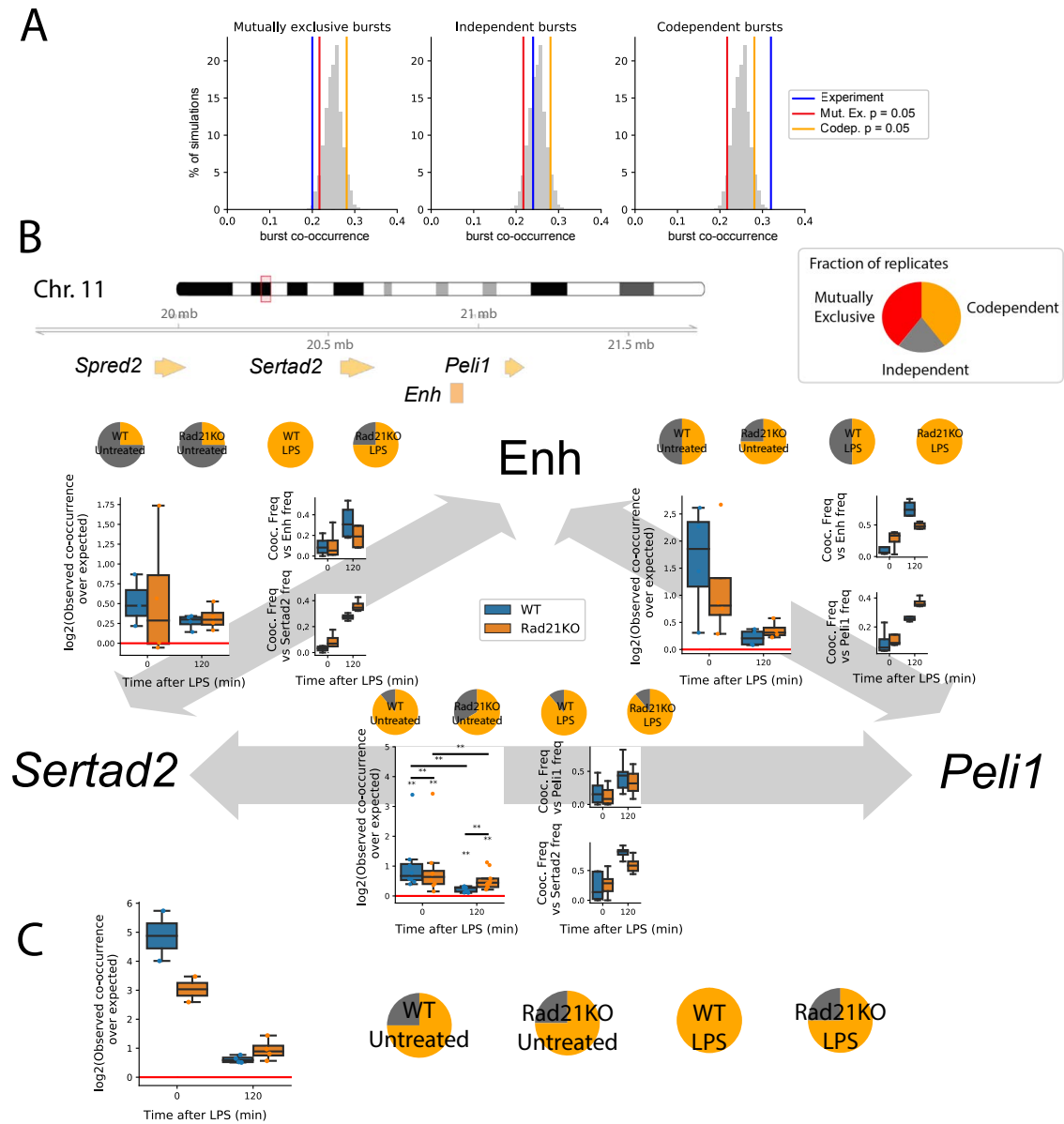


Figure 3.15: Co-dependence of burst frequencies. A) Graphical example of the statistical test that assessed whether burst frequencies are mutually exclusive, independent or co-dependent. Null hypothesis: gene burst independently. Burst fractions: $p_1 = 0.5$ and $p_2 = 0.5$. In grey, probability distribution constructed from 10,000 simulations of 500 loci sampled each. Blue line: burst co-occurrence sampled in an experiment. Red line: burst co-occurrence for mutually exclusive bursts at a p-value of 0.05. Orange line: burst co-occurrence for co-dependent bursts at a p-value of 0.05. B) Burst frequency coordination analysis by locus. Pie charts represent the fraction of replicates in which bursts are mutually exclusive, co-dependent or independent. Big boxplots: log2 co-occurrence of two bursts over expected co-occurrence, statistical significance: Wilcoxon rank test with Bonferroni correction for multiple testing. Stars at the top of the boxplots indicate whether the log2 co-occurrence ratio is different from 0, between boxplots that the mean difference is different from 0. Small boxplots: the fraction of times were a gene burst is detected with the other gene. C) Same as B but for the coordination of the three genes at once.

all *Peli1* and Enh bursts in WT LPS-activated macrophages). This is only significant in 2 out of our four replicates, when looking at each replicate independently. However, those samples displayed burst frequencies very close to 1 for *Peli1* (0.85 and 0.78), which reduce our power to detect co-dependencies, even though insignificant, their co-dependency p-values are relatively low (0.06, and 0.1). The situation is similar but less evident for *Sertad2*. In this case, we also sample *Sertad2* and Enh together around 20% more times than by chance in WT LPS-activated macrophages. Moreover, the probability of sampling them together is twice than random expectation (Figure 3.15C).

Enh bursts associate to *Peli1* or *Sertad2* bursts more often in WT than in *Rad21* KO macrophages

In LPS-activated *Rad21* KO macrophages, the co-dependency between Enh and *Peli1* is more evident because the burst frequency of *Peli1* is not as high as WT (Figure 3.11B). Nevertheless, in LPS-activated samples, 75.9% of Enh bursts on average are accompanied by a *Peli1* burst in WT, only 48.7% of Enh do so in *Rad21* KO. *Sertad2* displayed similar results. In this case, in LPS-activated WT cells, 33.1% of Enh bursts on average are accompanied by a *Sertad2* burst, only 18.9% of Enh do so in *Rad21* KO.

General features of TSS intensities and frequencies

Expression changes between *Rad21* KO and WT are explained by changes in burst frequencies

To identify whether changes in bursting dynamics could explain changes in gene expression, we compared our burst measures to fold changes by RNAseq in our 8 genes (*Spred2*, *Sertad2*, *Peli1*, *Actb*, *Hrpt*, *Fh1*, *Ifnb1*, *Il12b*) across 4 different conditions (Untreated, 2h LPS, 8h LPS and IFN- γ + 8h LPS). log2 fold changes in burst frequency correlated to log2 fold changes by RNAseq (Figure 3.16A, Spearman rho: 0.8, pvalue < 0.001), but not TSS intensities (Figure 3.16A, Spearman rho: 0.38, pvalue: 0.1293).

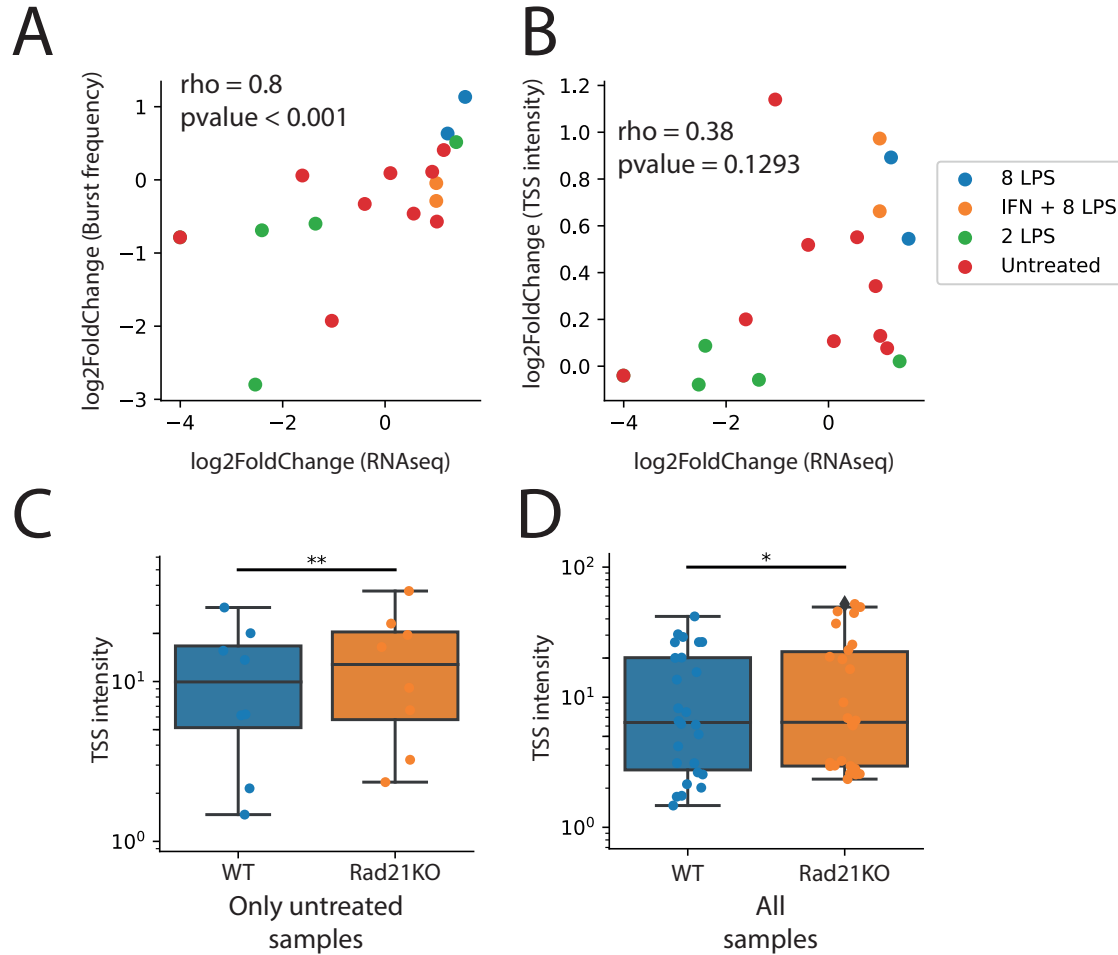


Figure 3.16: Expression changes between *Rad21* KO and WT are explained by changes in burst frequencies, but *Rad21* KO displayed bigger TSS intensities in the genes studied. Correlation between $\log_2$ fold changes by RNAseq (*Rad21* KO vs WT) and $\log_2$ fold change (*Rad21* KO vs WT) in A) burst frequency or B) TSS intensity (mean in *Rad21* KO divided by mean in WT, for all the conditions where RNAseq data is available). Rho from Spearman correlation, pvalue from R cor.test function. C) TSS intensities in untreated macrophages (mean of mean TSS intensity across replicates). D) TSS intensities in all conditions measured (mean of mean TSS intensity across replicates). smFISH data from 8 genes (*Spred2*, *Sertad2*, *Peli1*, *Actb*, *Hrpt*, *Fh1*, *Ifnb1*, *Il12b*). Statistical test for C and D: Wilcoxon Signed Rank test as implemented in HypothesisTest.jl.

Overall, *Rad21* KO macrophages displayed bigger TSS intensities

From the previous analysis, we realised that the fold changes between TSS intensities were always greater than 1 in *Rad21* KO macrophages. Per gene, TSS intensities could only explain expression changes for *Fh1* and *Hprt*, and mainly in activated macrophages (Figure 3.8). However, in general, *Rad21* KO macrophages presented bigger TSS intensities in untreated samples (Figure 3.16C), and also if you compare all the TSS intensities measured in all conditions (Figure 3.16D).

Distance to a TAD boundary

Recently, a link between burst frequency control of boundary-associated genes and cohesin has been established (Luppino et al. 2020). They used a synchronised cancer human cell line where a cohesin subunit can be proteolytically degraded via an auxin-degron to study conformation and gene expression changes upon acute cohesin depletion. Boundary-associated genes were more likely to be DE, and in 13 out of 14 cases studied, burst frequency differences explained expression changes. Using DNA FISH technology, they visualised the TSSs of the boundary-associated genes *MCM5* and *CREBL2*, and their neighbouring domains, observing in both cases a decrease in intermingling in the absence of cohesin and an increase in intermingling when WAPL or CTCF were depleted. Therefore, they suggested that boundary-associated genes rely on cohesin for their regulation because it increases their contacts with TADs across domain boundaries.

Long-term cohesin depletion does not associate with boundaries

In Luppino et al. system, the fold change difference between control and cohesin deficient cells anti-correlated with distance to the closest boundary. In our *Rad21* KO macrophages, it does not. We did not find a negative Spearman correlation between absolute log2FoldChanges and distance to the closest TAD boundary (Figure 3.17A), and in *Rad21* KO macrophages, there is a weak association between distance to a boundary and the likelihood of being DE. However, it is the opposite of what Luppino et al. found, genes far away from a boundary are more

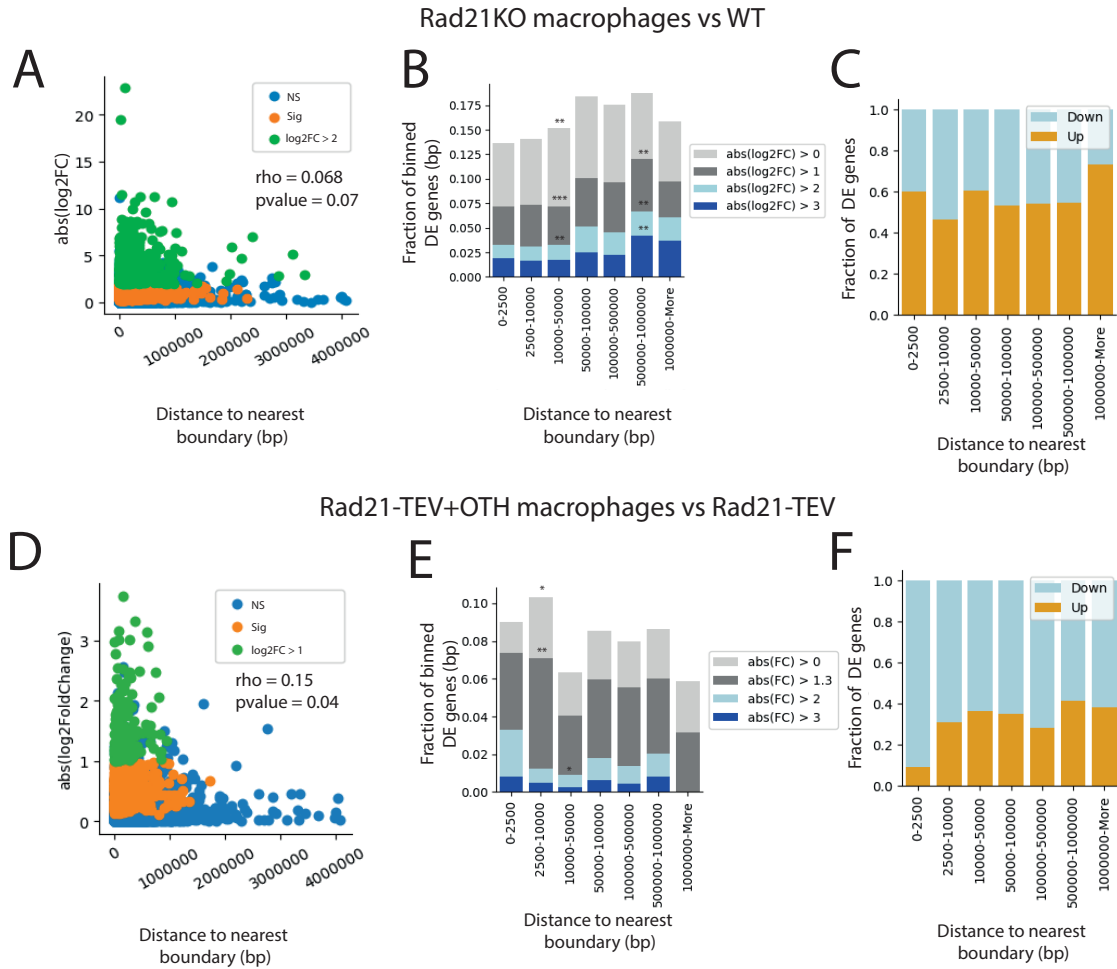


Figure 3.17: Long and short-term cohesin depletion display different associations between DE genes and distance to a TAD boundary. A and D) $\log_2\text{FoldChange}$ of differentially expressed genes vs distance between their TSS to the closest TAD boundary, ρ , Spearman rank correlation, p value calculated in R. B and E) Percentage of expressed genes at binned distances to the closest TAD boundary that are DE in *Rad21* KO macrophages, stars correspond to the p -value of a Fisher exact test. C and F) Fraction of genes that are up-regulated (Up) or down-regulated (Down) at binned distances to the closest TAD boundary. TADs were called as described (Rao et al. 2017) by Ya Guo from mouse BMD macrophages Hi-C data (GSE115524, Mumbach et al. 2019). DEseq2 data for A, B and C: *Rad21* KO macrophages vs WT. DEseq2 data for D, E and F: Rad21-TEV macrophages treated with 4-OHT, which activates the TEV protease and cleaves RAD21-TEV, compared to Rad21-TEV macrophages without 4-OHT. A gene is considered DE if $\text{padj} < 0.05$. TAD were called as described (Rao et al. 2017) by Ya Guo from mouse BMD macrophages Hi-C data (GSE115524, Mumbach et al. 2019).

susceptible to be DE (Figure 3.17B). Besides, they observed that the closer genes were to a boundary the more likely they were down-regulated, which we did not observe (Figure 3.17C).

Short-term cohesin depletion associates with boundaries

There is a significant difference between Luppino et al. system and our system that is the time cells spend without functional levels of cohesin. Our genetic depletion starts 3 days before we start experiments, in their system proteolytic depletion of cohesin starts 6 hours before they perform experiments. Therefore, they could detect the first effects of cohesin depletion, whereas we detect the long-term consequences for gene regulation. In the lab, we had macrophage data from a RAD21-TEV system (Cuartero et al. 2018), where the RAD21 subunit is cleaved by TEV protease when 4-OTH is added to the medium. After 8h of Rad21TEV proteolytic depletion, 1034 genes were DE. log2FoldChanges between genetically and proteolytically depleted cells correlate considering all DE genes in either of the systems (3119 genes, Spearman correlation: 0.45, p-value < 2.2e-16, R's cor.test) and especially in genes DE in both systems (353 genes, Spearman correlation: 0.779, p-value < 2.2e-16, R's cor.test). In this system of acute cohesin depletion, we found similar relations to DE and boundaries as in Luppino et al., like the association between DE genes and TAD boundaries (Figure 3.17E) and preferential down-regulation of DE genes at boundaries (Figure 3.17F), but not a negative Spearman correlation between distance to a boundary and log2FoldChange (Figure 3.17D).

If we look at the genes involved in LPS response, they do not display any particular preference for boundary position at baseline (Figure 3.18A and C), but after activation, the most likely genes to be DE between *Rad21* KO macrophages and WT are further from a boundary (Figure 3.18B and D).

scRNAseq analysis

smFISH provides a very detailed low-dimensional picture of single-cell expression constrained by the number of probes used. scRNAseq mRNA detection is less accurate, but many more genes can be analysed at once. Far from competing, these techniques can complement each other to understand cell-population heterogeneity. Most LPS-inducible genes appear bimodal

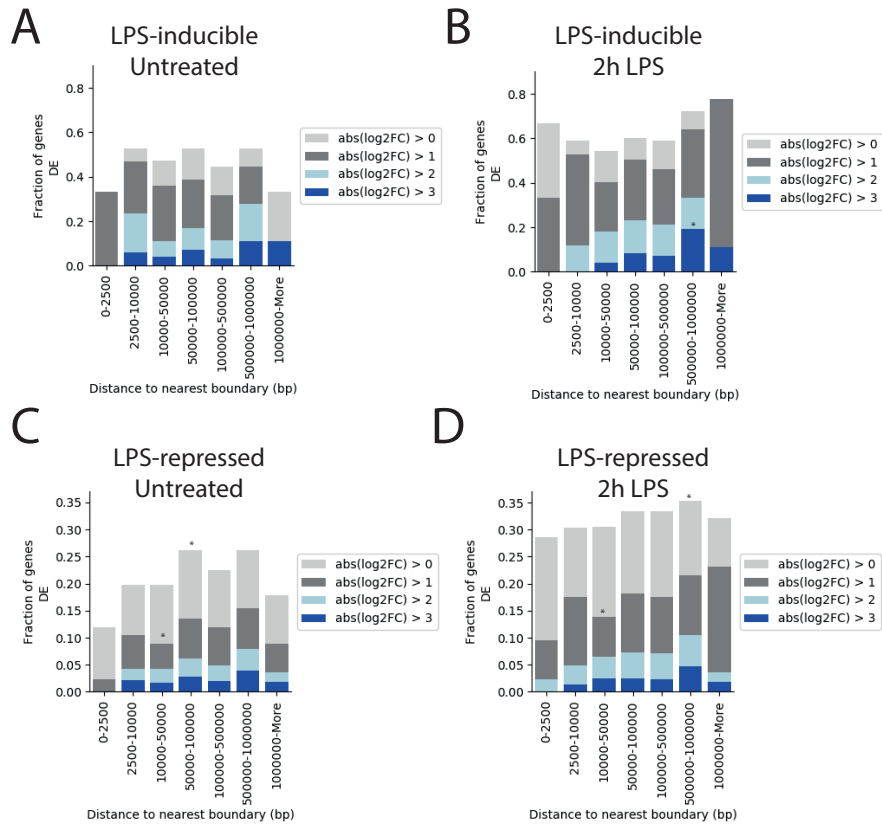


Figure 3.18: **Long and short cohesin depletion display different associations between differentially expressed genes and distance to a TAD boundary.** Percentage of expressed genes at binned distances that are differentially expressed. stars correspond to the p-value of two-sided Fisher exact test in: A) LPS-inducible genes *Rad21* KO untreated macrophages vs untreated WT. B) LPS-inducible genes *Rad21* KO LPS-activated macrophages vs LPS-activated WT. C) LPS-repressed genes *Rad21* KO untreated macrophages vs untreated WT. D) LPS-repressed genes *Rad21* KO LPS-activated macrophages vs LPS-activated WT. A gene is considered DE if $\text{padj} < 0.05$.

(Shalek et al. 2013) to a certain extent, which can be caused by transcriptional bursting per se (gene expression in bimodal in the steady-state), changes in bursting parameters (non-steady-state expression) or the presence of hidden cell states. Besides, as cohesin in the control of burst frequencies and low burst frequencies and big burst sizes associate to increased variability, it could be expected a change in cell-to-cell variability in *Rad21* KO macrophages. Therefore, we aimed to analyse variability and whether the *Rad21* KO decrease in LPS-inducible expression was general or restricted to a subpopulation.

There was scRNAseq data available in the lab. WT and *Rad21* KO macrophages after 0, 2 or 8 hours of LPS induction were FACS sorted and cryopreserved by Dr Sergi Cuartero. Then, cells were sequenced and aligned in the CNAG-CRG using the MARS-Seq methodology (Jaitin

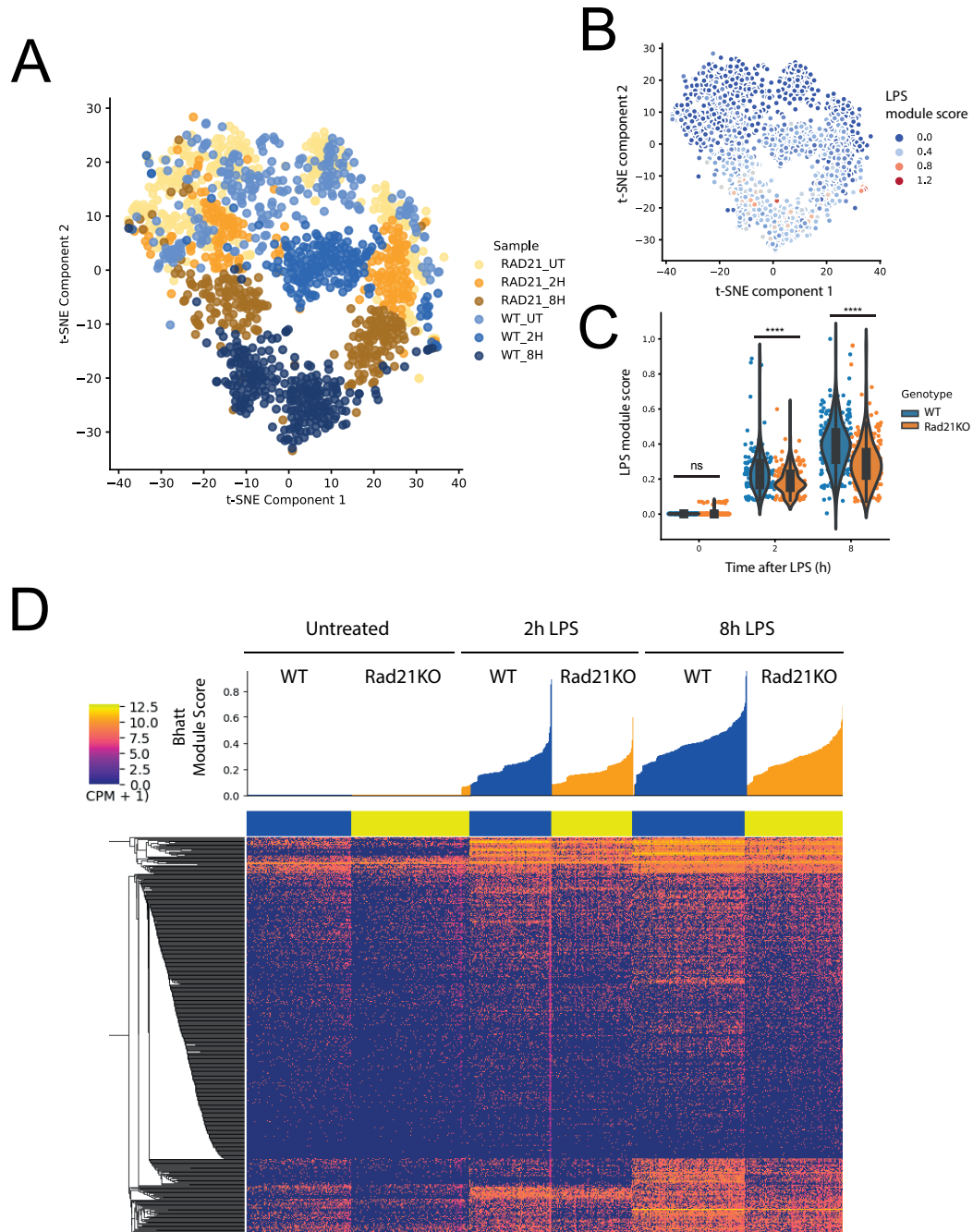


Figure 3.19: **Reduction in LPS response is widespread in *Rad21* KO macrophages.** A) scRNAseq data. tSNE of all genes CPMs detected in more than 20 cells, perplexity set as 50. B) LPS module score in each cell. C) LPS-inducible genes module score distribution in each sample. Statistical test: Tukey HSD Test (ANOVA design formula $\sim Genotype : Timepoint$). D) Time-course expression profiles for LPS-induced genes. Colour indicates $\ln(CPM + 1)$.

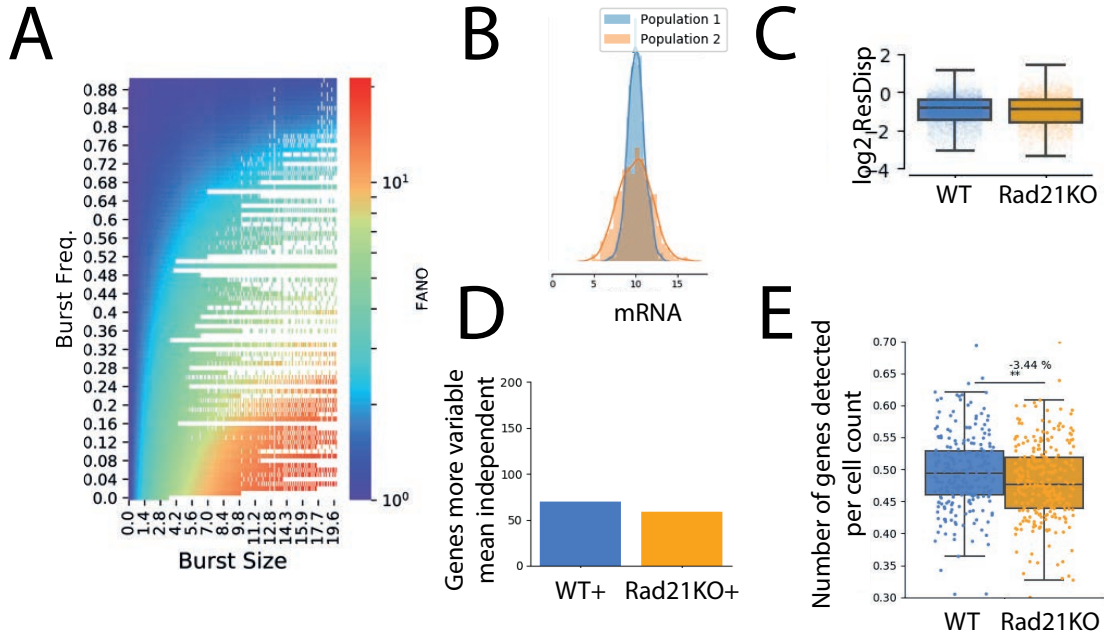


Figure 3.20: ***Rad21* KO macrophages did not show an increase in gene overdispersions.** A) High variability associates with low burst frequencies and high burst sizes. Fano factor of 1,030,301 simulations of the telegraph model at steady state. Pixel colour corresponds to the median value in cases where two simulations share burst size and frequency. See *Materials and Methods* for details. B) scRNAseq variability analysis focus on genes with no differences in mean expression. C) $\log_2$ residual dispersion of genes with no average difference between WT and *Rad21* KO. D) Number of genes differentially variable in untreated macrophages. E) *Rad21* KO macrophages have fewer genes detected per cell count than WT. Statistical test: Equal variance t-test from Julia package HypothesisTests.jl %, per cent change in *Rad21* KO.

et al. 2014).

***Rad21* KO defects in LPS-inducible expression are general**

The first question we addressed was whether the reduction in LPS-inducible expression in *Rad21* KO macrophages was a general or driven by a fraction of macrophages not responding to LPS. To measure total LPS-inducible expression per cell, we adapted an inducible gene module score developed in Shalek et al. 2014 (See *Materials and Methods* for details). Although *Rad21* KO cells at 2h post LPS segregate in 2 groups in a dimensionality reduced space, (Figure 3.19A), visual inspection of the overlayed LPS-inducible scores did not display any noticeable difference (Figure 3.19B), neither visual inspection of individual gene expression in each cell (Figure 3.19D). We detected an increase in LPS-inducible scores as we increased incubation

time with LPS. WT cells expressed higher levels of LPS-inducible scores compared to *Rad21* KO upon LPS-activation, but not at baseline. *Rad21* KO inducible scores grouped around the mean; therefore, the reduction in LPS-inducible expression is widespread in *Rad21* KO macrophages (Figure 3.19C).

Similar gene overdispersions in WT and *Rad21* KO macrophages

Inducible genes display lower burst frequencies in *Rad21* KO macrophages. In general, lower burst frequencies were characteristics of cohesin deficient systems, there would be an increase in the variability on the number of mRNA per cell (Figure 3.20A). Dr Nils Eling assessed scRNAseq variability in our data. Most variability measures are mean-dependent. His analysis corrected this dependence (Eling et al. 2018), and the resulting variability measure, residual dispersion, was validated by our smFISH experiments (Figure 3.21). The correlation between

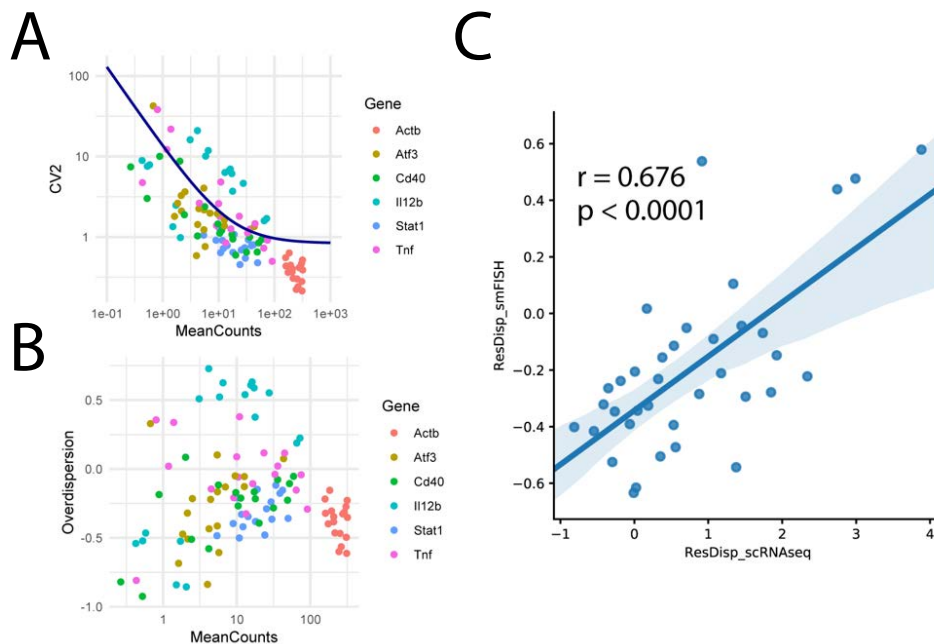


Figure 3.21: Validation of scRNAseq residual dispersion by smFISH. A) Line fit of a gamma family generalised linear model to the CV2 and the mean counts by smFISH (Residual dispersion is the deviation from this line). Each dot represents a replicate. B) smFISH overdispersions over mean counts. Each dot represents a replicate. C) Comparison between Residual dispersion calculated by scRNAseq and smFISH. To be able to do a one to one comparison, the residual dispersion among smFISH replicates was averaged. For more details, check the methods. r , Pearson correlation; p , p-value.

residual dispersion measured by smFISH and scRNAseq was 0.676 with a p-value > 0.0001 .

We focused on genes without differences in mean expression (Figure 3.20B), and we found We found no real differences in their residual dispersion between WT and *Rad21* KO macrophages (Figure 3.20C). Furthermore, the number of genes differentially more variable in WT is similar to the number of genes differentially more variable in *Rad21* KO (Figure 3.20D). The only remarkable finding was that the transcriptome of *Rad21* KO macrophages is less diverse than WT macrophages (Figure 3.20E).

Discussion

This chapter focussed on how cohesin depletion alters different aspects of RNA production in primary macrophages. We linked expression changes to alterations in transcriptional processes (Figure 3.1A). Burst sizes were slightly bigger in *Rad21* KO macrophages, but expression changes between WT and *Rad21* KO macrophages correlated with changes in burst frequencies (Figure 3.16). Cohesin-deficiency in macrophages led to an increase in RNA content per cell (Figure 3.1B) and an overall reduction in RNA pol II pause ratios (Figure 3.1C) which has also been reported in other systems (Schaaf et al. 2013). Reduction in RNA pol II pausing is often accompanied by increased expression levels. Upon LPS-activation, LPS-inducible genes reduced RNA pol II pausing. This reduction was more limited in *Rad21* KO than in WT macrophages (Figure 3.1B). Short-term effects of cohesin depletion associated with TAD boundaries, whereas long-term effects associate to genes far from TAD boundaries (Figure 3.17). Our data suggest that LPS-inducible expression is modulated through burst frequency for at least three genes, *Il12b*, *Peli1* and *Sertad2* (Figures 3.7 and 3.11). The LPS-inducible expression is reduced in every macrophage, not just in a fraction of them (Figure 3.19). Similarly, LPS-repression also acted through the reduction in burst frequencies, which was insufficient in *Rad21* KO macrophages (Figure 3.8). An interesting observation was that whereas LPS and IFN- γ priming either decreased or did not affect WT burst sizes of the LPS-repressed genes *Hprt* and *Fh1*, they increase burst sizes of *Hprt* and *Fh1* in cohesin-deficient macrophages (Figures 3.8F and K). Assuming that these genes are representative cases, an increase in burst sizes can compensate for the LPS-repression of burst frequencies in *Rad21* KO macrophages.

Our analysis of DE at TAD boundaries (Figure 3.17) suggests that the first genes to be DE after

cohesin depletion have their TSS in the vicinity of a TAD boundary, which is in accordance with Luppino et al. 2020, however long-term effects of cohesin malfunction preferentially affect genes at greater distances from TAD boundaries. Boundary regions are enriched with housekeeping genes and associate to high levels of transcription (Dixon et al. 2012). It is very plausible that a safety mechanism exists to keep the expression control of boundary regions in the event of a collapse of the genome architecture in a similar way a pTEFb feedback becomes activated in the event of Mediator depletion (Jaeger et al. 2020). Mediator depletion shared two similar features with cohesin depletion. First, it affected cell-type-specific expression rather than constitutive expression (Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Seitan et al. 2013). Second, Mediator depletion led to a decrease in RNA pol II pause ratios similar to what we found in *Rad21* KO macrophages (Figure 3.1). The bridge that connects the enhancer-recruited Mediator with its target promoter is cohesin. As cohesin bridges Mediator to its target genes, lack of cohesin might mimic the lack of Mediator. Therefore, it is possible that the same compensatory mechanism is active in cohesin-deficient macrophages, being potentially responsible for the increase in RNA content.

We hypothesised that cohesin depletion could lead to an increase in cell-to-cell variability for two reasons. First, cohesin is involved in burst frequency control by LPS, and an overall reduction in burst frequencies could lead to an increase in cell-to-cell variability. Second, burst sizes seem to be slightly bigger in *Rad21* KO macrophages (Figure 3.16). Third, cohesin depletion leads to an equalisation of the gene expression levels in bulk RNA-seq of thymocytes (Seitan et al. 2013), which is also seen in macrophages (Supplementary Figure 7.3). In this equalisation, lowly expressed genes tend to be up-regulated, whereas highly expressed genes tend to be down-regulated. At the single-cell level, this had two explanations; either equalisation occurs in every cell or a result of averaging a randomised expression. We found no evidence of an increase in cell-to-cell variability in our scRNAseq. *Rad21* KO cells presented a less diverse transcriptome than WT cells, and there was no difference in gene overdispersions (Figure 3.20). Therefore, cohesin alteration of burst frequencies did not increase gene dispersions, and there is no indication that cohesin depletion leads to a randomised expression. Visual inspection of our reduced scRNAseq data suggests the presence of subpopulations upon LPS-activation. However, a subpopulation analysis was not ready for the thesis submission deadline, and further analysis will be carried out to identify subsets of genes with increased variability (Figure 3.19A). We are currently carrying out a preliminary analysis, but we have not yet found a clear significant

pattern.

smFISH is the gold-standard for measuring gene expression in individual cells (Battich et al. 2013). Here, we combine intron and exon probes to distinguish nascent from mature expression. Our transcriptional burst identification and quantification have some limitations. First, splicing rate could affect our burst measures. Second, the difference between the identification of a single mRNA and a transcription burst relies on an intensity threshold set by a human, above which a dot is considered a burst. This threshold is easy to set for genes with fire-work shaped TSSs, like *Actb* or *Il12b*, consisting of a big blob surrounded by small dots. However, that was not always the case, particularly for small burst sizes, for example, for *Fh1*. This threshold affects burst fraction, but also burst quantification. Our burst quantification method relies on the distinction between bursts and individual molecules. Therefore, I want to emphasise that the number of mRNA per burst is an estimation normalised across different imaging sessions. Nevertheless, absolute burst sizes and frequencies should be considered in light of our limitations.

Il12b and *Peli1*, increase their expression in response to LPS. However, at the single-cell level, their induction is very different. *Il12b* expression occurs in big but infrequent bursts, whereas *Peli1* bursts are smaller and more frequent (Figure 3.7). As a consequence, *Il12b* increases in a fraction of cells whereas *Peli1* increases in practically every cell (Figure 3.5). Interestingly, *Il12b* burst frequency dropped from 90 to 120 min upon activation, suggesting the presence of a negative feedback inhibition of *Il12b* expression (Figure 3.7B and G, Table 3.4). On the one hand, *Il12b* is a subunit of Interleukin 12 (IL12) and 23 (IL23) (Oppmann et al. 2000), both proinflammatory cytokines. Engagement with its receptor activates INF- γ expression (Altare et al. 1998) and promotes Th1 growth and maturation (Hsieh et al. 1993). IL23 promotes the Th17 lymphocyte lineage (McGeachy et al. 2009). On the other hand, *Peli1* is an E3 ubiquitin ligase. It has a bivalent effect in immune responses that promotes inflammation through the ubiquitination of RIP1 (Chang et al. 2009). Thus, the product of *Il12b* is secreted, whereas *Peli1* functions inside the cell. Hence, whereas the cell population shares the consequences of *Il12b* expression through paracrine signalling, the consequences of *Peli1* expression affect only the cell that expresses *Peli1*. Therefore, it is interesting to hypothesise that components shared among cells can accept higher levels of variability than components that are not. More genes and a focused analysis on scRNAseq data could provide more convincing evidence for this hypothesis. Furthermore, our detailed timecourse data would be a valuable resource for

modelling approaches that can recover the time dimension from cross-sectional studies. We found contradictory results regarding *Peli1* burst sizes. In the three timecourse replicates where intronic and exonic *Peli1* probes were combined, *Peli1* burst sizes in the WT were larger than *Rad21* KO (Figure 3.7I). In the locus dataset when *Peli1* intron probe was combined with *Sertad2*, *Spred2*, and Enh probe (a total of nine replicates), *Peli1* burst sizes were bigger in the *Rad21* KO than in WT macrophages (Figure 3.11). However, the differences in burst frequencies in activated macrophages were consistent.

Cohesin intermingles contact domains (Luppino et al. 2020), shortening their distances. Accordingly, we found a general increase in burst-to-burst distances in *Rad21* KO macrophages, even though the distances between bursts of Enh and the inducible genes were not significantly different between WT and *Rad21* KO (Figure 3.14). Genes deregulated by cohesin depletion are enriched at enhancer vicinity (Cuartero et al. 2018; Ing-Simmons et al. 2015) and enhancer-promoter rewiring can explain their deregulation pattern (Thiecke et al. 2020). Enhancer Enh bursting is very similar between WT and *Rad21* KO cells. However, WT alleles are a third more likely to coexpress Enh and inducible genes than *Rad21* KO macrophages (Figure 3.15). In our experiments, the distance between TSSs can only be assessed when genes are expressed. Enhancers control bursts frequencies (Larsson et al. 2019) and, most studies suggest that only when a gene and an activated enhancer are at a certain distance, the gene bursts (Bartman et al. 2019, 2016; Chen et al. 2018; Fukaya et al. 2016). This probability might be reduced in *Rad21* KO because extrusion is not bringing promoters and enhancers close efficiently. Therefore, even though enhancer Enh activation occurs in *Rad21* KO macrophages, it is less likely to connect with *Peli1* and *Sertad2* promoters. Consequently, inducible genes burst less frequently, but enhancer-gene bursts appear at similar distances. Work in progress includes a DNA FISH experiment that will measure the loci distances independently on whether expression occurs or not.

Altogether, our results support a model where LPS alters macrophage expression by burst frequency modulation, and this modulation is weakened in *Rad21* KO macrophages. Cohesin mutations promote proliferation and alter differentiation (Fisher et al. 2017; Galeev et al. 2016; Mullenders et al. 2015; Viny et al. 2015). Inflammation also promotes differentiation (Pietras 2017). Thus, a weakened control of burst frequencies could lead to a resistance to differentiation, and that could be one of the contributions of cohesin mutations in hematopoietic cancers.

Chapter 4

Ifnb1 burst fraction directs collective behaviour in response to LPS

Collective behaviour of immune responses

Antimicrobial signals activate a new transcriptional landscape in macrophages (Bhatt et al. 2012; Kaikkonen et al. 2013; Ostuni et al. 2013). An immune reaction started by a fraction of cells can be amplified and coordinated at the population level through cell-to-cell communication. In this sense, the immune cell population responds collectively, where the first responders are the “leader cells” that enable the response of “follower cells”. One good example is *Ifnb1*, a key regulator of immune responses. During viral infections, *Ifnb1* is only expressed in a fraction of infected cells (Rand et al. 2012; Schmid et al. 2016). In this context, *Ifnb1* expression has consequences for the cell itself, by slowing down virus replication, but also for cells that detect the *Ifnb1* secreted by the *Ifnb1*-expressing cells. *Ifnb1* expression is stochastic, but response robustness is achieved by signal amplification through paracrine signalling that prevents viral spread (Rand et al. 2012). In our cohesin-deficient macrophages, the expression of *Ifnb1* is significantly reduced. In this chapter, I study how a reduced fraction of *Ifnb1*-producers affects the population as a whole.

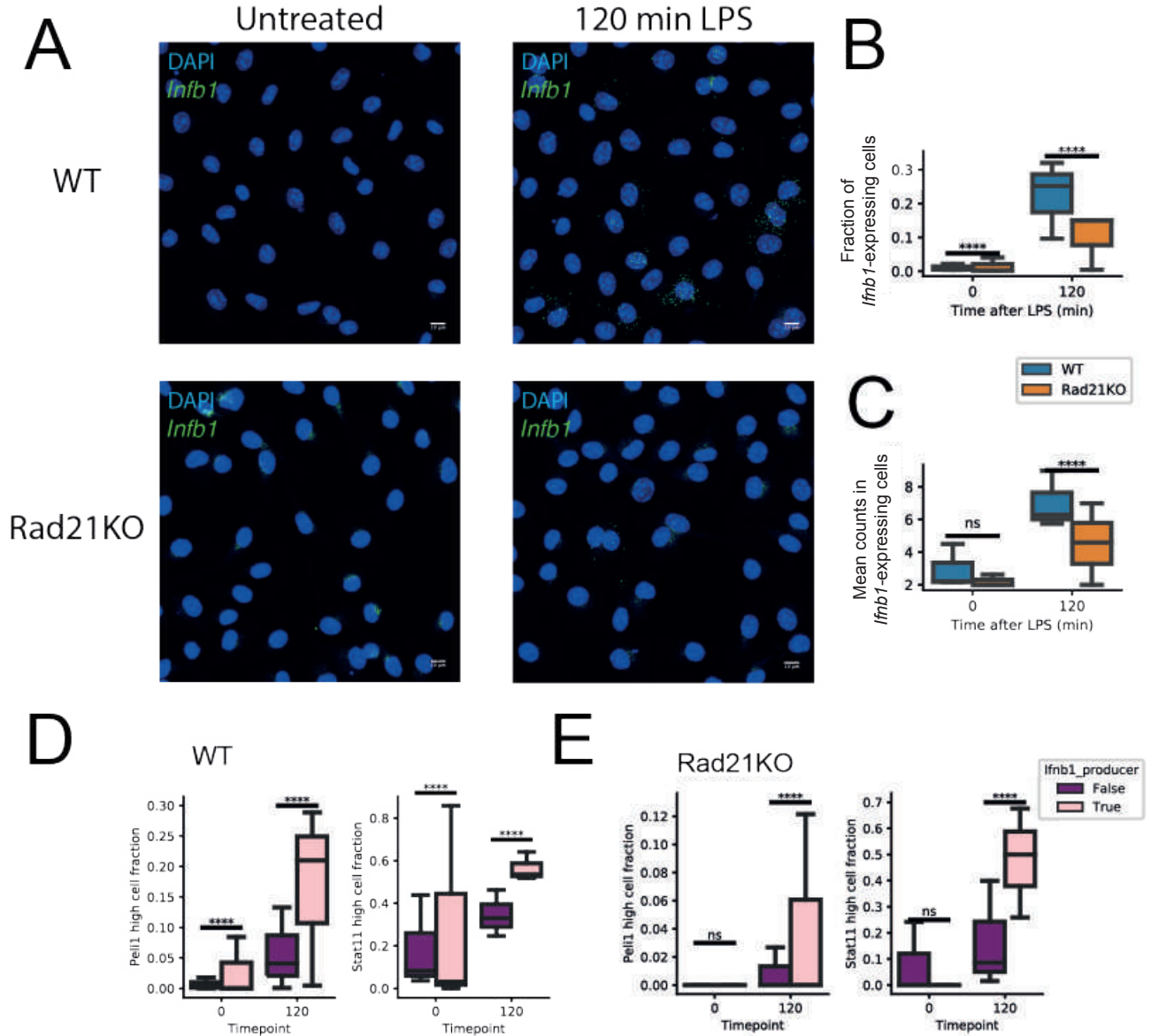


Figure 4.1: The fraction of *Ifnb1*-producers and their *Ifnb1* levels are reduced in *Rad21* KO macrophages. *Ifnb1*-producers express higher levels of *Peli1* and *Stat1*. A) Example smFISH images of *Ifnb1*-producers cells before and after stimulation. Scale bars represent 10 μ m. B) Fraction of cells with more than 1 *Ifnb1* count. Statistical test: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni multiple test correction. C) Mean *Ifnb1* expression in *Ifnb1*-producer cells in each replicate. pvalue: ANOVA (Counts in cells with at least 2 counts $\sim$ Sample + Replicate) with Tukey HSD for cross-comparisons and multiple testing correction. D and E) Fraction of cells with high levels of *Peli1* or *Stat1* for WT (D) and *Rad21* KO macrophages (E). Due to replicate variability, different thresholds were set for each replicate and gene, but the same threshold was apply to all WT and *Rad21* KO samples within the same replicate. Statistical test: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni multiple test correction. A cell is considered to be a *Ifnb1* producer if it contains 2 of more *Ifnb1* counts 3 biological replicates, $\sim$ 1000 cells imaged per sample (max = 5498, min = 452).

***Ifnb1* expression**

***Ifnb1* production occurs only in a fraction of cells**

In macrophages, *Ifnb1* expression peaks after 2h of LPS (Table 4.1). We used single-cell techniques to characterise the population of *Ifnb1* producers. In our scRNAseq data, no *Ifnb1* expressing cell was detected without or after 8h of LPS, but at 2h of LPS, 20.9% of WT macrophages presented *Ifnb1* reads. This fraction of *Ifnb1*-producer cells was confirmed by smFISH (Figure 4.1A), where *Ifnb1*-expressing cells went from $\sim 2\%$ in untreated macrophages to $\sim 20\%$ after 2h of LPS-activation.

In Shalek et al. 2014, they reported that *Ifnb1*-producers display high levels of an antiviral program. This program consisted of 98 genes that included *Peli1* and *Stat1*. This was corroborated by our smFISH experiments where *Ifnb1*-producers showed a higher expression of *Peli1* and *Stat1* compared to non-producers (Figure 4.1D).

***Ifnb1* producers are halved in *Rad21* KO macrophages**

Rad21 KO macrophages failed to induce *Ifnb1* as part of their response to LPS (Table 4.1). By scRNAseq, WT cells were 13 times more likely to have *Ifnb1* reads than *Rad21* KO macrophages. By smFISH, this difference is reduced, and WT macrophages were only twice more likely to express *Ifnb1* than *Rad21* KO (Figure 4.1B). However, *Rad21* KO *Ifnb1*-producers expressed less molecules of *Ifnb1* than WT *Ifnb1*-producers after 2h of LPS-activation (Figure 4.1C). *Rad21* KO *Ifnb1*-producers expressed higher levels of *Peli1* and *Stat1* compared to non-producers (Figure 4.1E), as WT macrophages (Figure 4.1D).

Comparison	log2FoldChange	padj
WT 2h LPS vs WT untreated	10.13,	3.51e-302
WT 8h LPS vs 2h LPS	-5.5	1.74e-113
<i>Rad21</i> KO 2h LPS vs WT 2h LPS	-4	1.95e-5

Table 4.1: **Differential expression analysis of *Ifnb1* by RNAseq.** DEseq2 comparisons of RNAseq for *Ifnb1*.

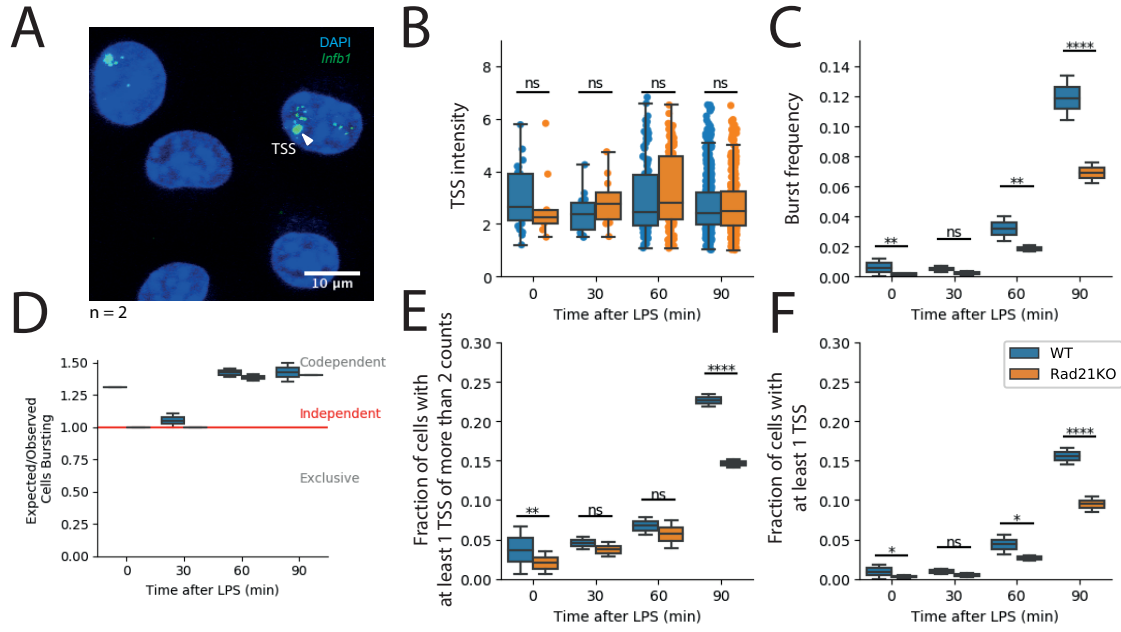


Figure 4.2: *Ifnb1* burst frequency is reduced in *Rad21* KO macrophages. A) Example *Ifnb1* TSSs. B) TSS intensities of all TSSs identified (radius 1). p-value: ANOVA (TSS intensity ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. C) Burst fraction in each sample. Statistical test: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni multiple test correction. D) Expected over observed cells with 2 TSSs. A value closer to 1 indicates that both alleles burst independently. Higher values indicate that their bursts are co-dependent, whereas lower values indicate that they are mutually exclusive. E) Fraction of cells expressing *Ifnb1*. Expressing cells have 1 TSS or more than 2 individual mRNAs detected. P-values calculate as C. F) Fraction of cells with at least 1 TSS. P-values calculate as C. 2 biological replicates, ~ 900 cells imaged per sample (max = 2184, min = 602).

Ifnb1 burst fraction is reduced in *Rad21* KO cells

Then, we examined *Ifnb1* transcriptional bursting patterns. To do so, we focus on half-hour steps after LPS activation to capture activation dynamics. *Ifnb1* lacks introns; thus, we were unable to use an intronic probe to identify nascent mRNA molecules and TSSs. However, TSSs can also be identified from mature signals because mRNAs cluster around their production site. Therefore, active TSSs signal is brighter and occupies more space than the signal coming from individual mRNAs (Figure 4.2A).

During LPS-activation, transcriptional burst size remains relatively unchanged (Figure 4.2B), but burst frequency is increased. First *Ifnb1* transcriptional bursts appear 1h after LPS in ~2% of cells. 30 min later, around 8% of WT macrophages presented at least 1 TSS (Figure 4.2F) and 12% of cells expressed *Ifnb1* to a certain extent (Figure 4.2E). Burst frequency is reduced

Genotype	Time after LPS (min)	Replicates	Cells	TSS intensity	Burst fraction	Cells expressing <i>Ifnb1</i>	Cells bursting <i>Ifnb1</i>
WT	0	2	1862	2.94	0.0059	0.037	0.009
<i>Rad21</i> KO	0	2	3115	3.24	0.0015	0.020	0.003
WT	30	2	1552	2.55	0.0050	0.045	0.009
<i>Rad21</i> KO	30	2	3021	2.81	0.0024	0.037	0.005
WT	60	2	1706	3.11	0.0316	0.067	0.044
<i>Rad21</i> KO	60	2	2998	3.13	0.0187	0.057	0.027
WT	90	2	1975	2.64	0.1189	0.227	0.156
<i>Rad21</i> KO	90	2	2687	2.57	0.0690	0.147	0.095

Table 4.2: **Summary of transcriptional burst measures of *Ifnb1* timecourse in WT and *Rad21* KO macrophages** All measures are the mean across replicates. Expressing cells have 1 TSS or more than 2 individual mRNAs detected.

in *Rad21* KO macrophages in untreated and 60-90 minutes post activation (Figure 4.2C), but no TSS intensity differences were detected (Figure 4.2B). Cells expressing *Ifnb1* from one allele were around $\sim 40\%$ more likely to express *Ifnb1* from the other allele (Figure 4.2D).

***Ifnb1* is induced in a subset of cells, whereas some *Ifnb1*-dependent genes are broadly expressed**

An increase in mean expression levels upon stimuli defines inducible gene expression. Nevertheless, this increase does not occur at the same level on all cells. Cell-to-cell response differences can be intrinsic or extrinsic. Intrinsic differences are due to probabilistic phenomena, such as transcriptional bursting, associated with low molecular numbers. Extrinsic differences are related to variations in cellular components, the micro-environment or the presence of hidden cellular states. These differences lead to different response thresholds in the cell population. Roughly speaking we can talk two different types of inducible expression: 1) Genes that increase its expression in a fraction of cells like *Il12b* and *Ifnb1*; and 2) Genes that increase its expression in more or less all cells, for example, *Ifnb1*-dependent genes like *Cxcl10* or *Ifit1* (Thomas et al. 2006).

scRNAseq or by smFISH techniques can study single-cell expression distributions. In Figure 4.3A, we have a snapshot of scRNAseq data of WT cells before and after 2h of LPS-activation. In these data, only a fraction of cells express *Ifnb1* and *Il12b*. On the contrast *Cxcl10* and *Ifit1* are broadly expressed (Figure 4.3B). If we look at smFISH data (Figure 4.3C), *Ifnb1* and *Il12b* counts do not group around the mean, whereas *Cxcl10* and *Ifit1* do (Figure 4.3D). Most LPS-

inducible genes look bimodal in most replicates due to transcriptional bursting and the fact that they have not reach steady-state expression when they were measured.

Identifying consequences of *Ifnb1* expression

WT and *Rad21* KO macrophages have different fraction of *Ifnb1*-producers, but *Ifnb1* targets are expressed in a higher number of cells. This suggests that *Ifnb1* expression has consequences not only for *Ifnb1*-expressing cells but also for the population as a whole.

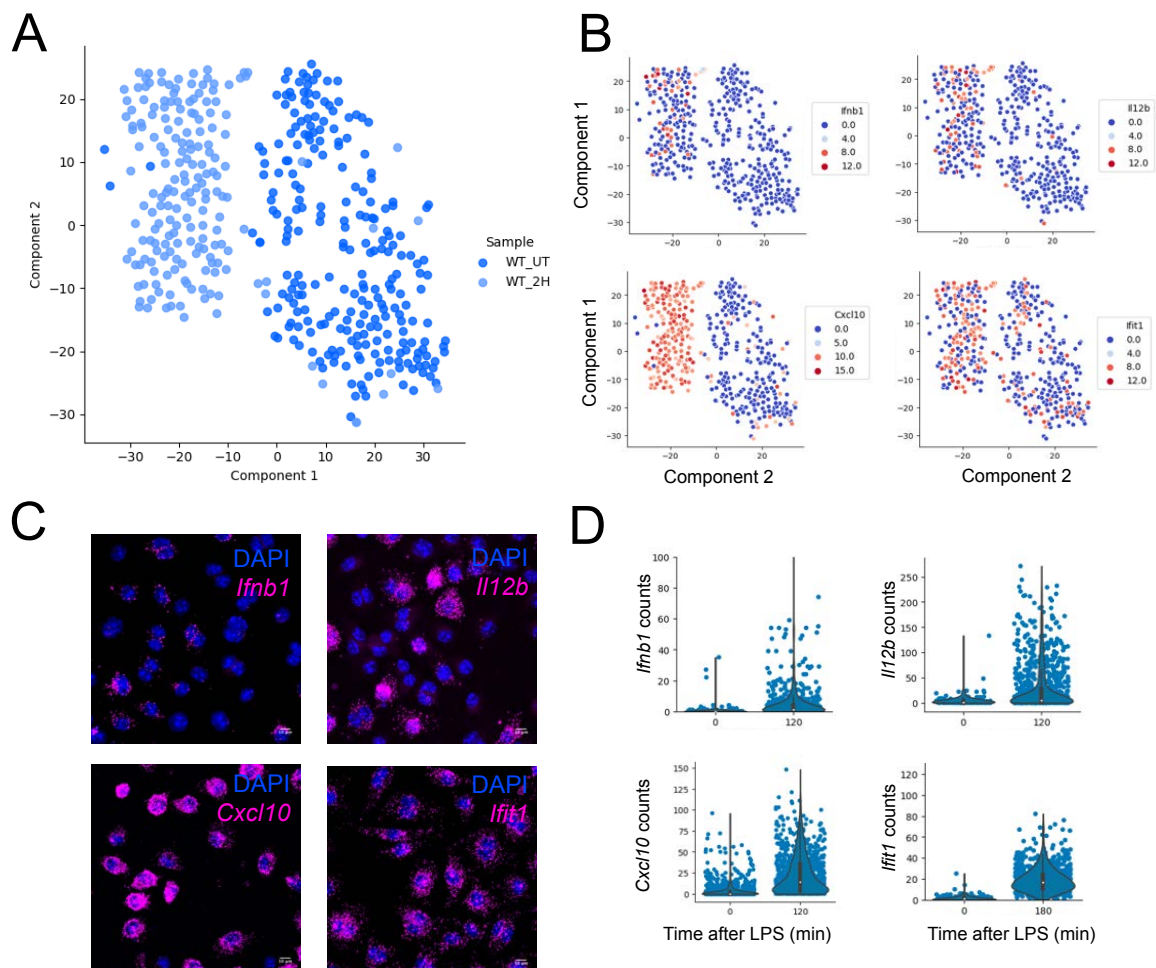


Figure 4.3: *Ifnb1* and *Il12b* are induced in a fraction of macrophages whereas *Ifnb1*-dependent genes are broadly induced upon LPS. A) tSNE with of scRNAseq WT macrophages data in untreated and 2h after LPS (on $\ln(\text{CPM} + 1)$, perplexity set as 20) B) $\ln(\text{CPM} + 1)$ expression of *Ifnb1* and *Il12b*, and the *Ifnb1*-dependent genes, *Cxcl10* and *Ifit1* C) smFISH example images of those genes in 2h LPS-activated WT macrophages. Scale bars represent $10\mu\text{m}$. D) smFISH example distributions of those genes.

At this point in the project, we wanted to measure *Ifnb1*-dependent collective behaviour. As a detector, we used macrophages that could respond to *Ifnb1* but cannot produce their own. In chamber slides, we mixed *Ifnb1* KO macrophages with either WT or *Rad21* KO macrophages. *Ifnb1* KO expression of *Ifnb1*-dependent genes measured *Ifnb1*-dependent collective behaviour. To identify the genotype of each cell in the mix population, we used two different CellTrackerTM (CT) dyes (Figure 4.5A). Image analysis of smFISH measured the expression of *Ifnb1*-dependent genes.

***Ifnb1*-dependent gene selection for smFISH**

We selected our *Ifnb1*-dependent genes based on available microarray data containing information on how *Ifnb1* KO macrophages respond to LPS. (Thomas et al. 2006). *Ifit1*, *Cxcl10*, and *Mx1* were selected as *Ifnb1*-dependent genes, based on the following criteria: 1) These genes were not expressed in untreated control macrophages; 2) Since in our data, *Ifnb1* production starts at 1h, their expression induction at 1h post-LPS is weak, but strong at 3h post-LPS in controls; 3) Failure to increase gene expression upon LPS activation in *Ifnb1* KO (Figure 4.4).

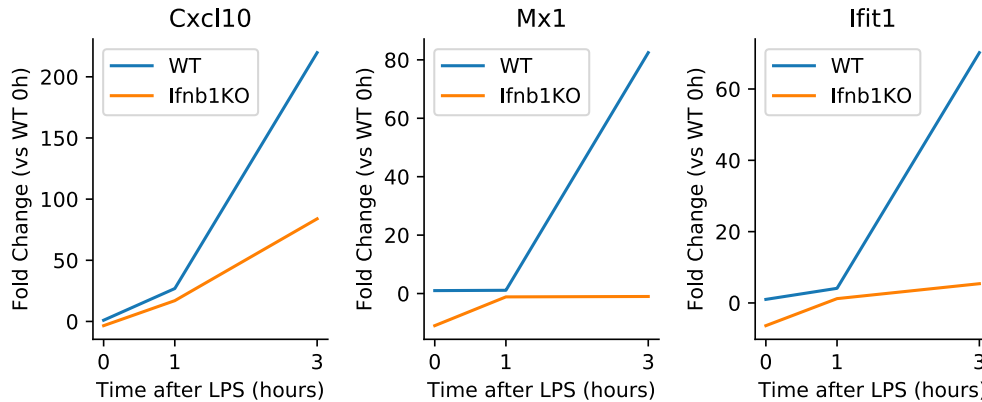


Figure 4.4: ***Ifnb1*-dependent genes for smFISH.** Fold Change compared to untreated WT cells vs time after LPS. Microarray data from Thomas et al. 2006.

***Ifnb1*-dependent expression in *Ifnb1*KO-mixed populations**

Follow-up macrophages that cannot produce their own *Ifnb1* were obtained from *Ifnb1* KO mice. *Ifnb1* KO frozen bone marrow (Deonarain et al. 2000) was differentiated in the same conditions as WT and *Rad21* KO. After 6 days of growth in conditional media, cells were stained with

green CT, washed, counted and mixed in a proportion 1:1 with red CT stained WT or *Rad21* KO macrophages (Figure 4.5B). LPS-activation and smFISH were performed on the following day.

In each image, red and green labels classified macrophages based on upper quartile intensities (Figure 4.5C). We tried other measures, such as maximum intensity, median intensity or mean intensity. However, maximum intensity is sensitive to pixel noise outliers, and, median and mean intensity were affected by intensity threshold used to identify cell outlines. The most robust method to identify green and red cells across replicates was the upper quartile green and red intensities.

Note that cells stained in green acquire red signal after co-culturing with red cells. To identify whether red dye transferal occurred due to cells genotype or due to co-culturing conditions and imaging, we took WT macrophages from the same mice (no genotype heterogeneity), separate them into two populations and stain them with red and green CT dyes. Then, we used a different technique (FACS) to quantify red and green intensity in each cell. Cells were stained 2 days before the 24h co-culture to exclude insufficient washing. As controls, we used single-colour labelled cells and a population mixed right before data collection. Red CT dye was able to transfer between populations after 24h co-culture, but not green CT, and not in the population that was mixed right before FACS analysis (Figure 4.5D). Therefore, red dye transferal occurred due to the nature of the red CT dye and the co-culture conditions, not due to the cells genotype or image acquisition. For smFISH, labelled cells were co-cultured for 12h before we started the experiments. We detected red dye transfer to the green labelled cells in images, but that did not impair the classification into green and red cells as the groups were still separated (Figure 4.5C).

***Cxcl10* induction in WT macrophages**

Cxcl10 encodes for a small secreted pro-inflammatory chemokine, and its regulatory regions have binding sites for NF κ , AP1, and an ISRE motif (Liu et al. 2011). It is a chemoattractant for cells with CXCR3 receptor, and it is involved in leukocyte trafficking and immune cell recruitment. . *Cxcl10* is barely expressed in untreated macrophages¹, but, after 2h of LPS,

¹Each WT untreated macrophage expressed an average of 1.19 *Cxcl10* mRNAs (Mean of means across replicates)

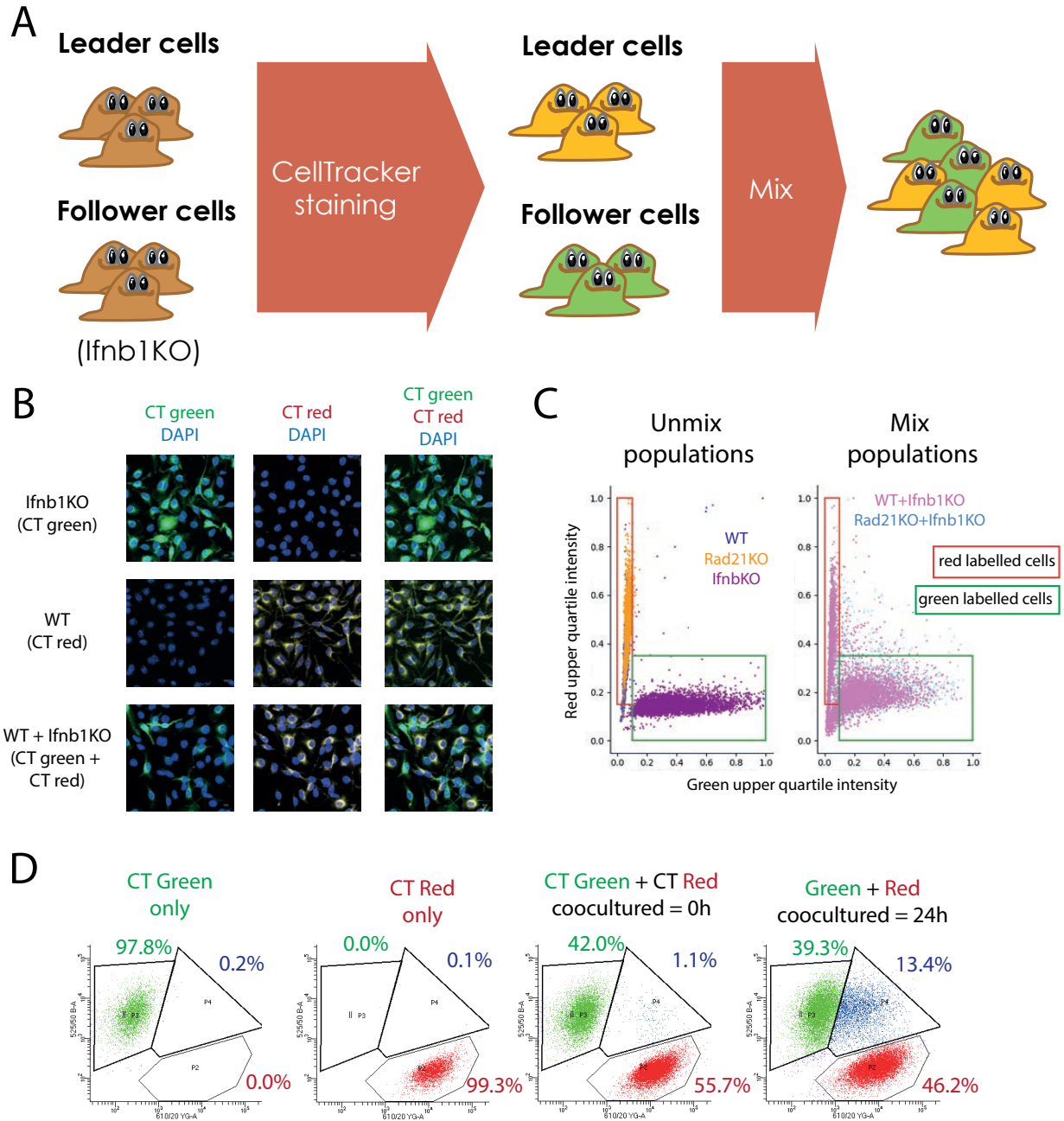


Figure 4.5: **Identifying follower and leader cells in a mix population.** A) Strategy. Leader cells can express *lfnb1* (either WT or *Rad21* KO), follower cells cannot. Different CT dyes stain the cytoplasm of the two populations so that genotype can be distinguished in a mix population. B) Example images of cells stained with green or red cell tracker in a mixed population. Scalebars represent 10 μ m. C) Classification into green and red cells by quantification of cell upper quartile intensity. CT staining occurs before mixing, and cells were co-cultured for 12h before analysis. WT and *Rad21* KO cells with red CT and *lfnb1* KO with green CT. Example images of CT stainings in mix and unmix populations. D) Red CT transfers among cells. For this experiment WT cells were stained with green or red CT at day 4 of differentiation and mixed one day before analysing the population by FACS; y axis, 525/50 B-A channel, x axis, 610/20 TG-A channel

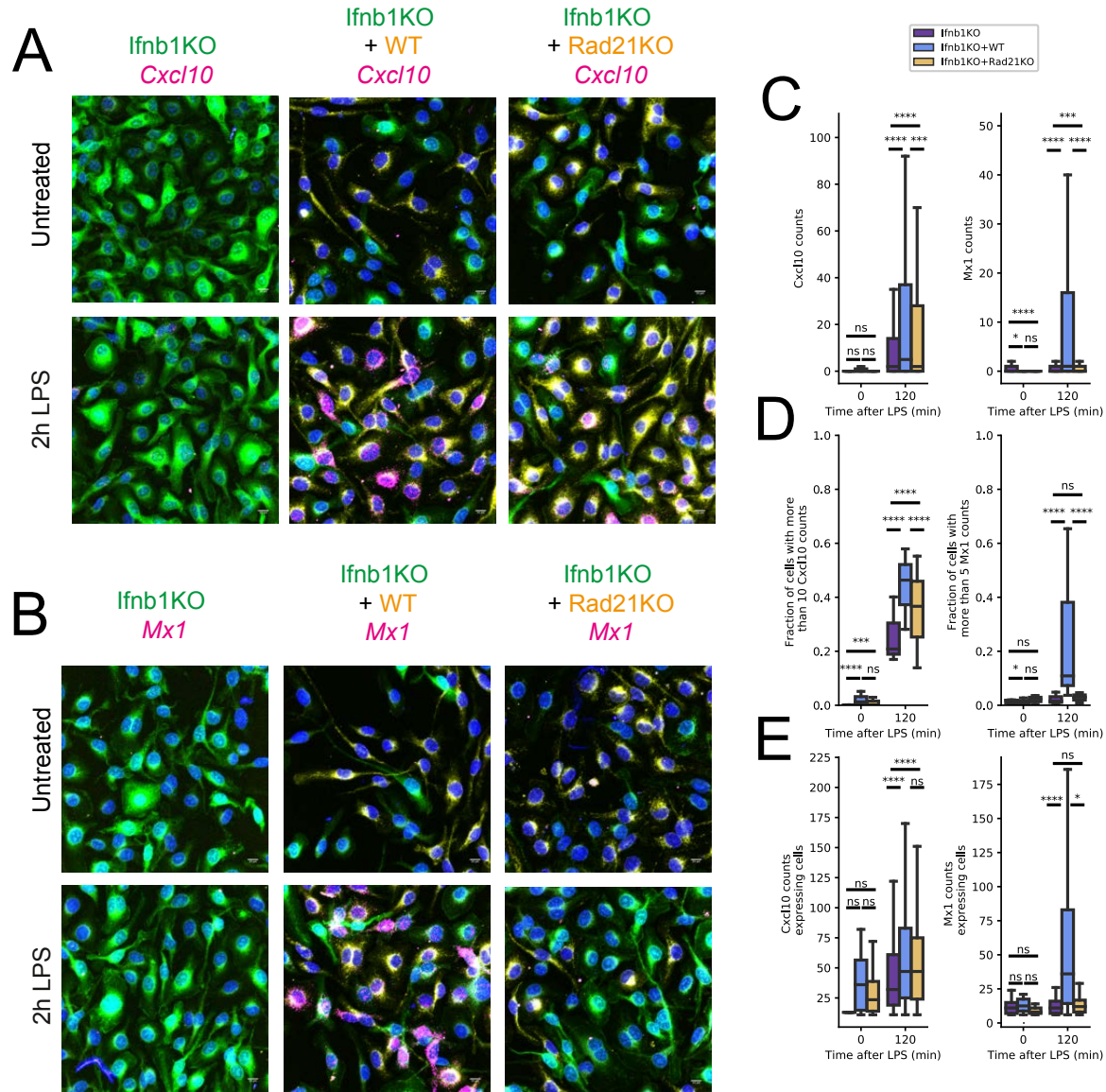


Figure 4.6: **Rescue of *Cxcl10* and *Mx1* expression in *Ifnb1* KO via co-culture with WT and *Rad21* KO cells.** A and B) Example images of *Ifnb1* KO cells (stained with CT green) mixed or not with WT or *Rad21* KO cells. Scalebars represent 10 μ m. WT and *Rad21* KO in yellow, *Ifnb1* KO in green *Cxcl10* smFISH signal in pink, DAPI in blue. C) Dot counts in mix and unmixed populations of *Ifnb1* KO macrophages. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. D) Fraction of expressing cells with more than 10 *Cxcl10* counts and fraction of cells with more than 5 *Mx1* counts. Cochran-Mantel-HaenszelTest (replicate as strata) with Bonferroni multiple test correction. E) Dot counts in mix and unmixed populations on cells expressing more than 10 *Cxcl10* or 5 *Mx1* mRNAs. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. Only cells classified as green considered (13698 in total for *Cxcl10*, 10656 for *Mx1*). 3 technical and biological replicates.

Cxcl10 expression increases 27.05 times² in WT macrophages (Figure 4.7A).

***Cxcl10* induction fails in *Rad21* KO and *Ifnb1* KO macrophages**

Neither *Rad21* KO or *Ifnb1* KO manage to increase *Cxcl10* expression to WT levels upon LPS. 54% of the WT cells expressed more than 10 *Cxcl10* counts by smFISH after LPS-activation (Figure 4.7A). In comparison only 31% of *Rad21* KO did it (Figure 4.7B) and only 26% of *Ifnb1* KO macrophages (Figure 4.6A). Considering *Cxcl10*-activated cells as cells with more than 10 *Cxcl10* counts, the mean expression of *Cxcl10*-activated macrophages dropped as well, to 63% WT levels in *Rad21* KO macrophages, and to 35% of WT levels in *Ifnb1* KO macrophages (Figure 4.6E).

Rescue of *Ifnb1*KO *Cxcl10* and *Mx1* induction via co-culture

Co-culture with *Ifnb1*-expressing cells partially restored *Cxcl10* induction in *Ifnb1* KO upon LPS-activation. This increase was higher when mixed with WT macrophages than with *Rad21* KO (Figure 4.6C). The fraction of expressing cells drove differences between WT and *Rad21* KO co-cultures. The number of *Ifnb1* KO cells expressing *Cxcl10* was higher in WT-*Ifnb1* KO co-cultures than in *Rad21* KO-*Ifnb1* KO co-cultures (Figure 4.6D). Nevertheless, *Ifnb1* KO cells that activated *Cxcl10* expression did not display any difference in mean expression depending on whether they were co-cultured with WT or *Rad21* KO macrophages (Figure 4.6E).

We also tested *Mx1* in *Ifnb1* KO unmixed and mixed populations. There, gene activation upon LPS induction was more variable across replicates (Figure 7.5), with one replicate where almost no rescue was detected. However, macrophages were activated because *Mx1* expression was measured in parallel to *Cxcl10*, and macrophages activated *Cxcl10*. We were able to detect some rescue level of *Mx1* LPD-induction in *Ifnb1* KO when they were cultured with WT macrophages but none with *Rad21* KO (Figure 4.6C, D and E). The rescue of *Mx1* in *Ifnb1* KO macrophages was statistically significant but less obvious than the *Cxcl10* rescue. We suspect that the apparent lack of *Mx1* induction is due to timing. In subsequent experiments, 3h of LPS instead of 2 lead to clear *Mx1* induction.

²After 2h of LPS, each WT macrophage expressed an average of 32.34 *Cxcl10* mRNAs (Mean of means across replicates)

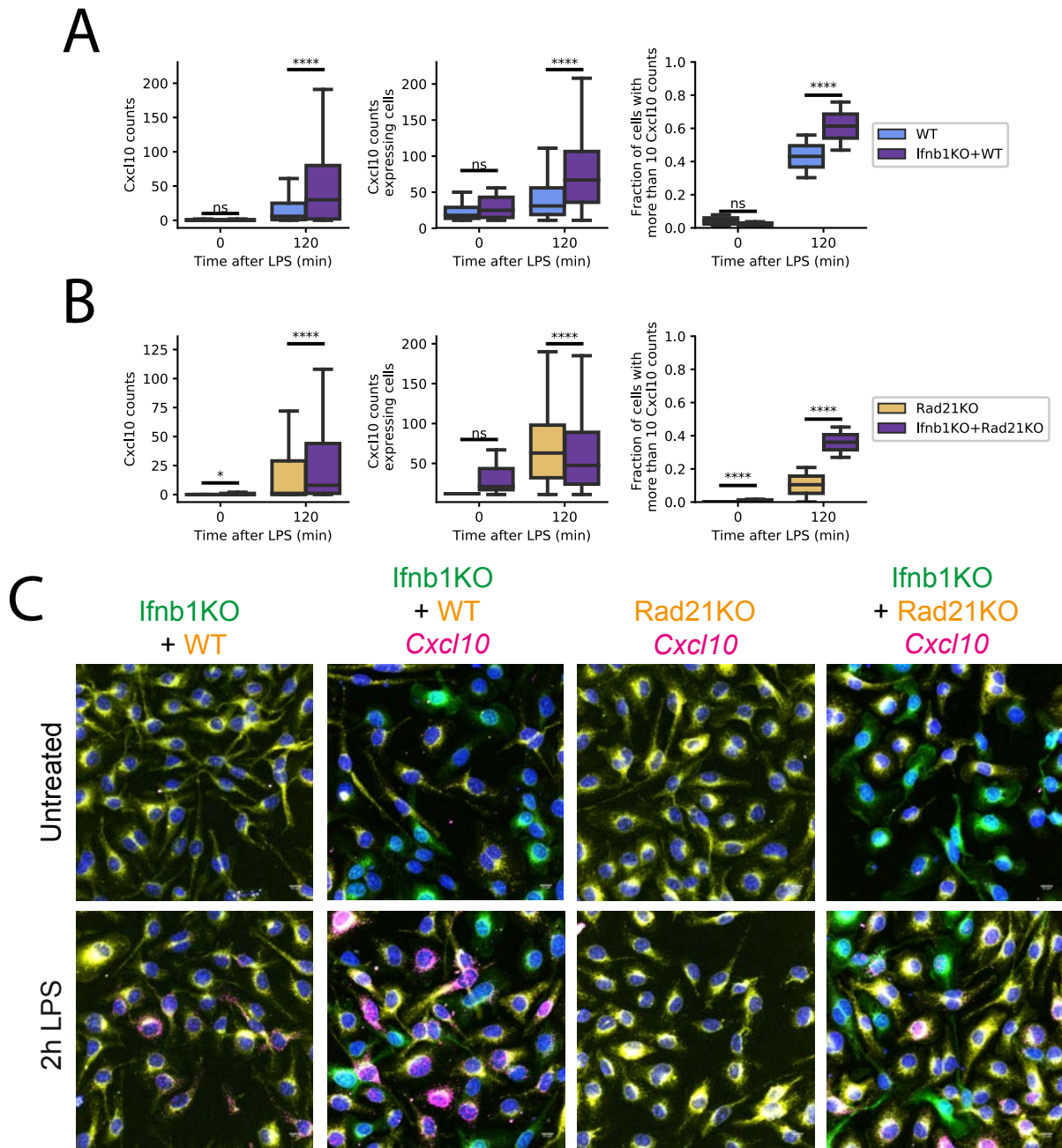


Figure 4.7: Co-culture with *Ifnb1* KO leads to an increase in *Cxcl10* expression upon LPS-activation. A) Expression changes in WT cells. Dot counts dot counts in cells with more than 10 *Cxcl10* counts and fraction of cells with more than 10 *Cxcl10* counts (5855 cells in total). B) Same as A but expression changes in *Rad21* KO cells (8031 cells in total). C) Example images of mixed populations. WT and *Rad21* KO in yellow, *Ifnb1* KO in green *Cxcl10* smFISH signal in pink, DAPI in blue. Scale bars represent 10 μ m. Only cells classified as red considered. 3 technical and biological replicates. Statistical test for dot counts comparisons: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. Statistical test for cell fraction comparison: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni multiple test correction.

Co-culture with *Ifnb1* KO cells strengthen *Cxcl10* induction

Perhaps, the most interesting observation in the co-culture experiments was the changes in *Cxcl10* expression in WT and *Rad21* KO macrophages. Co-culture with *Ifnb1*KO cells did not lead to significant differences in *Cxcl10* basal levels. However, in LPS-activated samples, WT and *Rad21* KO displayed a stronger *Cxcl10* when co-cultured with *Ifnb1* KO than in unmixed populations. WT *Cxcl10* displayed an increased numbers of cells activating *Cxcl10* at higher levels (Figure 4.7A). In the case of *Rad21* KO, the number of cells with more than 10 *Cxcl10* counts was almost three times higher, but the mean expression of the cells expressing *Cxcl10* dropped.

Inducible response in *Rad21* KO-WT co-cultures

In cancer, cells carrying carcinogenic mutations cohabit with somatic cells. This coexistence can alter somatic cell function and can promote cancer progression. Cohesin mutations are common in blood neoplasms and they favour proliferation in HPSCs. However, most studies focus on cell self-consequences of cohesin depletion, not whether cells with altered cohesin function could influence cells with WT cohesin. In this next section, we decided to study how the presence of *Rad21* KO macrophages affected the LPS response of WT macrophages, and vice-versa.

Ifit1 and *Mx1* induction after 3h of LPS

Ifit1 and *Mx1* are two *Ifnb1*-dependent genes which are induced upon LPS. In the previous experiments, we did not detect strong *Mx1* expression after 2h of LPS. *Ifnb1* expression starts at 1h in ~2-3% of cells (Figure 4.2F) , but at 2h around 20% of macrophages express *Ifnb1* (Figure 4.1B). Therefore, some *Ifnb1*-dependent events could take place after 2h of LPS, and we hypothesised that *Mx1* induction could be more robust at 3h, as in Thomas et al. 2006. Effectively, *Mx1* expression was strong and robust after 3h of LPS (Figure 4.8B). So strong, that dot signals merged and our analysis software could not count individual mRNAs in WT cells. Therefore, as an exception, in this set of experiments *Mx1* expression was quantified via cell integrated intensity. To normalise across replicates, *Mx1* intensity was relative to the mean

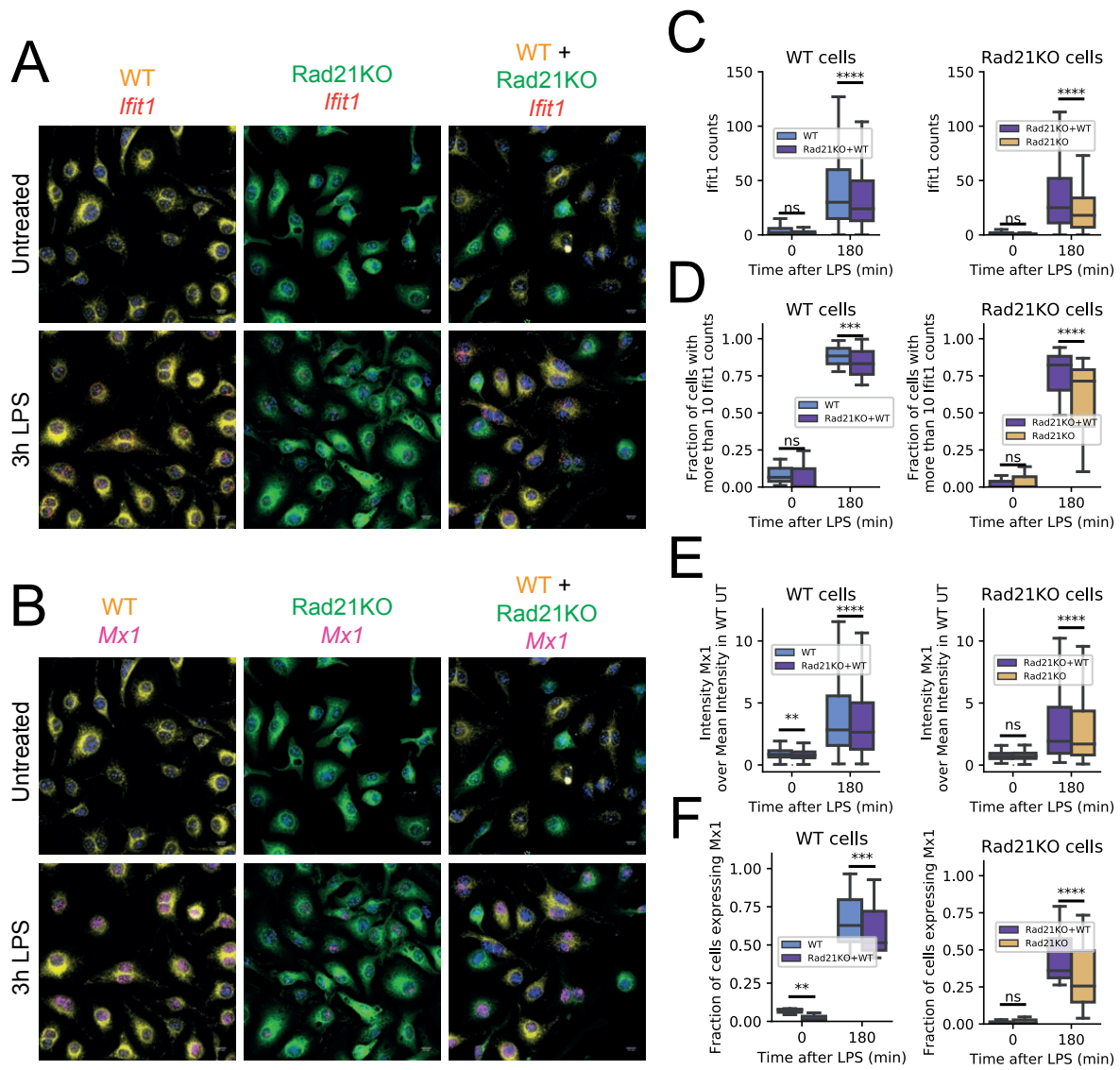


Figure 4.8: *Rad21* KO and WT co-cultures leads to a partial rescue of *Ifnb1*-dependent genes in *Rad21* KO. *Ifnb1*-dependent genes are reduced in WT macrophages via co-culturing with *Rad21* KO macrophages. A) Example images of *Ifit1* expression by smFISH in unmixed and mixed macrophage populations. WT and *Rad21* KO in yellow, *Rad21* KO in green *Ifit1* smFISH signal in pink, DAPI in blue. Only cells classified as red considered for WT and as green for *Rad21* KO. Scale bars represent 10 μ m. B) Example images of *Mx1* expression. Same colour code and scalebars as A. C) *Ifit1* expression in WT and *Rad21* KO cells in mixed and unmixed samples. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. D) Fraction of cells with more than 10 counts (right panel). Statistical-test: Cochran-Mantel-HaenszelTest (replicate as strata) with Bonferroni multiple test correction. E) *Mx1* expression in WT and *Rad21* KO cells in mixed and unmixed samples. Expression calculated over the mean integrated intensity of WT untreated cells in each replicate, as expression as for high expressing cells dot counting was not possible. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. F) Fraction of cells with more than 10 counts (right panel). Statistical-test: Cochran-Mantel-HaenszelTest (replicate as strata) with Bonferroni multiple test correction. Only cells classified as red considered for WT and as green for *Rad21* KO. 3 technical and biological replicates, 9687 cells analysed in total

integrated intensity in the WT untreated sample of the replicate. *Ifit1* expression was never as high and therefore, *Ifit1* mRNAs could be counted individually.

In WT cells, *Ifit1* mean of mean baseline expression was 3.45 mRNAs, whereas after 3h of LPS was 46.03, 13.35 times higher. In the case of *Mx1*, the expression intensity was 4.3 times higher than in untreated macrophages. For both genes, WT induction levels were twice the *Rad21* KO levels.

Differences between WT and *Rad21* KO *Ifnb1*-dependent genes induction are diluted in mixed populations

Co-culture with WT cells rescued *Ifit1* induction in *Rad21* KO cells (Figure 4.8B and C) and partially, the fraction of cells expressing *Mx1* upon 3h of LPS (Figure 4.8F). In the case of WT macrophages, co-culture with *Rad21* KO reduced the fraction of cells with high *Ifit1* and *Mx1* (Figure 4.8C and E). However, WT mean expression in the high expressing fraction is independent from whether WT cells were mix or not with *Rad21* KO macrophages (Figure 4.8D and F)

WT-*Rad21* KO co-culture drops *Ifnb1* and *Il12b* induction in WT cells

There was two possibilities for *Rad21* KO rescue in *Ifnb1*-dependent expression. Either co-culture with WT enabled them to increase their *Ifnb1* expression, or *Ifnb1* coming from WT cells enabled *Rad21* KO *Ifnb1*-dependent genes to be expressed in response to LPS. To test these hypothesis we measured *Ifnb1* in mix populations.

Co-culture with WT cells did not rescue *Rad21* KO *Ifnb1* induction. On the contrary, co-culture with *Rad21* KO macrophages dropped significantly WT *Ifnb1* induction levels and the fraction of cells expressing *Ifnb1* (Figure 4.9C and D). In mix populations, WT cells displayed increased levels of *Ifnb1* than *Rad21* KO cells. So the majority of the IFN present in the environment had a WT origin.

Il12b showed a similar picture to *Ifnb1*. Co-culture with WT cells did not modify neither *Il12b* induction levels (Figure 4.9E) or fraction of *Il12b*-expressing cells (Figure 4.9F). Nevertheless, co-culture with *Rad21* KO macrophages modified mean *Il12b* levels in WT macrophages (Figure

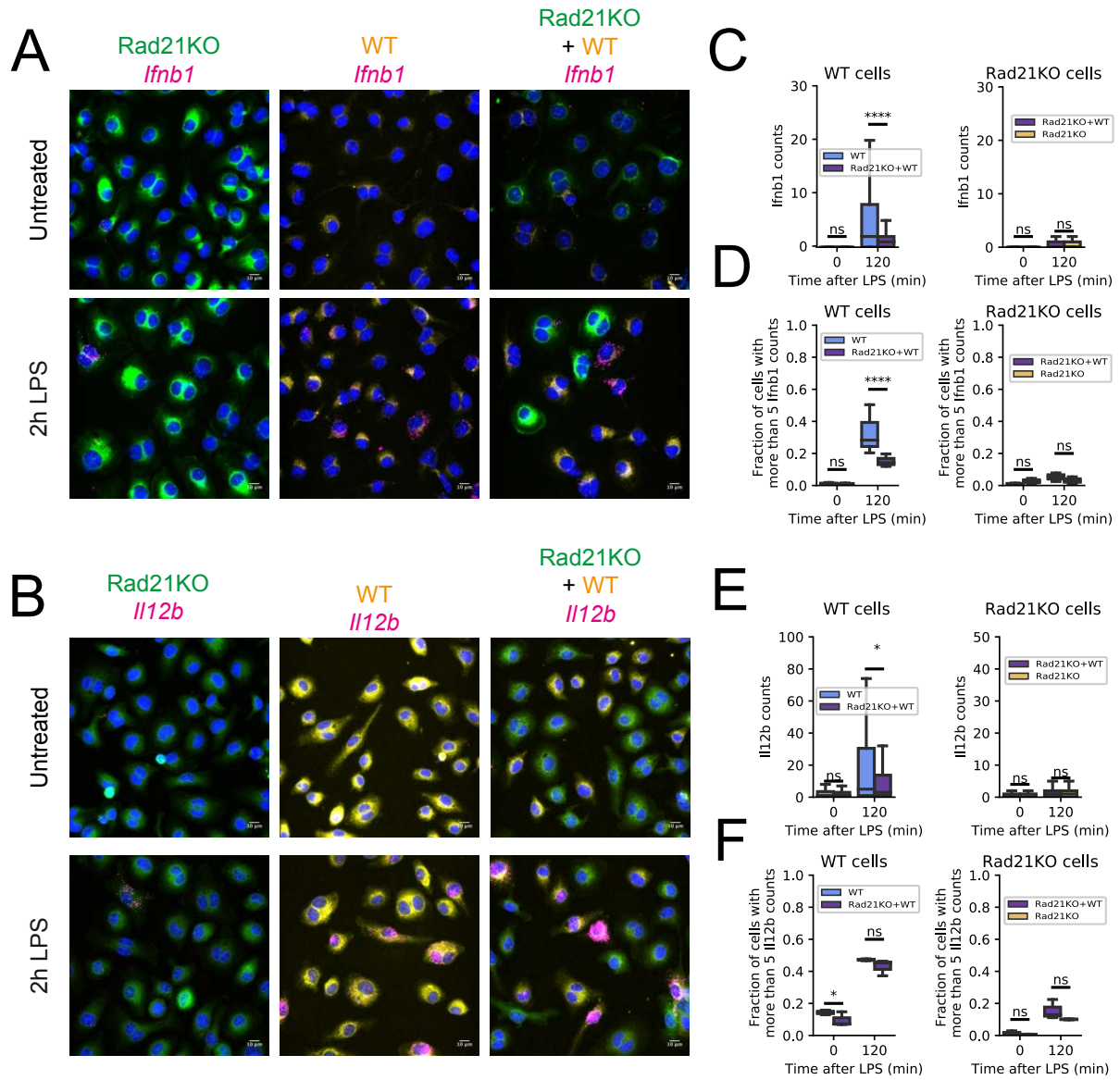


Figure 4.9: WT macrophages co-cultured with *Rad21* KO showed a reduction in the fraction of cells expressing *Ifnb1* and in the magnitude of *Il12b* expression upon LPS-activation. A) Example images of *Ifnb1* expression by smFISH in unmix and mix macrophage populations. WT and *Rad21* KO in yellow, *Rad21* KO in green *Cxcl10* smFISH signal in pink, DAPI in blue. Only cells classified as red considered for WT and as green for *Rad21* KO. Scale bars represent 10 μ m. B) Example images of *Il12b* expression. Same colour code and scale bars as A. C) *Ifnb1* expression in WT and *Rad21* KO cells in mix and unmix samples. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross-comparisons and multiple testing correction. D) Fraction of cells with more than 10 counts (right panel). Statistical-test: Cochran-Mantel-HaenszelTest (replicate as strata) with Bonferroni multiple test correction (5691 cells in total). E) *Il12b* expression in WT and *Rad21* KO cells in mix and unmix samples. Statistical test: ANOVA (Dot counts ~ Sample + Replicate) with Tukey HDS for cross comparisons and multiple testing correction. F) Fraction of cells with more than 10 counts (right panel). Statistical-test: Cochran-Mantel-Haenszel test (replicate as strata) with Bonferroni multiple test correction (5691 cells in total). Only cells classified as red considered for WT and as green for *Rad21* KO. For number of technical replicates refer to Table 7.2.

4.9E), but not the fraction of *Il12b*-expressing cells (Figure 4.9F). Again, in mixed populations, WT macrophage *Il12b* expression levels were higher than *Rad21* KO.

Discussion

Ifnb1 expression is activated in macrophages responding to LPS. At 2h LPS, there is a clear macrophage population that expresses *Ifnb1* and another that does not (Figure 4.1A). *Ifnb1* induction is blunted in *Rad21* KO macrophages due to changes in burst frequency. As a result, *Rad21* KO macrophages were half as likely to express *Ifnb1* after 2 hours of LPS-activation.

Ifnb1 induction heterogeneity can have two sources: intrinsic, due to transcriptional burst fluctuation; or extrinsic, due to hidden cellular states. We have evidence that intrinsic and extrinsic fluctuations play a role in *Ifnb1* expression. Hidden states are supported by the higher levels of inflammatory gene expression in *Ifnb1*-producer macrophages (Figure 4.1D), and also by the increased likelihood of the two alleles from a given cell to burst *Ifnb1* together. On the other hand, *Ifnb1* expression is bursty, and thus, intrinsic noise also shapes *Ifnb1* variability.

In our mixing experiments, we demonstrated that two populations of cells could modify the behaviour of each other. *Ifnb1* functions through paracrine signalling. Therefore, *Ifnb1* production does not have only consequences for the cell itself but also the surrounding cells. Co-culture with IFN-producing cells activated *Ifnb1*-dependent genes in *Ifnb1*KO. This increase was higher when mixed with WT macrophages, which had a larger proportion of *Ifnb1*-producers.

Unexpectedly, co-culture with *Ifnb1* KO macrophages elevated the induction levels of *Cxcl10* in WT and *Rad21* KO macrophages, but not at baseline. In mixed populations, WT expression was always higher than *Ifnb1* KO. One hypothesis is that this is due to broken negative feedback. *Ifnb1* KO cells cannot make *Ifnb1*, but that does not mean that feedback mechanisms that promote *Ifnb1* expression are not in place. These feedback mechanisms could be hyper-activated, as *Ifnb1* KO cannot express *Ifnb1*. If these feedback mechanisms involve molecules secreted to the environment or cell-to-cell contacts, they might promote a super-inflammatory state in neighbouring cells. These cell-to-cell interactions hyperactivate WT cells and assist *Rad21* KO macrophages in their response, as both *Ifnb1* KO and *Rad21* KO macrophages performed better together than separated. This is interesting as it suggests that cells with

different mutations can compensate each other through cell communication.

The fact that mutations can have effects in surrounding cells raised an obvious question: How can cohesin mutations in only a fraction of cells alter normal cell behaviour? We have studied 4 genes in unmixed and mixed *Rad21* KO-WT macrophage co-cultures: two LPS-inducible genes at 2 hours, and two IFN-dependent genes at 3 hours post LPS. This had consequences for WT and for *Rad21* KO cells. We initially expected that co-culture expression changes in *Rad21* KO would be similar to those caused by 24h IFN- γ pre-treatment. That was true for *Il12b*, where mixing did not modify *Rad21* KO *Il12b* induction (Figure 4.9C and D) or IFN- γ pre-treatment. However, results from IFN- γ pre-treatment and mixing did not agree for *Rad21* KO *Ifnb1* induction. 24h IFN- γ pre-treatment did partially rescue *Ifnb1*-induction in *Rad21* KO macrophages, whereas co-culture with WT cells did not (Figure 4.9C and D). Furthermore, in mixed populations, WT *Ifnb1* secretion partially rescued IFN-dependent expression in *Rad21* KO macrophages. IFN- γ pre-treatment was only able to modify the induction of *Mx1* in *Rad21* KO macrophages, but not *Ifit1* (Supplementary Table 7.3).

In summary, induction levels of all four genes were reduced in WT macrophages, especially in the number of high-expressing cells. Co-culture with WT cells did not rescue induction of neither *Il12b* nor *Ifnb1* in *Rad21* KO macrophages but did partially rescue IFN-dependent genes. These results indicate that *Rad21* KO can modify LPS responsiveness of WT cells. There are several open questions in this regard: First, whether co-culture conditions affect *Rad21* KO-WT interaction in terms of timing and mixing proportions. Second, the nature of *Rad21* KO-WT crosstalks. Candidates include transmembrane proteins or secreted cytokines and metabolites. Ongoing work in the lab will involve RNAseq of mixes and unmixed populations to identify possible cell crosstalk mechanisms.

Chapter 5

Summary and final discussion

The control of gene expression by the cohesin complex is the focus of this thesis. In *Chapter 2*, I analysed which type of genes become up and down-regulated and their enriched motifs when cohesin is depleted from macrophages. We established that cohesin mutation in AMLs presented a characteristic expression pattern that resembles an inversion of the expression changes that occur during the macrophage response to LPS. Up and down-regulated hallmarks in human AML were shared with cohesin-deficient LSK and our working primary cell model, mouse BMD macrophages. These macrophages showed a decreased response to antimicrobial signals, reduced IFN-related expression and an increase in mitochondrial content.

In *Chapter 3*, I examined transcriptional events altered by cohesin deficiency. Transcription played a major role in gene expression changes between WT and *Rad21* KO macrophages. Globally, *Rad21* KO macrophages had more RNA, increased burst sizes and less paused RNA pol II than WT macrophages. Expression changes between WT and *Rad21* KO macrophages were explained by changes in burst frequencies. BMD macrophages orchestrate the LPS-response through burst frequency modulation which was affected in *Rad21* KO macrophages. WT macrophages did not alter burst sizes in response to LPS, whereas activated *Rad21* KO macrophages increased burst sizes of LPS-repressed genes, alleviating the expression change caused by a reduction in their burst frequency. IFN γ priming in *Rad21* KO macrophages partially restored LPS-repression by burst frequency reduction. Furthermore, I studied bursting patterns of a 2Mb region with an inducible SE (Enh) and three genes DE genes in the absence of cohesin. Enh activation was very similar between WT and *Rad21* KO macrophages, but the

fraction of times that Enh burst with the inducible genes was significantly reduced.

In *Chapter 4*, I studied the consequences of an altered transcriptional bursting pattern for the collective behaviour and interactions between WT and *Rad21* KO macrophages in a mixed population. *Ifnb1* induction was lower in *Rad21* KO macrophages due to differences in burst frequencies. As a result, the fraction of *Ifnb1*-expressing cells was reduced in *Rad21* KO cells. WT and *Rad21* KO macrophages were able to activate *Ifnb1*-dependent expression in *Ifnb1* KO macrophages. Interestingly, *Ifnb1* KO macrophages increased the induction of *Cxcl10* in WT and *Rad21* KO macrophages. Moreover, co-culture of WT and *Rad21* KO macrophages had consequences for both of them. On the one hand, *Rad21* KO macrophages partially restored *Ifnb1*-dependent expression but not *Ifnb1* expression itself. On the other hand, the LPS-induced genes studied displayed a slight reduction in their induction levels in WT macrophages when co-cultured with *Rad21* KO macrophages.

Cohesin, inflammation and cancer

Despite the role of cohesin in mitosis, cohesin mutations are common in cancer (Cancer Genome Atlas Research Network 2013; Corces-Zimmerman et al. 2014; Kon et al. 2013), partly because the cell cycle can progress with reduced cohesin levels in yeast (Heidinger-Pauli et al. 2010). As cohesin mutations do not associate with abnormal karyotypes, they contribute to cancer by altering the regulation of gene expression (Corces-Zimmerman et al. 2014; Cuartero et al. 2019; Kon et al. 2013).

In the bone marrow, HPSCs differentiate into the different hematopoietic lineages while renewing themselves to maintain their population. Therefore, normal hematopoiesis implies a delicate balance between differentiation and self-renewal. Cohesin depletion pushes that balance towards self-renewal (Fisher et al. 2017; Galeev et al. 2016; Mullenders et al. 2015; Viny et al. 2015), but the mechanism remains unknown. Furthermore, patient AML samples with cohesin mutations had similar gene expression alterations to patients with mutations in the *WT1* and *FLT3* genes (Figure 2.1), suggesting that these mutations share mechanisms to drive cancer.

In *Chapter 2* we established a link between cohesin and inflammation (Cuartero et al. 2018).

MSigDB hallmarks that are induced by microbial signals in mouse macrophages are down-regulated in three cohesin deficient systems, whereas hallmarks that are repressed are up-regulated (Figure 2.4). Inflammation alters the fate of HPSCs, reducing their self-renewal and promoting differentiation, in order to produce enough mature immune cells to fight an infection (Pietras 2017). Cohesin depletion in macrophages reduced their response to microbial signals (Cuartero et al. 2018), suggesting that cohesin deficiency provide resistance to stimuli that foment differentiation.

Cohesin, enhancers, and bursts

Cell-state transitions, such as signal response or differentiation, are governed by enhancers, *cis*-regulatory elements that regulate transcription of their target genes (Schoenfelder and Fraser 2019). Which enhancer regulates which gene is determined by their proximity in the 3D nuclear space. The ring-shaped cohesin complex organises the distribution of chromatin fibers through loop extrusion (Davidson et al. 2019; Fudenberg et al. 2016; Kim et al. 2019), enabling enhancer-recruited Mediator to reach its target promoters (Jaeger et al. 2020; Kagey et al. 2010). Accordingly, genomic interactions are modulated during differentiation (Beagan et al. 2017; Bonev et al. 2017; Dixon et al. 2015; Hanssen et al. 2017). Furthermore, cohesin depletion is more likely to affect cell-type-specific genes than housekeeping (Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017; Seitan et al. 2013) and their deregulation associates with enhancer proximity (Cuartero et al. 2018; Ing-Simmons et al. 2015). Collectively this evidence let us conclude that cohesin collaborates with enhancers in the context-dependent control of gene expression, although its specific role remains to be determined.

Enhancers establish long-distance interactions with their targets, recruiting a set of co-activators, including Mediator. These long-distance interactions depend on spatial proximity, and cohesin, through loop extrusion, increases or decreases the probability of encounter. Most published work relate enhancer-driven control of gene expression to burst initiation, which depends on enhancer promoter-proximity (Bartman et al. 2016; Fukaya et al. 2016; Larsson et al. 2019). In the absence of cohesin, the position of TAD boundaries is randomised (Bintu et al. 2018), and short-range interactions are diminished (Luppino et al. 2020; Rao et al. 2017; Schwarzer et al. 2017).

Macrophage LPS-response is governed by enhancers (Kaikkonen et al. 2013; Ostuni et al. 2013), and the four LPS-inducible genes analysed (*Il12b*, *Ifnb1*, *Peli1*, *Sertad2*) increase their burst frequencies upon LPS exposure. All four genes presented lower burst frequencies upon induction in *Rad21* KO macrophages compared to WT. However, this does not seem to be related to a failure in enhancer activation. The inducible SE Enh between *Peli1* and *Sertad2*, is transcribed with similar transcriptional dynamics in WT and *Rad21* KO macrophages. However, the number of times that Enh bursts together with *Peli1* or *Sertad2* is reduced in *Rad21* KO macrophages compared to WT. Nevertheless, the distances between Enh bursts and inducible gene bursts are similar between WT and *Rad21* KO macrophages, suggesting that cohesin is involved in the enhancer-control of burst frequencies by increasing the probability of an enhancer-promoter encounter. In this regard, even in the absence of cohesin, when the distance requirement is met, a productive gene burst might occur. To address this question, we are planning to extend this analysis to other regions with inducible enhancers and genes coupled to DNA FISH experiments, that would inform where the genomic locations are at all times.

One possible caveat is that Enh bursts less frequently than *Peli1*. However, our probes detect transcripts from a certain region. There are several enhancers around *Peli1*, including within the gene body region (Figure 5.1, Ostuni et al. 2013), 4 of which belong to the SE Enh. These enhancers can have an additive effect in the modulation of *Peli1* burst frequency, but our experimental approach only detect activity from one.

The LPS response of macrophages is modulated preferentially through modulation of burst frequencies rather than sizes, because LPS-induction increased burst frequencies (*Il12b*, *Peli1*,

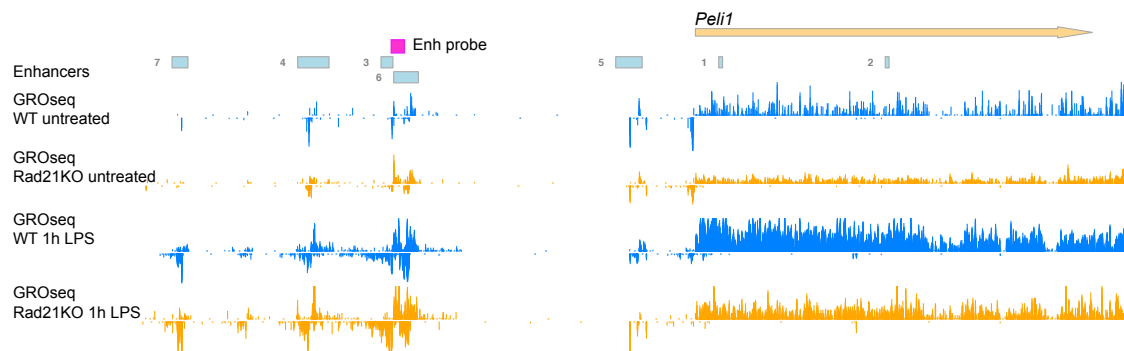


Figure 5.1: **Macrophage enhancers around *Peli1***. In pink, localisation of the Enh smFISH probe. Enhancers around *Peli1* where we detected transcription by GRO-seq (from Ostuni et al. 2013). GRO-seq reads from Cuartero et al. 2018.

Ifnb1, *Sertad2*) and LPS-repression decreased burst frequencies (*Hprt1*, *Fh1*). Stimulus-dependent modulation of burst frequencies enables fast and sensitive responses that can adapt to a range of stimulus intensities (Li et al. 2018), and the vast majority of published work that analyse bursting in response to a stimulus, reported changes in burst frequencies (Bartman et al. 2016; Chen et al. 2019b; Fritzsche et al. 2018; Fukaya et al. 2016; Larson et al. 2013; Molina et al. 2013; Senecal et al. 2014).

According to our data, cohesin reduced the magnitude of the burst frequency changes caused by LPS. In this regard, cohesin depletion decreased the response to LPS of *Rad21* KO macrophages. As inflammatory signalling cascades promote differentiation of hematopoietic cells, cohesin mutations could shift the balance between self-renewal and differentiation towards self-renewal by reducing responsiveness to environmental cues that induce cell-state transitions.

Burst frequency and paracrine signalling

Transcriptional induction of *Il12b* and *Ifnb1* generated characteristic population distributions with only a subset of macrophages with a high level of expression surrounded by non-expressing cells. However, the protein product of both genes interact with their respective receptor on the surface of cells. Changes in burst frequencies without changes in burst sizes associate with the modulation of the k_{on} parameter of the telegraph model (Peccoud and Ycart 1995; Raj et al. 2006b), which determines how many bursts occur per time unit. For genes involved in paracrine signalling with very low burst frequencies even in cells where they have been induced, burst frequency modulation determines how many cells secrete a factor that modifies the behaviour of cells independently of whether they secreted the factor, using cell-to-cell communication to amplify and coordinate the response of the population as a whole. As an example, the number of *Ifnb1* secreting macrophages was lower in *Rad21* KO than in WT macrophages. As a consequence, WT macrophages activate more *Ifnb1*-dependent expression in *Ifnb1* KO than *Rad21* KO.

Frequency modulation of a stochastic reaction can transform the single-cell digital output into an analogical output at the population level. For example, $\text{NF}\kappa\beta$ remains in the cell cytoplasm at baseline, but $\text{TNF}\alpha$ stimulation leads to its translocation to the nucleus which can be visualised by coupling $\text{NF}\kappa\beta$ to GFP. The translocation strength can be modulated,

but in general, cells can be classified by whether they have translocated $\text{NF}\kappa\beta$ or not. Cell heterogeneity in the population leads to different cells having different signal intensity thresholds to perform the translocation. As a consequence, different dosages of $\text{TNF}\alpha$ modulate the fraction of cells that translocate $\text{NF}\kappa\beta$ which is a TF that will induce expression changes in the cells where it has been translocated, including processes in cell-to-cell communication and allowing the cell population to adjust its response to a range of levels of signal input (Tay et al. 2010).

Mutations and cell-to-cell communication

Our expectations at the initiation of the mixed cultures experiments were to observe changes in *Ifnb1*-dependent expression in *Ifnb1* KO macrophages. Indeed, *Ifnb1* expression induction was higher in WT than in *Rad21* KO macrophages and, accordingly, co-culture with WT macrophages increased more *Ifnb1*-dependent expression in *Ifnb1* KO macrophages than co-culture with *Rad21* KO. Initially, the possibility that *Ifnb1* KO macrophages would modify the expression of WT and *Rad21* KO was not taken into account. However co-cultured with *Ifnb1* KO macrophages lead to an hyperactivation of *Cxcl10* in WT and *Rad21* KO macrophages compared to their respective unmixed populations. Intrigued by this observation, we co-cultured WT and *Rad21* KO macrophages in order to examine whether they could modify each other behaviour. Indeed, in LPS-activated samples, there was a partial rescue of *Ifnb1*-dependent expression in *Rad21* KO cells and a decrease on the four LPS-inducible genes tested in the WT macrophages in mixed populations.

Cohesin is not directly involved in cell-to-cell communication as it is a nuclear complex, nevertheless, our results suggests that regulatory changes caused by cohesin defects can affect the behaviour of neighbouring cells, at least in the context of their response to LPS. This opens the possibility that cohesin mutations in leukaemias could promote proliferation by altering the bone marrow niche, maybe by inducing resistance to differentiation in general, not constrained exclusively to cohesin mutated cells. Cancers cells often hijack somatic cells to promote tumour growth and enhance angiogenesis. For example, it has been reported that exosomes released from melanoma cells can induce a pre-metastatic niche in the bone marrow (Peinado et al. 2012).

Conclusion

DNA elements orchestrate cell-state transitions by differential recruitment of TFs and by interacting between them. Cohesin mediates spatial interaction between promoter and enhancers through loop extrusion. Inflammation promotes the differentiation of HPSCs, whereas cohesin deficiency promotes the self-renewal of HPSCs, where it can induce a pre-leukemic state. Loss of cohesin function associates with a transcriptional landscape with low inflammation and high OXPHOS and *Myc* target expression. Expression changes triggered by LPS are primarily modulated by changes in transcriptional burst frequencies. Burst frequency is under enhancer control and genes deregulated in the absence of cohesin associates with enhancer presence. In the absence of cohesin, macrophages can detect LPS, but they fail to establish a strong response because of a more constrained modulation of burst frequencies compared to WT. The alteration of burst frequencies has consequences for the cell itself, but also for the population as a whole, especially if the gene products are involved in paracrine signalling. WT and cohesin-deficient cells had different fractions of *Ifnb1* producers which can activate *Ifnb1*-dependent expression in *Ifnb1* KO macrophages. Co-culture with cohesin deficient macrophages can reduce LPS-inducible expression in WT macrophages. In short, our results suggest that cohesin deficiency leads to a defective modulation of burst frequencies at cell-state transitions which could contribute to malignant transformation by providing resistance to differentiation stimuli.

Chapter 6

Materials and Methods

Bone marrow derived macrophages

WT and *Rad21* KO macrophages

Bone marrow progenitors were extracted from mouse tibia and femur of *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{wt/wt}* mice and *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{lox/lox}* mice (Seibler et al. 2003; Seitan et al. 2011). Cells were then cultured for 7 days in DMEM (supplemented with 0.25mM sodium pyruvate, 0.05mM β -mercaptoethanol and 100U/ml Penicillin and 100 μ g/ml Streptomycin) + 20% L929 conditional medium at 37 °C. 96 hours after extraction medium was changed but this time with 0.2 μ M of 4-OHT to promote creERT translocation to the nucleus, where Rad21 exons 5 and 6 are excised in the case of the *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{lox/lox}* macrophages (*Rad21* KO macrophages) but not in *Rosa26^{ERT2Cre/ERT2Cre} Rad21^{wt/wt}* macrophages (WT macrophages).

Ifnb1 KO macrophages

Cryopreserved bone marrow from *Ifnb1^{-/-}* mice (Thomas et al. 2006) was obtained from David Brooks lab (The Princess Margaret Hospital, University of Toronto). Cells were thawed and then treated and differentiated as the section above.

Cell treatments

L929 conditional media production

L929 cells are thawed and plated in DMEM supplemented with 0.25mM sodium pyruvate, 0.05mM β -mercaptoethanol and antibiotics (100U/ml Penicillin and 100 μ g/ml Streptomycin). When they achieved 80 % confluence each 10 cm dish was split into 20 40ml flasks. 10 days later, L929 conditional media is collected from the flasks and filtered through a sterile 0.22 μ m filter. Then, L929 media was aliquoted and stored at -20 °C.

LPS-activation

For LPS-activation, cells were treated with 10ng/mL of LPS (Sigma-Aldrich L7895).

IFN- γ pre-treatment

Cells were treated with 10ng/ml Recombinant IFN- γ (Invitrogen, PMC4031) for 24h.

CellTracker staining

Cells were stained for 15 min at 37 °C with 5 μ M CellTracker Red (Invitrogen, C34552) or CellTracker Green (Invitrogen, C2925) in serum-free media. Then, cells were washed twice with phosphate-buffered saline (PBS).

Cell mixing

Cell populations were stained as described above. Then, cells were lifted using PBS 5 μ M for 10 min and centrifuged 1200 revolutions per minute (rpm) for 5 min. After counting, cell populations were mixed 1:1. Unless stated, macrophages were stained and mixed on day 6. Subsequent experiments were performed on day 7.

Genomic DNA extraction

First, cells were lifted with PBS + 5mM EDTA for 10 min and then centrifuged for 5 min at 3000 rpm. The supernatant was discarded and the cell pellet was stored at -80 °C until used.

Genomic DNA was extracted from cells using DNeasy Blood & tissue kit (QIAGEN, 60504) as indicated by the manufacturer's protocol. Briefly, the cell pellet was thawed and resuspended in 200 μ L of PBS. 200 μ L of Buffer AL + 20 μ L of proteinase K were added and pipetted up and down. The mixture was incubated for 10 min at 56 °C, and 200 μ L of ethanol (96-100%) were then added. Next, the mixture was pipetted into a DNeasy Mini spin column and centrifuged at 8000 rpm for 1 min. In the following step, 500 μ L of buffer AW1 was added, and the column was spun again at 8000 rpm for 1 min. Afterwards, 500 μ L of buffer AW2 was added, and the spin column was centrifuged for 3 min at 14000 rpm. Finally, DNA was eluted by adding 100 μ L of buffer AE and incubated for 1 min at room temperature (RT). Then, the column was spun at 8000 rpm for 1 min. Eluent contained genomic DNA, which was then measured by nanodrop.

RNA extraction

For lysis, adherent cells were washed twice with PBS. PBS was removed and 1 mL of RNAbee (Bio Connect, CS-501B) was added, and the lysate was pipetted several times up and down to ensure lysis, which could be kept at -20 °C.

For phase separation, 0.2 mL chloroform was added, and the mixture was shaken vigorously for 20 seconds. Afterwards, the mixture was kept on ice for 5 min and then, centrifuged at 12000g for 5-15 min at 4 °C. The RNA remained in the aqueous phase, whereas DNA and proteins were in the interphase and organic phase.

For RNA precipitation, the aqueous phase was transferred into a clean tube with 0.5 ml of isopropanol + 1 μ L of GlycoBlue (Invitrogen, AM9515). GlycoBlue co-precipitated with the RNA and stained the pellet in blue. Incubate the mixture for 5-10 min at RT and then centrifuge at 12000g for 5 min at 4 °C. Supernatant was removed.

To wash RNA, 1 mL 75% ethanol was added. Subsequently, the pellet from the side of the

tube was dislodged by vortexing. Then, the tube was centrifuged for 5 min at 7500g at 4 °C. The supernatant was removed, and RNA pellet was air-dried (5 - 10 min at RT).

Then, RNA was DNase treated and purified using the RNeasy MinElute Cleanup Kit (Qia-gen, 74204) (Following manufacturer's Appendix C of "DNase digestion of RNA before RNA cleanup"). Then RNA was stored at -80 °C.

cDNA synthesis

To generate complementary DNA (cDNA), we followed the protocol for SuperScript™ III Reverse Transcriptase (Thermo Fisher, 18080085). 200 ng of RNA was used as starting material, and the reaction volume per sample was 20 μ L. In short, RNA volume was diluted to 11 μ L in nuclease-free water. Then, we added 1 μ L of random primers (100 ng/ μ L) and 1 μ L of dNTPs (10 mM). The mixture was denatured for 5 min at 65 °C and then cooled at 4 °C. Next, we added 2 μ L of FS 5x, 2 μ L of DTT, 1 μ L of RNaseOut, and 1 μ L of Superscript III. Retrotranscription occurred in the thermocycler where samples were incubated first for 10 min at 25 °C, then, for 50 min 50 °C; and finally the reaction was inactivated at 70 °C for 5 min.

Real time quantitative PCR

RT-PCR was performed using IQ SYBR Green Supermix (Bio-Rad, 1708880) and a CFX Real-time PCR system (Bio-Rad). Each reaction contained: 6 μ L of SYBR Green, 2 μ L of cDNA, 0.4 μ L primers (10mM), and 3.4 μ L of nuclease-free water. Primers used on gDNA are listed in Table 6.1. Primers for gene expression analysis are listed in Table 6.2. Values were normalised through the $\Delta\Delta$ Ct method.

Name	Sequence	Aim
mtND1 Fw	CCTATCACCCCTTGCCATCAT	Mitochondrial DNA
mtND1 Rv	GAGGCTGTTGCTTGTGTGAC	Mitochondrial DNA
trnah chr2 F	CCCGGGCTAATTCTCCTATGTC	DNA input normalisation
trnah chr2 R	GCTGTAAAGTCAGTCGCTCGTG	DNA input normalisation
Sec1lox qu F	CATGGTTGGCAGATGAGCAAC	Rad21 genomic depletion quantification
Sec1lox qu R	CTACTTTCCCGCTAGCAACTG	Rad21 genomic depletion quantification

Table 6.1: Genomic qPCR primers.

Name	Sequence	Aim
Actb Fw	GGCTGTATTCCCCTCCATCG	Housekeeping
Actb Rv	CCAGTTGGTAACAATGCCATGT	Housekeeping
Esrrb Fw	GCACCTGGGCTCTAGTTGC	
Esrrb Rv	TACAGTCCTCGTAGCTCTTGC	
Gabpa Fw	CTGCTGCACTGGAAGGCTATAG	
Gabpa Rv	GGTGAGGTCTATATCGGTCATGC	
Gabpb2 Fw	AGAGATGCCAGGACCAAAGTGG	
Gabpb2 Rv	GTTGCCCAATGCAGAGCTGTCA	
Lmna Fw	ACCCCGCTGAGTACAACCT	Housekeeping
Lmna Rv	TTCGAGTGACTGTGACACTGG	Housekeeping
Myc Fw	GCCCAAATCCTGTACCTCGTCC	Amplifies Myc
Myc Rv	CTCTTCTCCACAGACACCACATCA	Amplifies Myc
Max Fw	ACCATAATGCACTGGAACGAAA	Amplifies Myc partner
Max Rv	GTCCCGCAAACCTGTGAAAGC	Amplifies Myc partner
Mdx1 Fw	AGATGCCTTCAAACGGAGGAA	Amplifies Myc partner
Mdx1 Rv	CAAGCTCAGAGTGGTGTGTGCG	Amplifies Myc partner
Nrf1 Fw	GGCAACAGTAGCCACATTGGCT	
Nrf1 Rv	GTCTGGATGGTCATTTCACCGC	
Tfam Fw	GAGGCAAAGGATGATTCGGCTC	
Tfam Rv	CGAATCCTATCATCTTTAGCAAGC	
Tfb1m Fw	CTGGTGGTTGAAAAGGACACTCG	
Tfb1m Rv	CCACTGTCTTCTAATGTTGCCTG	
Tfb2m Fw	GCACCAAAACCCATCCCGTCAA	
Tfb2m Rv	CTTTCAAAGGCAACCACTCTGGC	
Polg2 Fw	GCCTGGCTATTTCGAAACTGTC	
Polg2 Rv	CCGTTTTCCAGCGTAGTCTCAG	
Polrmt Fw	AGACCTGTGCTCCTATGCAGCT	
Polrmt Rv	ACCAGGTCAGTGAAGAGCAGCT	
Ppara Fw	ACCACTACGGAGTTCACGCATG	
Ppara Rv	GAATCTTGCAGCTCCGATCACAC	
Ppard Fw	GGACCAGAACACACGCTTCCTT	
Ppard Rv	CCGACATTCCATGTTGAGGCTG	
Pparg Fw	GTACTGTCGGTTTCAGAAGTGCC	
Pparg Rv	ATCTCCGCCAACAGCTTCTCCT	
Ppargc1a Fw	GAATCAAGCCACTACAGACACCG	
Ppargc1a Rv	CATCCCTCTTGAGCCTTTCGTG	
Ppargc1b Fw	TCCTGTAAAAGCCCGGAGTAT	
Ppargc1b Rv	GCTCTGGTAGGGGCAGTGA	
Pprc1 Fw	CAGGAGAAGAAGCCCTTAGACC	
Pprc1 Rv	CTTTCGCCAAGAGTGAGACAG	
Ptpmt1 Fw	CGATTCCCTGTGCAACACCTCGA	
Ptpmt1 Rv	CCTGGTACTTGAGAGCAAACCTGG	
Rad21 ex6-7 Fw	AGCACCAGCAACCTGAATGA	Rad21 region cut by CreERT2
Rad21 ex6-7 Rv	GATCGTCAAAGATGCCACCA	Rad21 region cut by CreERT2
Ubc Fw	AGGAGGCTGATGAAGGAGCTTGA	Housekeeping
Ubc Rv	TGGTTTGAATGGATACTCTGCTGGA	Housekeeping
Zfx Fw	CCTCGGCAGACTGGCTAAACAA	
Zfx Rv	GCAAGGGTAGACAGTCAGAGGA	

Table 6.2: RT qPCR primers.

Western blot

Protein samples

First, cells were detached by incubating them with PBS + 5mM EDTA (Invitrogen 15575020) for 10 min at RT. Then, cells were centrifuged (1200 rpm, 5 min) and resuspended in 100 μ L of protein sample buffer (5% SDS, 10% glycerol, 60mM Tris-HCl pH). Samples were sonicated, and protein content was quantified using the Qubit Protein Assay Kit protocol (Thermofisher, Q33211). Then, 0.001% bromophenol blue and 5% beta-mercaptoethanol were added.

Gel electrophoresis

20 μ g of protein were loaded per sample and 10 μ L of protein ladder (Life Technologies, 10748010) were loaded into a precast 7.5% acrylamide precast resolving gel (Bio-Rad, 455-1086). Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out with using electrophoresis running buffer (1.5% Tris, 7.2% glycine, 0.5% SDS). Gels were then run at 25 mA/gel, 120V for 1h.

Membrane transfer

Once finished, proteins were transferred to an Immobilon-FL PVDF Membrane (Sigma-Aldrich, IPFL00010) cut to fit the gel. The transfer was performed using a Semi-Dry Electrophoretic Transfer Cell system (Bio-Rad, 1703940) in transfer buffer (48mM Trizma base, 39mM glycine, 0.037% SDS and 20% methanol). Membrane transfer was run at 140 mA/gel, 20V for 1 h and 10 min.

Primary antibody incubation

First, to avoid unspecific antibody binding, membranes were blocked for 1h in blocking solution. Then, incubation with the primary antibody was performed at 4oC overnight. For RAD21 protein detection, a dilution in blocking buffer of 1:1000 of primary antibody (Abcam ab154769)

was used, and for ACTB 1:20000 of primary antibody (Santa Cruz sc-69879). The next day, membranes were washed three times with TBS + 0.1% Tween 20.

Secondary antibody incubation

Secondary antibody incubation was performed at 1:10000 dilution in blocking solution for 1h. The antibodies used were anti-rabbit (Invitrogen, A11011) and anti-mouse (Invitrogen, A11004). Then, membranes were washed three times with TBS + 0.1% Tween 20 for 15 min.

Membrane imaging and quantification

Membranes were imaged using a LI-COR Odyssey imaging system, and analysis was performed on Image Studio Lite.

Immunofluorescence

NF κ B immunofluorescence

200.000 cells were seeded onto sterile 13 mm coverslips and left in the 37 °C incubator overnight. The next day, they were activated with LPS for 0 (untreated), 1 and 2h. After 3 washes with PBS, fixation was performed with 4% formaldehyde for 15 min. Then, cells were permeabilised with 0.1% Triton X-100 for 10 min. To block unspecific antibody binding, cells were incubated with 10% serum for 30 min. Then, cells were incubated with p65 (NF κ B subunit) ab7970 antibody at 1:100 dilution in 10% serum for 1h. Afterwards, cells were incubated with goat anti-rabbit 488 at 1:750 dilution in 10% serum. From the fixation, all incubations were performed at RT, and three PBS washes were performed in between all different steps. In the end, samples were mounted using ProLongTM Gold Antifade Mountant with DAPI (Invitrogen).

NF κ B immunofluorescence image acquisition and analysis

Four 3D stacks were imaged in each sample with a Leica SP8 microscope (between 113 to 340 cells imaged per sample, 1024x1024 pixels per image, with a pixel size of 0.2027x0.2027 μ m, x40 oil objective). Their maximum projections were analysed using a pipeline in CellProfiler (Carpenter et al. 2006) v.2.2.0 measuring fluorescence coming from the secondary antibody from the nuclei and cytoplasm of cells. In short, the pipeline first identifies nuclei using the IdentifyPrimaryObjects and then the cell outline is delimited with the module IdentifySecondaryObjects. Then, we measured the correlation between the DAPI signal and the fluorescence from the secondary antibody. A cell was considered positive for translocation if this correlation was higher than 0.5. Statistical significance was assessed using a 2-way ANOVA test with Sidak Correction using GraphPad.

RAD21 immunofluorescence

200.000 cells were seeded onto sterile 13 mm coverslips and left in the 37 °C incubator overnight. A day after, cells were washed with PBS and fixed for 5 min with 300 μ L ice-cold 100 % methanol at RT. Then, three washes with PBS were performed, and cells were stored at 4 °C overnight. The next day, cells were permeabilised with 200 μ L/sample 0.5% Triton X-100 in PBS 10 min. Cells were incubated with 10% serum for 30 min, to prevent unspecific antibody binding. Primary antibody incubation was performed at 4 °C overnight using Rad21 ab154769 1:500 in 10% serum. The next morning, cells were washed three times with PBS. Secondary antibody incubation was performed with goat α -rabbit 488 A11034 1:500 for 1h at RT.

RAD21 immunofluorescence image acquisition and analysis

3D stacks were acquired Leica SP8 microscope with a voxel size of 0.144x0.144x0.99 μ m, x40 oil objective). For the analysis, nuclei and cell outlines were identified from maximum projections by CellProfiler v2.2.0 (ibid.) as primary and secondary objects. Then, for each cell, the RAD21 level for each cell is defined as the mean nuclear intensity of the antibody channel minus the cytoplasmic mean. This metric normalised against variable levels of background.

FACS

Cells were lifted, centrifuged at 1200rpm for 5 min and resuspended in 1mL of FACS buffer (PBS Mg²⁺ and Ca²⁺ free, 2% FCS, 2mM EDTA). FACS analysis was carried out by a BD LSR II.

Mitochondria quantification

Relative mtDNA content

Relative mtDNA content was calculated as described in Rooney et al. 2015. Briefly, RTqPCR was performed on genomic DNA. mtDNA was quantified by amplifying a region of the mtND1 gene, whereas nuDNA was quantified by amplifying a chromosome 2 region. Primer sequences are presented in Table 6.1.

$$Relative_{mtDNA} = 2x2^{mtDNA Ct - nuDNA Ct} \quad (6.1)$$

Mitochondria staining

100.000 cells were plated in each well of a chamber slide (Ibidi, 80826). The next day cells were incubated in serum-free media with 10 μ M MitoTracker Red CMH2XRos (ThermoFisher M7512) and 5 μ M CellTracker Green (Invitrogen C2925) at 37°C for 20 minutes. Then, cells were washed thoughtfully with PBS. Cells were imaged within the same day.

Mitochondria imaging

Images here acquired using a Leica TCS SP5 inverted microscope using a dry 20x/0.75 NA objective. 5-10 manually selected sites were imaged per sample to ensure that they have similar cell confluence. The pixel size was 189 nm, and the z-step size was 500 nm.

Mitochondria image analysis

Images were analysed using a pipeline in CellProfiler v3.0.0 Carpenter et al. 2006 First, cells were identified as primary objects from the Green CellTracker channel. Then, mitochondria signal was enhanced by enhancing speckle like signal using the EnhanceOrSuppressFeatures module. Subsequently, mitochondria were identified from the resulting image as primary objects. All mitochondria that overlap with a cell were assigned to that cell. Then, area and intensity were measured in each object. The fraction of cell covered by mitochondria was the mitochondria area over the cell area.

Single molecule Fluorescent In Situ Hybridization

Validation of bDNA FISH

smFISH for validations

200.000 cells/well were seeded per 13mm glass coverslips. The next day, cells were activated with LPS, and their cytoplasms were stained with Red CellTracker. ViewRNA™ ISH Cell Assay Kit (QVC0001), and gene-specific probes were purchased from Affymetrix (Table 6). We followed the smFISH protocol as described by the manufacturer. In the last step, coverslips were mounted, and nuclei were stained using ProLong Gold Antifade (Thermo Scientific).

smFISH image acquisition and analysis

Images were acquired using a Leica TCS SP8 inverted microscope using an oil-immersion 40x/1.30 NA objective. The imaging fields and the number of z-planes were selected manually. We imaged between 15 to 30 images per sample. The pixel size was 101.34 nm, and the z-step size was 340 nm. For image analysis, the MATLAB version of CellProfiler (v1.0.0, *ibid.*) was used with specific image-based transcriptomics modules (Battich et al. 2013, Stoeger et al. 2015) for nuclei, cytoplasm and dot identification.

High throughput smFISH

We seeded 100.000 cells/well in μ -Slide 8 Well IbiTreat Chamber Slides (Ibidi, 80826) on day 6. On day 7, cells were activated with LPS. If probes against exons were used, we stained with Red CellTracker, whereas if only intronic probes were used, we skipped the Red CellTracker staining step. To perform bDNA FISH, we followed the QuantiGene View RNA HC Screening Assay (QVP0011, Affymetrix) protocol with the following modifications: The fixation buffer included 2.5% of Glacial Acetic Acid which enabled nuclear detection of transcripts, and PBS buffers contained 2nM of an RNase inhibitor (Ribonucleoside Vanadyl Complex, NEB S1402S). Finally, nuclei were stained using DAPI (ThermoScientific, 62248) 1:3000 in PBS for 5 min at RT. Samples were stored in PBS at 4°C, and imaged within the next 5 days. Probes used are depicted in (Table 6).

target (μ L)	λ excitation (nm)	catalog #
<i>Actb</i>	488	VB4-10432
<i>Actb</i> intron	550	VB1-20973
<i>Atf3</i>	650	VB6-15721
<i>Cd40</i>	550	VB1-15088
<i>Cxcl10</i>	650	VB6-10663
<i>Fh1</i> intron	488	VPNKRGW
<i>Enh</i>	650	VPMFWXN
<i>Hprt</i> intron	550	VPMFWWY
<i>Ifit1</i>	550	VB1-12995-VC
<i>Il12b</i>	650	VB6-20182
<i>Il12b</i> intron	488	VB4-20975
<i>Ifnb1</i>	488	VB4-12687-VC
<i>Ifnb1</i>	550	VB1-13004-VC
<i>Mx1</i>	650	VB6-16985-06
<i>Peli1</i>	488	VB6-3205469
<i>Peli1</i>	550	VB1-3035672-VC
<i>Peli1</i> intron	488	VB4-6000742
<i>Sertad2</i>	650	VB6-3204369-VC
<i>Sertad2</i> intron1	550	VPDJXFM
<i>Spred2</i>	488	VB4-6000742
<i>Spred2</i> intron1	650	VB6-6000741
<i>Stat1</i>	550	VB1-11676
<i>Tnf</i>	488	VB4-18211

Table 6.3: **bDNA FISH probes.** All probes were purchased from Affymetrix eBiosciences.

High throughput smFISH image acquisition

Unless stated at the figure legend, a Leica SP5 inverted microscope acquired the images with a dry 20x/0.75 NA objective. The voxel size was 0.189 μm in xy and 0.5 μm in xz. We used the microscope in the resonant mode with a bidirectional scan. A line average of 8-16 was used to reduce noise. Imaging fields were selected automatically, with human supervision to ensure that there were cells. The Ibidi μ -Slide allowed to image up to 8 samples in one night coupled to a MatrixScreener software. Fluorescence reflection allowed automatic focusing by identifying the bottom of the slide.

High throughput smFISH image analysis

Cell, nuclei and mature mRNA identification

In short, nuclei and cell outlines were identified from maximum projections using CellProfiler v2.2.0 modules as primary and secondary objects (Carpenter et al. 2006). To avoid considering two cells as one, size filter was applied to exclude any cell with nuclei or a cytoplasm bigger than a threshold. Outlines were visually inspected, and then loaded in FISHQuant (FQ) (Mueller et al. 2013). FQ counted mRNA molecules by fitting 3D Gaussian in the 3D stacks.

Identification and quantification of transcriptional bursts

Nascent mRNA molecules accumulate at TSSs during a transcriptional burst. Due to their signal accumulation, active TSSs can be identified as big blobs within nucleus compared to small spots which correspond to individual mRNA molecules. Active TSSs and individual molecules were identified by FQ using the intensity from intronic probes. No more than 2 TSSs were allowed to be found in a single cell. However, FQ failed to fit a Gaussian function for most active TSSs, thus, it could not perform the quantification. To overcome this issue, we compared each burst to the average image of all individual nascent mRNA molecules by identifying the brightest pixel in the TSS and the average spot. Then, we select all the pixels around it up to a radius of 1 or 2 pixels and calculate their integrated intensity. Burst size estimation was the ratio between the integrated intensity of these pixels in the TSS and the same area in the averaged individual mRNA image.

Association of multiple gene burst to a particular locus

Our macrophages are diploid cells. Therefore, a burst can only belong to one of the two chromosomes (locus1 and locus2), and each locus can only have one active TSS from each gene. The TSSs are assigned based on their distances prioritising that the closest TSS pairs belong to the same locus. This approach is based on active TSSs distances and was written in Julia v1.1.4.

1. Identify TSS outlines using FQ (maximum of 2 TSS per cell).
2. Set a maximum distance at which TSSs from the same locus can be. This threshold associate active TSSs to a different locus in cases where one locus is expressing one gene and the other, only one other gene. The threshold distance was set as 5 μm , which is the typical distance at which two active TSSs from the same gene were in a single macrophage.
3. For each cell:
 - (a) Measure TSS distances as distances between their brightest pixels in the 3D stack. Do this between each gene active TSS and all the other genes active TSS present in the cell. Make a table.
 - (b) Order the TSS pairs by their distances.
 - (c) Unless the first distance is larger than the distance threshold, associate the closest pair TSS to locus1. Otherwise, associate each one of the TSSs to different locus.
 - (d) Then, for each TSS pair from the second closest to the furthest:
 - i. If a particular active TSSs has already been associated to one locus, add the other pair TSSs to the same locus unless the distance between them is larger than the threshold, or it has already been associated to the other locus, or the locus already has a TSS of the same gene.
 - ii. If one locus has already associated an active TSS from one of the genes, but it is not one of the TSS from the pair, then add that active TSSs to the other locus. Add also the other TSS from the pair, unless the distance between them is larger than the threshold or it has already been associated with the opposite locus, or the locus has already a TSS from the same gene.

smFISH overdispersion

smFISH overdispersion data

Two sets of genes, set 1 (*Actb*, *Il12b*, *Stat1*) and set 2 (*Tnf*, *Cd40*, *Atf3*) were measured in WT and *Rad21* KO macrophages in untreated samples, 2h and 8h after activation with LPS (10ng/ml, Sigma-Aldrich L789). Each sample had three replicates. Cells were then imaged using a SP8 microscope (x40 oil objective, 0.3 μm xy resolution, 0.3 μm xz resolution), images were then analysed using Lukas Pelksman lab pipelines (Battich et al. 2013; Stoeger et al. 2015).

smFISH overdispersion calculation

CV2 and the mean number of mRNA dots is calculated in each replicate. Then, using the R package statmod, we fitted a gamma family generalized linear model (glmGamFit function) of CV2s with a0 as 1 and altilde corresponding to 1/mean counts. Then, the residual dispersion is calculated as the log10 of the CV2 minus the log10 of the fitted values.

The code (In R):

```
library("statmod")
df.FISH <- read.csv("smFISH_MEAN_CV.csv") fit <- glmGamFit(
cbind(a0 = 1, altilde = 1/df.FISH$MeanCounts), df.FISH$CV2)
res <- log10(df.FISH$CV2) - log10(fit$fitted.values)
df.FISH$Overdispersion <- res
write.csv(df.FISH, "smFISH_MEAN_CV_Overdispersion.csv")
```

Telegraph model simulations

Telegraph model

A promoter of a gene can switch between an on and an off state with rates k_{on} and k_{off} . When a gene is on the on state, it can produce mRNA molecules at a rate k_{tc} . Then, those mRNA

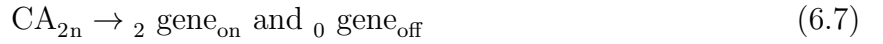
molecules can be degraded at a rate k_{deg} .



Where $\text{gene}_{\text{on}} + \text{gene}_{\text{off}} = \text{ploidity}$.

Telegraph model states

Assuming that cells are diploid, so $\text{ploidity} = 2$. Then, if n = number of mRNA molecules, each cell (CA) state can be represented as:



Telegraph model ODE

$$\frac{dCA_{0n}}{dt} = (n+1)k_{deg}CA_{0n+1} + k_{\text{off}}CA_{1n} - (2k_{\text{on}} + nk_{deg})CA_{0n} \quad (6.8)$$

$$\begin{aligned} \frac{dCA_{1n}}{dt} = & (n+1)k_{deg}CA_{1n+1} + 2k_{\text{off}}CA_{2n} - (k_{\text{on}} + k_{\text{off}} + nk_{deg} + k_{tc})CA_{1n} \\ & + 2k_{\text{on}}CA_{0n} + k_{tc}CA_{1n-1} \end{aligned} \quad (6.9)$$

$$\begin{aligned} \frac{dCA_{2n}}{dt} = & (n+1)k_{deg}CA_{2n+1} + k_{\text{on}}CA_{1n} - (2k_{tc} + 2k_{\text{off}} + nk_{deg})CA_{2n} \\ & + 2k_{tc}CA_{2n-1} \end{aligned} \quad (6.10)$$

Telegraph model steady-state simulations

1030301 Ordinary differential equation (ODE) simulations were performed. k_{deg} was constant and equal to 0.01 and $k_{\text{on}}, k_{\text{off}}, k_{tc}$ acquired values from 0 to 1 with an interval of 0.01. An ODE

transition matrix is constructed using telegraph model ODE and the steady-state is found using the nullspace function from LinearAlgebra.jl. The solution is divided by its sum to produce probabilities between 0 and 1. The maximum number of mRNA molecules that a cell can have was set as 300. Code was written in Julia v0.7.0.

Calculation of bursting parameters from the simulations

$$BurstFrequency = \frac{k_{on}}{k_{on} + k_{off}} \quad (6.11)$$

$$BurstHalfLife = \frac{\ln(2)}{k_{off}} \quad (6.12)$$

$$BurstSize = BurstHalfLife * k_{tc} = \frac{\ln(2)}{k_{off}} * k_{tc} \quad (6.13)$$

Expression and mutation data of human AML

Dataset contained 672 AML tumour samples from 562 individuals was published in Tyner et al. 2018. From those, 452 patients have expression data available and for 408, expression and survival data information. Data downloaded from Synapse and Vizome. Bioinformatic analysis carried out by Dr Yi-Fang Wang.

Mutation gene sets

Mutations in STAG2, SMC3, SMC1A, RAD21, STAG1, PDS5B and PDS5A were considered as cohesin mutations. Mutations in CSTF2T, DDX1, DDX23, DHX32, HNRNPK, METTL3, PLRG1, POLR2A, PRPF3, PRPF8, RBMX, SF3B1, SNRNP200, SRRM2, SRSF2, SRSF6, SUPT5H, TRA2B, U2AF1, U2AF1L4 and U2AF2 were considered as spliceosome mutations. In ACVR2B, ADRBK1, AKAP13, BUB1, CPNE3, DCLK1, MAPK1, MYLK2, MYO3A, NRK, PRKCG, RPS6KA6, SMG1, STK32A, STK33, STK36, TRIO, TTBK1, WNK3 and WNK4 were considered as mutations in Serine/threonine kinases (STK). Mutations in ABL1, DYRK4,

EPHA2, EPHA3, JAK3, MST1R, OBSCN, PDGFRB and WEE1 were considered as Other tyrosine kinases (OTK) mutations. NRAS and KRAS were considered as RAS mutations. Finally, mutations in IDH included defects in IDH1 and IDH2 genes. Dr Sergi Cuartero defined these gene subsets.

Patient gene expression processing

Genes counts for each patient samples were converted to counts per million (CPM) using function CPM from the R/Bioconductor package edgeR (3.16.5; McCarthy et al. 2012; Robinson et al. 2010). Lowly expressed genes were removed if CPM was smaller than 1 in one of all group samples. Normalisation was then performed by the trimmed mean of M-values (TMM; Robinson and Oshlack 2010) using the calcNormFactors function in edgeR. R/Bioconductor package limma (3.30.12; Ritchie et al. 2015) which was applied for the differential expression analysis, and the function voom (Law et al. 2014) was used to transfer raw counts to log2-counts per million (logCPM) using the library sizes and normalise factors obtaining from the normalisation step. Differential expression analysis between cohesin mutant and non-cohesin mutant patients was performed by the lmFit and eBayes (Phipson et al. 2016) functions in limma.

Expression data of mouse LSK cohesin knockdown

Published RNA-Seq dataset for RNA-sequencing control (Ren) and cohesin (Stag2, Smc1a, and Smc3) knockdown cells (Mullenders et al. 2015) were retrieved from Gene Expression Omnibus (GEO) under accession GSE72368.

Single-end 50bp RNAseq reads were aligned against Ensembl mouse genome (NCBIM37) with Tophat2 (2.0.11; Kim et al. 2013). Reference sequence assembly and transcript annotation that obtained from Illumina iGenomes (<https://support.illumina.com/sequencing/sequencing-software/igenome.html>). Gene-based read count was obtained using featureCounts function from Rsubread Bioconductor package (1.24.2; Liao et al. 2013, 2014). Normalisation was performed in DESeq2 Bioconductor package (1.14.1; Love et al. 2014). Differentially expressed

gene analysis was also performed with DESeq2 and defined with Benjamini-Hochberg adjusted p-value<0.05.

***Rad21* KO macrophage data**

Macrophage RNAseq data

Data from our lab published in Cuartero et al. 2018. In sort, 2 million cells were sorted for each condition from three biological replicates. RNA was extracted with an RNeasy Minikit and using Qiashredder (Qiagen). RNA was assessed for quality (Bioanalyzer, Agilent) and quantified (Qubit, Invitrogen). ERCC RNA Spike-Ins (Invitrogen, 4456740) were added, and strand-specific libraries prepared from 750 ng of total RNA using TruSeq Stranded Total RNA Kit (Illumina, RS-122-2201). RNA was prepared by Dr Sergi Cuartero. Libraries were sequenced on an Illumina Hiseq2500 (v4 chemistry), generating 40 million paired-end 100-bp reads per sample. Reads were aligned to mouse genome mm9 using Tophat2 (Kim et al. 2013) with arguments `-library-type fr-firststrand -b2-very-sensitive -b2-L 25`, with gene annotations from Ensembl version 67. RUVseq Bioconductor package was applied with RUVs function and $k=3$ (1.18.0; Risso et al. 2014) for ERCC spike-ins and replicate samples as controls, then, differential expression analysis was performed using DESeq2 (Love et al. 2014). FPKM values were computed in R. RNA content calculated as in Kouzine et al. 2013 by Dr Gopuraja Dharmalingam. Differential expression analysis and normalisation were carried out by Dr Yi-Fang Wang.

The RNA from mouse Rad21-TEV macrophages was extracted by DrFelix Weiss and published in (Cuartero et al. 2018). RNA was extracted using the PicoPure RNA Isolation Kit (ThermoFisher) and then treated with RNase-Free DNase Set (Qiagen). Three biological replicates were sequenced per condition. Libraries were prepared from 100ng of RNA per sample using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina using ribozero ribosomal RNA (rRNA) depletion. Sequencing was carried out on an Illumina Hiseq2500 (v4 chemistry), generating 40 million paired-end 100-bp reads per sample. Tophat (Kim et al. 2013) was used for alignment to the mouse genome mm9 with arguments `-library-type fr-firststrand -b2-very-sensitive -b2-L 25`, with gene annotation from Ensembl version 67. Read counts on genes were

summarised using HTSeq-count. Differentially expressed genes were identified using DESeq2 (Love et al. 2014).

GROseq data

Data published in Cuartero et al. 2018 and produced by Dr Sergi Cuartero. GROseq libraries were prepared from two biological replicates per condition from 5 million cells. First, a nuclear run-on was conducted. Then, RNA was extracted using Trizol (Ambion) and treated with Turbo DNase (Ambion AM1907). Subsequently, RNA was fragmented (Ambion, AM8740) and purified on P-30 columns (Bio-Rad 732-6250). After RNA phosphorylation with PNK (New England Biolabs, Y904L), RNA was purified again using anti-BrdU beads (SantaCruz, sc-214314). cDNA was obtained through reverse transcription with barcoded oligonucleotides. This cDNA was then purified, amplified by PCR and purified again from a gel (Novex 10% TBE gel). cDNA was cleaned following the ChIP DNA Clean Concentrator Kit (Zymo D5205). Sequencing products were 50-bp single-end reads. The 10 most 3 bases were discarded based on fastqc quality assessment. Reads were aligned to mouse genome mm9 using bowtie with arguments -l 30 -m 10 -n 2 -trim3 10. Read counts on enhancers were computed using summarizeOverlaps function from GenomicAlignments R Package. DESeq2 (Love et al. 2014) identified DE genes and enhancers using the enhancer macrophage set described in Ostuni et al. 2013.

Bioinformatic analysis

LPS-inducible gene subset

560 genes strongly induced by LPS in macrophages from Figure 3 of Bhatt et al. 2012.

Constitutive gene subset

Constitutive genes were 5481 genes non-DE by RNA-seq between untreated WT macrophages and 2h and 8h LPS-activated WT macrophages (DEseq2, $\text{padj} > 0.05$ and $|\log_2\text{FoldChange}| < 0.5$)

LPS-repressed gene subset

LPS-repressed genes were 4123 genes up-regulated in WT 8h vs WT untreated macrophages (DESeq2, $\text{padj} < 0.05$ and $\log_2\text{FoldChange} > -0.5$).

Macrophage enhancers

Macrophage enhancers come from Ostuni et al. 2013. Superenhancers were identified using ROSE (Whyte et al. 2013) in Cuartero et al. 2018.

Gene Set Enrichment Analysis

GSEA (Mootha et al. 2003; Subramanian et al. 2005) was performed using as ranks $\log_2\text{FoldChanges}$ from DESeq2 results using MSigDB gene sets (Liberzon et al. 2015). Low expressed genes were excluded by using the DESeq2 independent filtering approach.

RNA pol II pause ratio calculation

Calculated from GROseq reads as in Day et al. 2016:

$$\text{Pausing Ratio} = \frac{\text{ReadCount}(TSS)/L1}{\text{ReadCount}(\text{Gene Body})/L2} \quad (6.14)$$

Where L1 is the length of the gene TSS region ($TSS \pm 500$ bp) and L2 is the length of gene body region (+500 bp past TSS to gene TES).

Known Motif Promoter Enrichment Analysis

Enrichment of known transcription factor motifs in gene promoters ($TSS \pm 2\text{kb}$) was performed using Homer's findMotifs.pl program (Heinz et al. 2010).

Voronoi maps

KEGG gene ontologies (Kanehisa and Goto 2000) were visualised using Voronto (Santamaría and Pierre 2012) using the mean FPKM across replicates as input.

Single cell RNAseq

WT and *Rad21* KO macrophage scRNAseq

WT and *Rad21* KO macrophages were sorted using FACS into 96 well plates by Sergi Cuartero and sent to CNAG-CRG where they sequenced and aligned them using the MARS-Seq technique (Jaitin et al. 2014). This approach uses a two-level indexing of 192 cells per pool and multiple pools per sequencing lane. Sequencing was carried out as paired-end reads, wherein the first read contains the transcript sequence and the second read the cell barcode and UMIs. Quality check of the generated reads was performed with the FastQC quality control suite. Samples that reached the quality standards were then processed to deconvolute the reads to single-cell level by de-multiplexing according to the cell and pool barcodes. Reads were filtered to remove polyT sequences. Sequencing reads from human, mouse, or canine cells were mapped with the RNA pipeline of the GEMTools 1.7.0 suite (Marco-Sola et al. 2012) using default parameters (6% of mismatches, minimum of 80% matched bases, and minimum quality threshold of 26) and the genome references for human (Gencode release 24, assembly GRCh38.p5), mouse (Gencode release M8, assembly GRCm38.p4), and dog (Ensembl v84, assembly CanFam3.1). The analysis of spike-in control RNA content allowed us to identify empty wells and barcodes with more than 15% of reads mapping to spike-in artificial transcripts were discarded. In addition, cells with less than 60% of reads mapping on the reference genome or more than 2×10^6 total reads were discarded. Gene quantification was performed using UMI corrected transcript information to correct for amplification biases, collapsing read counts for reads mapping on a gene with the same UMI (allowing an edit distance up to two nucleotides in UMI comparisons). Only unambiguously mapped reads were considered. Genes not expressed in at least 5% of the cells were discarded. Thresholds were set to reduce technical noise, but to conserve the sensitivity to identify low frequency outlier cell populations and to capture differences between fresh and cryopreserved cells. Sequence alignment and initial analysis were performed by Gustavo

Variability analysis

Dr Nils Eling carried out scRNAseq variability analysis at Jonh Marioni's lab in Cambridge. He used BASiCS to detect differentially expressed and differentially variable genes (Eling et al. 2018; Vallejos et al. 2015). Genes were filtered, so all genes considered had an average posterior mean expression > 1 in each population. Only genes with no change in mean expression were considered (FDR cutoff of 0.05) for Figure 3.20.

Module score

Measure developed in Shalek et al. 2014 and used in this thesis with some modifications. It measures the level of expression of a subset of genes in scRNAseq. First, for each cell, we transform the gene counts into counts per million (CPM). Then, for each gene, we calculate α , σ , and μ for each sample. α is the fraction of cells with the gene detected. μ is the mean CPM of cells with counts detected for that gene, and σ is the standard deviation. Then, each cell is fitted into a logarithmic regression using the ScikitLearn python module to obtain B_0 , and B_1 for each cell. Then, we calculate the gene weight (ω) as:

$$\omega(C, G) = \frac{1}{1 + e^{-(B_{C,0} - B_{C,1} * \mu)}} \quad (6.15)$$

Weight is set as 1 if the transcript is detected in the cell. Alternatively $\text{weight} = 1 - \omega(C, G)$ if there is no transcript detected in the cell. Then, for each gene in the module, we calculate its induction values as the CPM of that gene in the cell divided by the average expression + 1 of the gene in a control sample (often WT untreated).

Finally, the gene module score for each cell is the weighted mean of the induction values of all the genes in the subset in the cell.

Statistical analysis for comparisons

qPCR data comparisons

Equal variance t-test as implemented in HypothesisTests.jl Julia package with the Bonferroni method to correct for multiple testing as implemented in MutipleTesting.jl Julia package. Equal variance t-test function is a two-sample t-test where the null hypothesis is the two distributions have the same mean and variance against the alternative hypothesis that they share the variance but not the mean. Similar results were found when it was not assumed that variances were equal.

Comparison between mitochondria imaging data

Same as the

qPCR data comparisons

section. Replicates were imaged on the same day and were quite similar among them, so there was no need to use a statistical test that considers technical differences among replicates.

Comparisons between the expression log2FoldChanges in one comparison vs the log2FoldChanges in another

Wilcoxon Rank test against the null hypothesis that the distribution of the differences has 0 as the median as implemented in the Julia package HypothesisTests.jl with the Bonferroni method as implemented in MutipleTesting.jl Julia package to correct for multiple testing.

Burst frequency and size comparisons

We used a one-way ANOVA with Tukey HSD correction for TSS intensity comparisons as implemented in R. Replicate is considered an independent variable. For burst frequency compar-

isons, we opted for the Cochran–Mantel–Haenszel test as implemented in R, with the Bonferroni method as implemented in `MutipleTesting.jl` Julia package to correct for multiple testing. This test tests the association between two nominal variables independently of a third called the strata. In this case, the null hypothesis is that the burst frequency between samples is equal. The Bonferroni method corrected for multiple testing. Both statistics had the advantage of taking into account the number of individual measures and the variability across technical replicates.

Chapter 7

Supplementary Material

Gene	Distance to nearest TAD boundary (bp)
<i>Hprt</i>	58686
<i>Spred2</i>	99378
<i>Sertad2</i>	206744
<i>Il12b</i>	211435
<i>Actb</i>	231567
<i>Ifnb1</i>	306302
<i>Peli1</i>	341294
<i>Fh1</i>	1005766

Table 7.1: **Distance to nearest TAD boundary of genes studied for burst frequency and size.** TAD were called as described (Rao et al. 2017) by Ya Guo from mouse BMD macrophages Hi-C data (GSE115524, Mumbach et al. 2019)

Genotype	LPS (min)	<i>Il12b</i>	<i>Ifnb1</i>
WT	0	2	2
WT	120	2	3
<i>Rad21</i> KO	0	2	2
<i>Rad21</i> KO	120	2	3
<i>Rad21</i> KO+WT	0	3	3
<i>Rad21</i> KO+WT	120	3	3

Table 7.2: Number of technical replicates for replicates for mix and unmix samples in *Il12b* and *Ifnb1* smFISH experiments.

Comparison	Gene	log2FoldChange	adjusted p-value
<i>Rad21</i> KO 2h LPS + IFN γ vs <i>Rad21</i> KO 2h LPS	<i>Il12b</i>	-0.39	0.64
<i>Rad21</i> KO 2h LPS + IFN γ vs <i>Rad21</i> KO 2h LPS	<i>Ifnb1</i>	2.82	3.3e-35
<i>Rad21</i> KO 2h LPS + IFN γ vs <i>Rad21</i> KO 2h LPS	<i>Mx1</i>	1.867	1.09e-4
<i>Rad21</i> KO 8h LPS + IFN γ vs <i>Rad21</i> KO 8h LPS	<i>Mx1</i>	1.28	9.8e-36
<i>Rad21</i> KO 2h LPS + IFN γ vs <i>Rad21</i> KO 2h LPS	<i>Ifit1</i>	0.51	0.24
<i>Rad21</i> KO 8h LPS + IFN γ vs <i>Rad21</i> KO 8h LPS	<i>Ifit1</i>	0.71	0.075

Table 7.3: Differential expression analysis in LPS-activated *Rad21* KO macrophages with and without IFN priming by RNAseq.

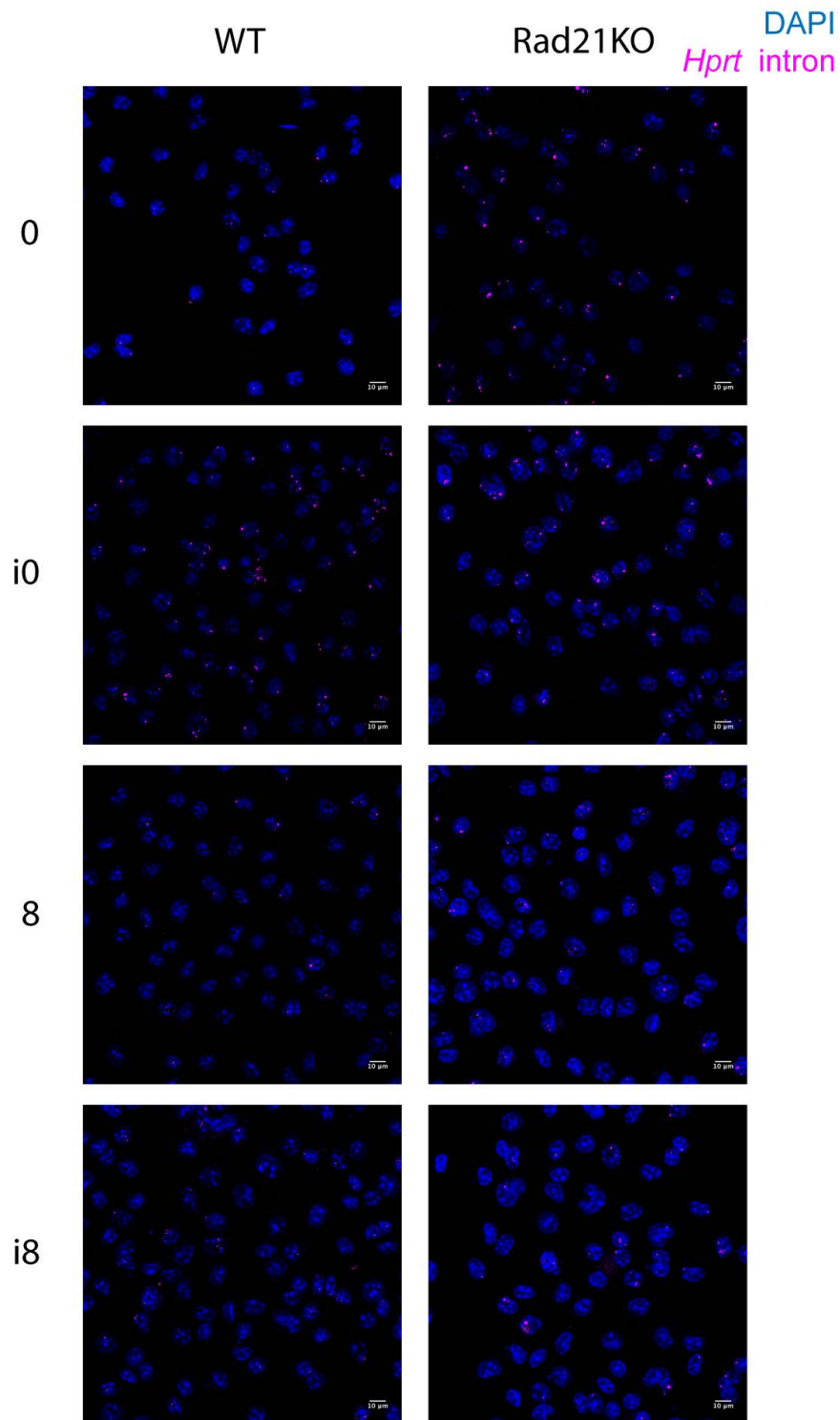


Figure 7.1: *Hprt* bursts. 0, untreated macrophages. i0, 24h pre-treatment with IFN- γ . 8, 8h LPS. i8, 24h pre-treatment with IFN- γ + 8h LPS. i8. Scale bars represent 10 μ m.

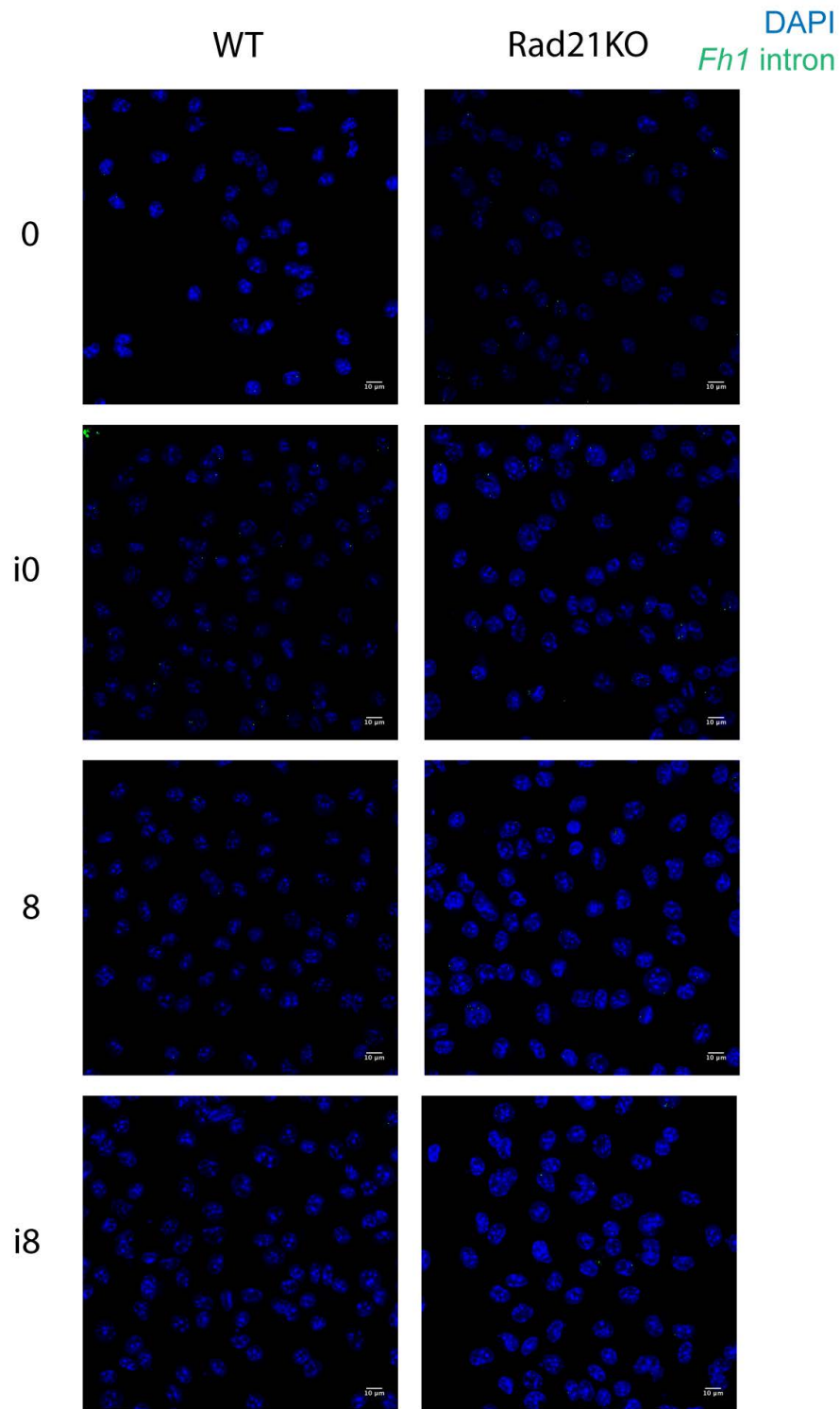


Figure 7.2: ***Fh1* bursts.** 0, untreated macrophages. i0, 24h pre-treatment with IFN- γ . 8, 8h LPS. i8, 24h pre-treatment with IFN- γ + 8h LPS. i8. Scale bars represent 10 μ m.

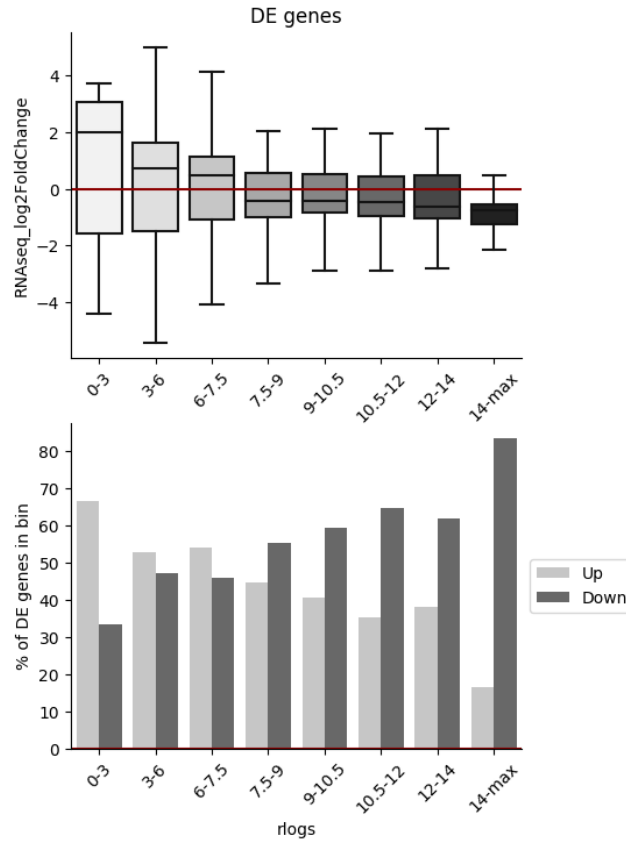
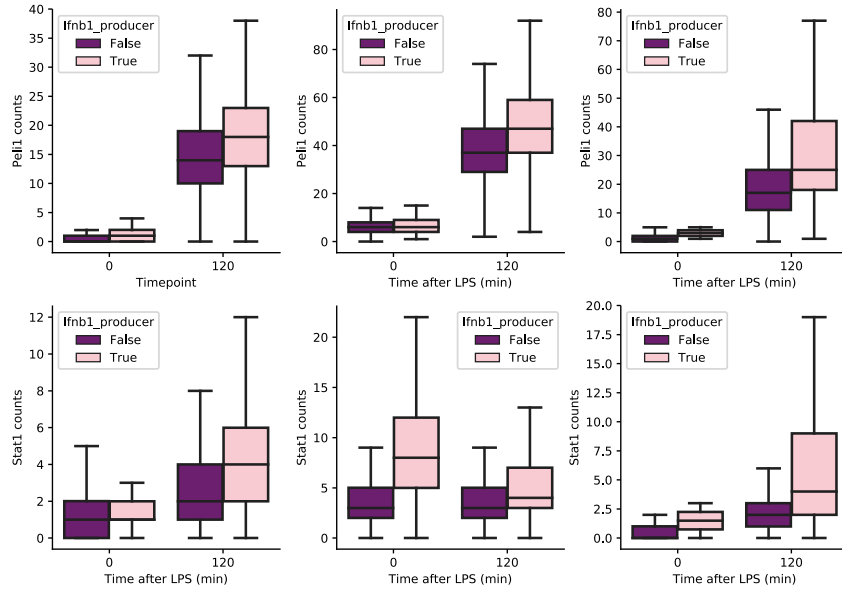


Figure 7.3: **Lowly express genes tend to be up-regulated whereas highly express genes tend to be down-regulated in *Rad21* KO.** First, all genes are sorted into intervals based on the rlog log2 mean between control and cohesin-depleted macrophages. Then, DE genes are selected ($\text{padj} < 0.05$). TOP) Box plots showing the log2 fold changes in each bin. BOTTOM) Proportion of up-regulated (Up) and down-regulated genes (Down) in each bin.

WT macrophages



Rad21KO macrophages

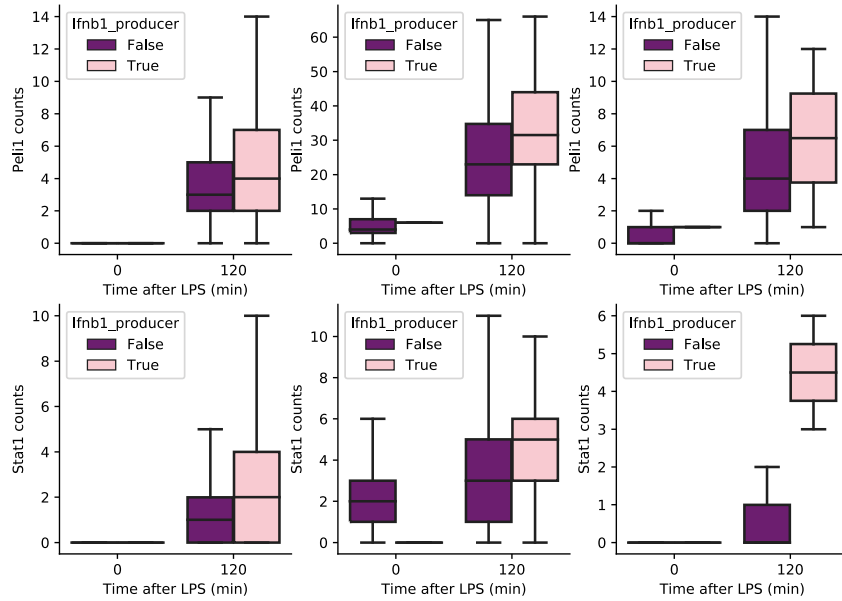


Figure 7.4: **Increase inducible gene expression in *Ifnb1*-producer cells by replicate.** Each column correspond to a replicate. A) WT macrophages. B) *Rad21* KO macrophages. A cell is considered to be a *Ifnb1* producer if it contains 2 of more *Ifnb1* counts

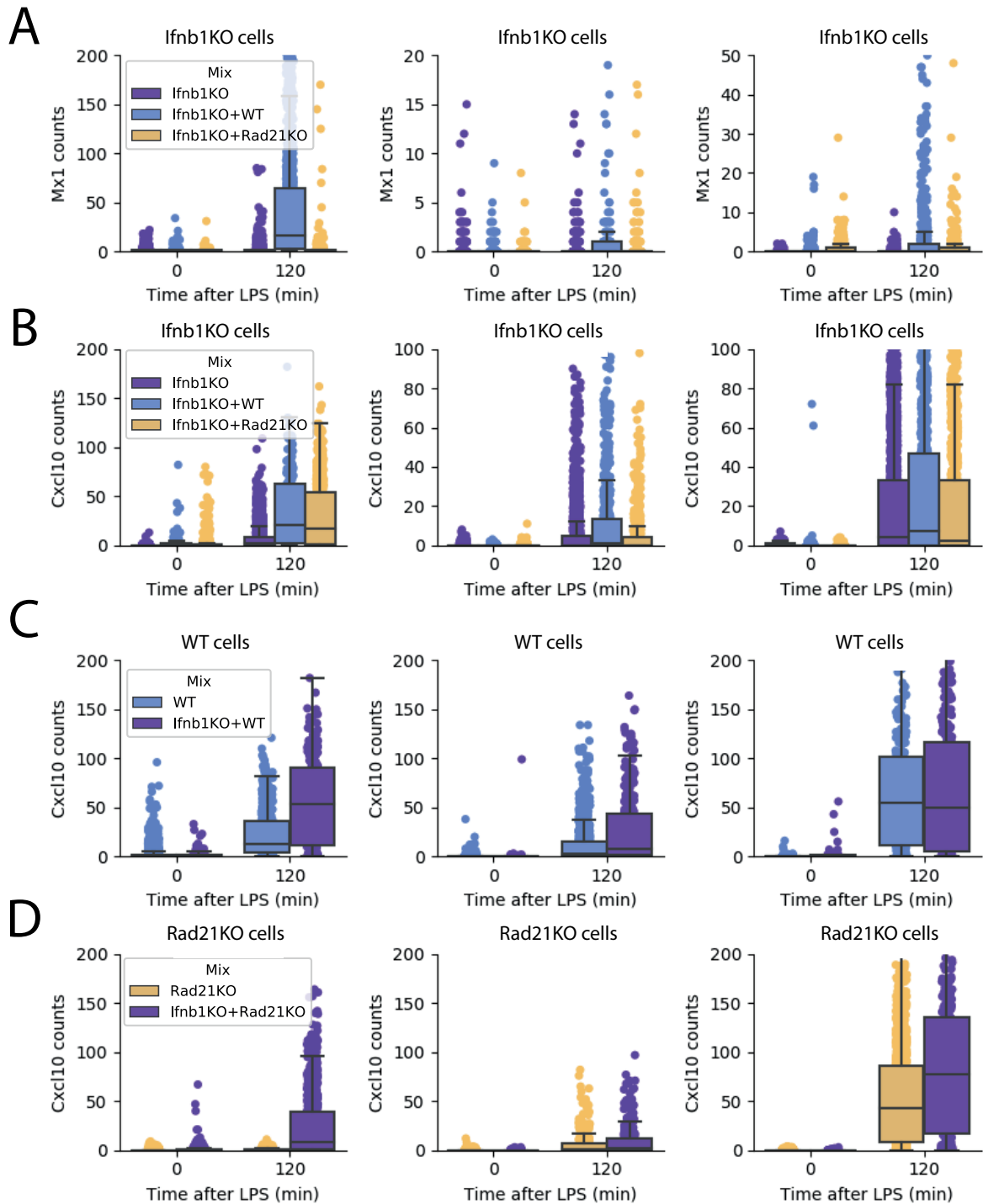


Figure 7.5: *Mx1* and *Cxcl10* counts in Ifnb1KO, WT, *Rad21* KO and Mix populations by replicate. A) *Mx1* counts for Ifnb1KO cells (Green classified cells in Ifnb1KO, Ifnb1KO+WT and Ifnb1KO+*Rad21* KO mixes). B) *Cxcl10* counts for Ifnb1KO cells (Green classified cells in Ifnb1KO, Ifnb1KO+WT and Ifnb1KO+*Rad21* KO mixes). C) *Cxcl10* counts for WT cells (Red classified cells in WT and Ifnb1KO+WT mixes). C) *Cxcl10* counts for *Rad21* KO cells (Red classified cells in *Rad21* KO and Ifnb1KO+*Rad21* KO mixes). Each column is a replicate. 3 biological replicates. 48993 cells images in total.

Bibliography

- Adelman, Karen and John T Lis (2012). “Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans”. In: *Nature Reviews Genetics* 13.10, pp. 720–731.
- Alexander, Jeffrey M, Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D Weiner (2019). “Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity”. In: *elife* 8, e41769.
- Alomer, Reem M, Eulália ML da Silva, Jingrong Chen, Katarzyna M Piekarz, Katherine McDonald, Courtney G Sansam, Christopher L Sansam, and Susannah Rankin (2017). “Esco1 and Esco2 regulate distinct cohesin functions during cell cycle progression”. In: *Proceedings of the National Academy of Sciences* 114.37, pp. 9906–9911.
- Alon, Uri (2007). “Network motifs: theory and experimental approaches”. In: *Nature Reviews Genetics* 8.6, pp. 450–461.
- Altare, Frédéric et al. (1998). “Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency”. In: *Science* 280.5368, pp. 1432–1435.
- Andersson, Robin, Peter Refsing Andersen, Eivind Valen, Leighton J Core, Jette Bornholdt, Mette Boyd, Torben Heick Jensen, and Albin Sandelin (2014). “Nuclear stability and transcriptional directionality separate functionally distinct RNA species”. In: *Nature communications* 5.1, pp. 1–10.
- Andersson, Robin and Albin Sandelin (2019). “Determinants of enhancer and promoter activities of regulatory elements”. In: *Nature Reviews Genetics*, pp. 1–17.
- Arzate-Mejía, Rodrigo G, Angel Josué Cerecedo-Castillo, Georgina Guerrero, Mayra Furlan-Magaril, and Félix Recillas-Targa (2020). “In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*”. In: *Nature communications* 11.1, pp. 1–16.
- Banigan, Edward J and Leonid A Mirny (2020). “Loop extrusion: theory meets single-molecule experiments”. In: *Current Opinion in Cell Biology* 64, pp. 124–138.

- Bannister, Andrew J and Tony Kouzarides (2011). “Regulation of chromatin by histone modifications”. In: *Cell research* 21.3, pp. 381–395.
- Bartholomeeusen, Koen, Yanhui Xiang, Koh Fujinaga, and B Matija Peterlin (2012). “Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein”. In: *Journal of Biological Chemistry* 287.43, pp. 36609–36616.
- Bartman, Caroline R, Nicole Hamagami, Cheryl A Keller, Belinda Giardine, Ross C Hardison, Gerd A Blobel, and Arjun Raj (2019). “Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation”. In: *Molecular cell* 73.3, pp. 519–532.
- Bartman, Caroline R, Sarah C Hsu, Chris C-S Hsiung, Arjun Raj, and Gerd A Blobel (2016). “Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping”. In: *Molecular cell* 62.2, pp. 237–247.
- Battich, Nico, Thomas Stoeger, and Lucas Pelkmans (2013). “Image-based transcriptomics in thousands of single human cells at single-molecule resolution”. In: *Nature methods* 10.11, pp. 1127–1133.
- Bauman, JGJ, J Wiegant, P Borst, and P Van Duijn (1980). “A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA”. In: *Experimental cell research* 128.2, pp. 485–490.
- Beagan, Jonathan A, Michael T Duong, Katelyn R Titus, Linda Zhou, Zhendong Cao, Jingjing Ma, Caroline V Lachanski, Daniel R Gillis, and Jennifer E Phillips-Cremens (2017). “YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment”. In: *Genome research* 27.7, pp. 1139–1152.
- Beagrie, Robert A et al. (2017). “Complex multi-enhancer contacts captured by genome architecture mapping”. In: *Nature* 543.7646, pp. 519–524.
- Becker, Nicole A, Justin P Peters, Troy A Lionberger, and L James Maher III (2013). “Mechanism of promoter repression by Lac repressor–DNA loops”. In: *Nucleic acids research* 41.1, pp. 156–166.
- Beliveau, Brian J et al. (2012). “Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes”. In: *Proceedings of the National Academy of Sciences* 109.52, pp. 21301–21306.

- Bell, Adam C, Adam G West, and Gary Felsenfeld (1999). “The protein CTCF is required for the enhancer blocking activity of vertebrate insulators”. In: *Cell* 98.3, pp. 387–396.
- Bertrand, Edouard, Pascal Chartrand, Matthias Schaefer, Shailesh M. Shenoy, Robert H. Singer, and Roy M. Long (1998). “Localization of ASH1 mRNA Particles in Living Yeast”. In: *Molecular Cell* 2.4, pp. 437–445. ISSN: 1097-2765. DOI: [https://doi.org/10.1016/S1097-2765\(00\)80143-4](https://doi.org/10.1016/S1097-2765(00)80143-4). URL: <http://www.sciencedirect.com/science/article/pii/S1097276500801434>.
- Bhatt, Dev M, Amy Pandya-Jones, Ann-Jay Tong, Iros Barozzi, Michelle M Lissner, Gioacchino Natoli, Douglas L Black, and Stephen T Smale (2012). “Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions”. In: *Cell* 150.2, pp. 279–290.
- Bintu, Bogdan, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang (2018). “Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells”. In: *Science* 362.6413, eaau1783.
- Bonev, Boyan et al. (2017). “Multiscale 3D genome rewiring during mouse neural development”. In: *Cell* 171.3, pp. 557–572.
- Bose, Daniel A, Greg Donahue, Danny Reinberg, Ramin Shiekhata, Roberto Bonasio, and Shelley L Berger (2017). “RNA binding to CBP stimulates histone acetylation and transcription”. In: *Cell* 168.1-2, pp. 135–149.
- Branco, Miguel R and Ana Pombo (2006). “Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations”. In: *PLoS biology* 4.5.
- Buettner, Florian, Kedar N Natarajan, F Paolo Casale, Valentina Proserpio, Antonio Scialdone, Fabian J Theis, Sarah A Teichmann, John C Marioni, and Oliver Stegle (2015). “Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells”. In: *Nature biotechnology* 33.2, pp. 155–160.
- Buheitel, Johannes and Olaf Stemmann (2013). “Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3–Scc1 gate”. In: *The EMBO journal* 32.5, pp. 666–676.
- Bürmann, Frank et al. (2019). “A folded conformation of MukBEF and cohesin”. In: *Nature structural & molecular biology* 26.3, pp. 227–236.

- Cacciatore, Ángela Sedeño and Benjamin D Rowland (2019). “Loop formation by SMC complexes: Turning heads, bending elbows, and fixed anchors”. In: *Current opinion in genetics & development* 55, pp. 11–18.
- Cancer Genome Atlas Research Network (2013). “Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia”. In: *New England Journal of Medicine* 368.22, pp. 2059–2074.
- Cancer Research UK. *AML statistics*. URL: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml> (visited on 03/31/2020).
- Carpenter, Anne E et al. (2006). “CellProfiler: image analysis software for identifying and quantifying cell phenotypes”. In: *Genome biology* 7.10, R100.
- Chan, Thalia E, Michael PH Stumpf, and Ann C Babbie (2017). “Gene regulatory network inference from single-cell data using multivariate information measures”. In: *Cell systems* 5.3, pp. 251–267.
- Chang, Mikyoung, Wei Jin, and Shao-Cong Sun (2009). “Peli1 facilitates TRIF-dependent Toll-like receptor signaling and proinflammatory cytokine production”. In: *Nature immunology* 10.10, p. 1089.
- Chao, Jeffrey A, Yury Patskovsky, Steven C Almo, and Robert H Singer (2008). “Structural basis for the coevolution of a viral RNA–protein complex”. In: *Nature structural & molecular biology* 15.1, pp. 103–105.
- Chao, William CH et al. (2017). “Structure of the cohesin loader Scc2”. In: *Nature communications* 8.1, pp. 1–8.
- Chen, Geng, Baitang Ning, and Tielu Shi (2019a). “Single-cell RNA-seq technologies and related computational data analysis”. In: *Frontiers in genetics* 10, p. 317.
- Chen, Hongtao, Michal Levo, Lev Barinov, Miki Fujioka, James B Jaynes, and Thomas Gregor (2018). “Dynamic interplay between enhancer–promoter topology and gene activity”. In: *Nature genetics* 50.9, pp. 1296–1303.
- Chen, Kok Hao, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang (2015). “Spatially resolved, highly multiplexed RNA profiling in single cells”. In: *Science* 348.6233.
- Chen, Liang-Fu et al. (2019b). “Enhancer histone acetylation modulates transcriptional bursting dynamics of neuronal activity-inducible genes”. In: *Cell reports* 26.5, pp. 1174–1188.

- Chen, Zhiyang et al. (2019c). “Cohesin-mediated NF- κ B signaling limits hematopoietic stem cell self-renewal in aging and inflammation”. In: *Journal of Experimental Medicine* 216.1, pp. 152–175.
- Chong, Shasha, Chongyi Chen, Hao Ge, and X Sunney Xie (2014). “Mechanism of transcriptional bursting in bacteria”. In: *Cell* 158.2, pp. 314–326.
- Chubb, Jonathan R, Tatjana Trcek, Shailesh M Shenoy, and Robert H Singer (2006). “Transcriptional pulsing of a developmental gene”. In: *Current biology* 16.10, pp. 1018–1025.
- Ciosk, Rafal, Masaki Shirayama, Anna Shevchenko, Tomoyuki Tanaka, Attila Toth, Andrej Shevchenko, and Kim Nasmyth (2000). “Cohesin’s binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins”. In: *Molecular cell* 5.2, pp. 243–254.
- Clapier, Cedric R, Janet Iwasa, Bradley R Cairns, and Craig L Peterson (2017). “Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes”. In: *Nature reviews Molecular cell biology* 18.7, pp. 407–422.
- Cook, Daniel L, Anthony N Gerber, and Stephen J Tapscott (1998). “Modeling stochastic gene expression: implications for haploinsufficiency”. In: *Proceedings of the National Academy of Sciences* 95.26, pp. 15641–15646.
- Corces-Zimmerman, M Ryan, Wan-Jen Hong, Irving L Weissman, Bruno C Medeiros, and Ravindra Majeti (2014). “Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission”. In: *Proceedings of the National Academy of Sciences* 111.7, pp. 2548–2553.
- Core, Leighton and Karen Adelman (2019). “Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation”. In: *Genes & development* 33.15-16, pp. 960–982.
- Cremer, Thomas and Marion Cremer (2010). “Chromosome territories”. In: *Cold Spring Harbor perspectives in biology* 2.3, a003889.
- Creyghton, Menno P et al. (2010). “Histone H3K27ac separates active from poised enhancers and predicts developmental state”. In: *Proceedings of the National Academy of Sciences* 107.50, pp. 21931–21936.
- Cuartero, Sergi, Andrew J Innes, and Matthias Merkenschlager (2019). “Towards a better understanding of cohesin mutations in AML”. In: *Frontiers in Oncology* 9, p. 867.
- Cuartero, Sergi et al. (2018). “Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation”. In: *Nature immunology* 19.9, p. 932.

- Davidson, Iain F, Benedikt Bauer, Daniela Goetz, Wen Tang, Gordana Wutz, and Jan-Michael Peters (2019). “DNA loop extrusion by human cohesin”. In: *Science* 366.6471, pp. 1338–1345.
- Davis, Beckley K. (2019). “Derivation of Macrophages from Mouse Bone Marrow”. In: *Mouse Models of Innate Immunity: Methods and Protocols*. Ed. by Irving C. Allen. New York, NY: Springer New York, pp. 41–55. ISBN: 978-1-4939-9167-9. DOI: 10.1007/978-1-4939-9167-9_3. URL: https://doi.org/10.1007/978-1-4939-9167-9_3.
- Day, Daniel S, Bing Zhang, Sean M Stevens, Francesco Ferrari, Erica N Larschan, Peter J Park, and William T Pu (2016). “Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types”. In: *Genome biology* 17.1, p. 120.
- De Santa, Francesca et al. (2010). “A large fraction of extragenic RNA pol II transcription sites overlap enhancers”. In: *PLoS Biol* 8.5, e1000384.
- Deardorff, Matthew A et al. (2012). “HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle”. In: *Nature* 489.7415, pp. 313–317.
- Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner (2002). “Capturing chromosome conformation”. In: *science* 295.5558, pp. 1306–1311.
- DeMare, Laura E., Jing Leng, Justin Cotney, Steven K. Reilly, Jun Yin, Richard Sarro, and James P. Noonan (2013). “The genomic landscape of cohesin-Associated chromatin interactions”. In: *Genome Research*. ISSN: 10889051. DOI: 10.1101/gr.156570.113.
- Deng, Wulan, Jongjoo Lee, Hongxin Wang, Jeff Miller, Andreas Reik, Philip D Gregory, Ann Dean, and Gerd A Blobel (2012). “Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor”. In: *Cell* 149.6, pp. 1233–1244.
- Deonarain, Raj, Antonio Alcamí, Maria Alexiou, Margaret J Dallman, Dirk R Gewert, and Andrew CG Porter (2000). “Impaired antiviral response and alpha/beta interferon induction in mice lacking beta interferon”. In: *Journal of virology* 74.7, pp. 3404–3409.
- Despang, Alexandra et al. (2019). “Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture”. In: *Nature genetics* 51.8, pp. 1263–1271.
- Diebold-Durand, Marie-Laure et al. (2017). “Structure of full-length SMC and rearrangements required for chromosome organization”. In: *Molecular cell* 67.2, pp. 334–347.
- Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren (2012). “Topological domains in mammalian genomes identified by analysis of chromatin interactions”. In: *Nature*. ISSN: 00280836. DOI: 10.1038/nature11082.

- Dixon, Jesse R. et al. (2015). “Chromatin architecture reorganization during stem cell differentiation”. In: *Nature*. ISSN: 14764687. DOI: 10.1038/nature14222.
- Dolgosheina, Elena V, Sunny CY Jeng, Shanker Shyam S Panchapakesan, Razvan Cojocaru, Patrick SK Chen, Peter D Wilson, Nancy Hawkins, Paul A Wiggins, and Peter J Unrau (2014). “RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking”. In: *ACS chemical biology* 9.10, pp. 2412–2420.
- Donovan, Benjamin T, Anh Huynh, David A Ball, Heta P Patel, Michael G Poirier, Daniel R Larson, Matthew L Ferguson, and Tineke L Lenstra (2019). “Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting”. In: *The EMBO journal* 38.12.
- Dorschner, Michael O et al. (2004). “High-throughput localization of functional elements by quantitative chromatin profiling”. In: *Nature methods* 1.3, pp. 219–225.
- Dostie, Josée et al. (2006). “Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements”. In: *Genome research* 16.10, pp. 1299–1309.
- Downen, Jill M et al. (2014). “Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes”. In: *Cell* 159.2, pp. 374–387.
- El Khattabi, Laila et al. (2019). “A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers”. In: *Cell* 178.5, pp. 1145–1158.
- Eling, Nils, Arianne C Richard, Sylvia Richardson, John C Marioni, and Catalina A Vallejos (2018). “Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data”. In: *Cell systems* 7.3, pp. 284–294.
- Falo-Sanjuan, Julia, Nicholas C Lammers, Hernan G Garcia, and Sarah J Bray (2019). “Enhancer priming enables fast and sustained transcriptional responses to Notch signaling”. In: *Developmental cell* 50.4, pp. 411–425.
- Fang, Rongxin, Miao Yu, Guoqiang Li, Sora Chee, Tristin Liu, Anthony D Schmitt, and Bing Ren (2016). “Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq”. In: *Cell research* 26.12, pp. 1345–1348.
- Femino, Andrea M, Fredric S Fay, Kevin Fogarty, and Robert H Singer (1998). “Visualization of Single RNA Transcripts in Situ”. In: *Science* 280.5363, pp. 585–590. ISSN: 0036-8075. DOI: 10.1126/science.280.5363.585. URL: <https://science.sciencemag.org/content/280/5363/585>.

- Fisher, Joseph B et al. (2017). “The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9”. In: *Leukemia* 31.3, pp. 712–719.
- Fitzgerald, Katherine A and Jonathan C Kagan (2020). “Toll-like receptors and the control of immunity”. In: *Cell*.
- Flyamer, Ilya M, Johanna Gassler, Maxim Imakaev, Hugo B Brandão, Sergey V Uljanov, Nezar Abdennur, Sergey V Razin, Leonid A Mirny, and Kikue Tachibana-Konwalski (2017). “Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition”. In: *Nature* 544.7648, pp. 110–114.
- Franke, Martin et al. (2016). “Formation of new chromatin domains determines pathogenicity of genomic duplications”. In: *Nature* 538.7624, pp. 265–269.
- Fritsch, Christoph, Stephan Baumgärtner, Monika Kuban, Daria Steinshorn, George Reid, and Stefan Legewie (2018). “Estrogen-dependent control and cell-to-cell variability of transcriptional bursting”. In: *Molecular systems biology* 14.2, e7678.
- Fu, Yi, Pedro P Rocha, Vincent M Luo, Ramya Raviram, Yan Deng, Esteban O Mazzoni, and Jane A Skok (2016). “CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci”. In: *Nature communications* 7.1, pp. 1–8.
- Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny (2016). “Formation of chromosomal domains by loop extrusion”. In: *Cell reports* 15.9, pp. 2038–2049.
- Fukaya, Takashi, Bomyi Lim, and Michael Levine (2016). “Enhancer control of transcriptional bursting”. In: *Cell* 166.2, pp. 358–368.
- Fullwood, Melissa J et al. (2009). “An oestrogen-receptor- α -bound human chromatin interactome”. In: *Nature* 462.7269, pp. 58–64.
- Galeev, Roman et al. (2016). “Genome-wide RNAi screen identifies cohesin genes as modifiers of renewal and differentiation in human HSCs”. In: *Cell Reports* 14.12, pp. 2988–3000.
- Gall, Joseph G (2016). “The origin of in situ hybridization—a personal history”. In: *Methods* 98, pp. 4–9.
- Gall, Joseph G. and Mary Lou Pardue (1969). “FORMATION AND DETECTION OF RNA-DNA HYBRID MOLECULES IN CYTOLOGICAL PREPARATIONS”. In: *Proceedings of the National Academy of Sciences* 63.2, pp. 378–383. ISSN: 0027-8424. DOI: 10.1073/pnas.

- 63.2.378. eprint: <https://www.pnas.org/content/63/2/378.full.pdf>. URL: <https://www.pnas.org/content/63/2/378>.
- Ganji, Mahipal, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker (2018). “Real-time imaging of DNA loop extrusion by condensin”. In: *Science* 360.6384, pp. 102–105.
- Gassler, Johanna, Hugo B Brandão, Maxim Imakaev, Ilya M Flyamer, Sabrina Ladstätter, Wendy A Bickmore, Jan-Michael Peters, Leonid A Mirny, and Kikuë Tachibana (2017). “A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture”. In: *The EMBO journal* 36.24, pp. 3600–3618.
- Ghavi-Helm, Yad, Aleksander Jankowski, Sascha Meiers, Rebecca R Viales, Jan O Korbel, and Eileen EM Furlong (2019). “Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression”. In: *Nature genetics* 51.8, pp. 1272–1282.
- Gioannini, Theresa L, Athmane Teghanemt, DeSheng Zhang, Nathan P Coussens, Wendie Dockstader, S Ramaswamy, and Jerrold P Weiss (2004). “Isolation of an endotoxin–MD-2 complex that produces Toll-like receptor 4-dependent cell activation at picomolar concentrations”. In: *Proceedings of the National Academy of Sciences* 101.12, pp. 4186–4191.
- Giorgio, Elisa et al. (2015). “A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD)”. In: *Human molecular genetics* 24.11, pp. 3143–3154.
- Golding, Ido, Johan Paulsson, Scott M Zawilski, and Edward C Cox (2005). “Real-time kinetics of gene activity in individual bacteria”. In: *Cell* 123.6, pp. 1025–1036.
- Grosberg, A, Y Rabin, S Havlin, and A Neer (1993). “Crumpled globule model of the three-dimensional structure of DNA”. In: *EPL (Europhysics Letters)* 23.5, p. 373.
- Gruber, Stephan, Christian H Haering, and Kim Nasmyth (2003). “Chromosomal cohesin forms a ring”. In: *Cell* 112.6, pp. 765–777.
- Guo, Ya et al. (2015). “CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function”. In: *Cell*. ISSN: 10974172. DOI: 10.1016/j.cell.2015.07.038.
- Gupta, Kapil, Duygu Sari-Ak, Matthias Haffke, Simon Trowitzsch, and Imre Berger (2016). “Zooming in on transcription preinitiation”. In: *Journal of molecular biology* 428.12, pp. 2581–2591.
- Haarhuis, Judith HI et al. (2017). “The cohesin release factor WAPL restricts chromatin loop extension”. In: *Cell* 169.4, pp. 693–707.

- Haas, Simon, Andreas Trumpp, and Michael D Milsom (2018). “Causes and consequences of hematopoietic stem cell heterogeneity”. In: *Cell Stem Cell* 22.5, pp. 627–638.
- Halpern, Keren Bahar, Sivan Tanami, Shanie Landen, Michal Chapal, Liran Szlak, Anat Hutzler, Anna Nizhberg, and Shalev Itzkovitz (2015). “Bursty gene expression in the intact mammalian liver”. In: *Molecular cell* 58.1, pp. 147–156.
- Hansen, Anders S, Iryna Pustova, Claudia Cattoglio, Robert Tjian, and Xavier Darzacq (2017). “CTCF and cohesin regulate chromatin loop stability with distinct dynamics”. In: *Elife* 6, e25776.
- Hanssen, Lars LP et al. (2017). “Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo”. In: *Nature cell biology* 19.8, pp. 952–961.
- Hark, Amy T, Christopher J Schoenherr, David J Katz, Robert S Ingram, John M Leverage, and Shirley M Tilghman (2000). “CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus”. In: *Nature* 405.6785, pp. 486–489.
- Hassler, Markus, Indra A Shaltiel, and Christian H Haering (2018). “Towards a unified model of SMC complex function”. In: *Current Biology* 28.21, R1266–R1281.
- Hastie, Nicholas D (2017). “Wilms’ tumour 1 (WT1) in development, homeostasis and disease”. In: *Development* 144.16, pp. 2862–2872.
- Heidinger-Pauli, Jill M, Ozlem Mert, Carol Davenport, Vincent Guacci, and Douglas Koshland (2010). “Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair”. In: *Current Biology* 20.10, pp. 957–963.
- Heintzman, Nathaniel D et al. (2007). “Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome”. In: *Nature genetics* 39.3, pp. 311–318.
- Heinz, Sven et al. (2010). “Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities”. In: *Molecular cell* 38.4, pp. 576–589.
- Herschman, Harvey R (1991). “Primary response genes induced by growth factors and tumor promoters”. In: *Annual review of biochemistry* 60.1, pp. 281–319.
- Hnisz, Denes, Brian J Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A Sigova, Heather A Hoke, and Richard A Young (2013). “Super-enhancers in the control of cell identity and disease”. In: *Cell* 155.4, pp. 934–947.

- Hnisz, Denes et al. (2016). “Activation of proto-oncogenes by disruption of chromosome neighborhoods”. In: *Science* 351.6280, pp. 1454–1458.
- Hsieh, Chyi-Song, Steven E Macatonia, Catherine S Tripp, Stanley F Wolf, Anne O’Garra, and Kenneth M Murphy (1993). “Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages”. In: *Science* 260.5107, pp. 547–549.
- Hwang, Byungjin, Ji Hyun Lee, and Duhee Bang (2018). “Single-cell RNA sequencing technologies and bioinformatics pipelines”. In: *Experimental & molecular medicine* 50.8, pp. 1–14.
- Ing-Simmons, Elizabeth, Vlad C Seitan, Andre J Faure, Paul Flicek, Thomas Carroll, Job Dekker, Amanda G Fisher, Boris Lenhard, and Matthias Merkenschlager (2015). “Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin”. In: *Genome research* 25.4, pp. 504–513.
- Ivashkiv, Lionel B and Laura T Donlin (2014). “Regulation of type I interferon responses”. In: *Nature reviews Immunology* 14.1, pp. 36–49.
- Jackson, David A, Robert H Symons, and Paul Berg (1972). “Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli”. In: *Proceedings of the National Academy of Sciences* 69.10, pp. 2904–2909.
- Jaeger, Martin G et al. (2020). “Selective Mediator dependence of cell-type-specifying transcription”. In: *Nature Genetics*, pp. 1–9.
- Jaitin, Diego Adhemar et al. (2014). “Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types”. In: *Science* 343.6172, pp. 776–779.
- Kaaij, Lucas JT, Robin H van der Weide, Rene F Ketting, and Elzo de Wit (2018). “Systemic loss and gain of chromatin architecture throughout zebrafish development”. In: *Cell reports* 24.1, pp. 1–10.
- Kadonaga, James T (2012). “Perspectives on the RNA polymerase II core promoter”. In: *Wiley Interdisciplinary Reviews: Developmental Biology* 1.1, pp. 40–51.
- Kagan, Jonathan C, Venkat Giri Magupalli, and Hao Wu (2014). “SMOCs: supramolecular organizing centres that control innate immunity”. In: *Nature Reviews Immunology* 14.12, pp. 821–826.
- Kagey, Michael H et al. (2010). “Mediator and cohesin connect gene expression and chromatin architecture”. In: *Nature* 467.7314, pp. 430–435.

- Kaikkonen, Minna U et al. (2013). “Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription”. In: *Molecular cell* 51.3, pp. 310–325.
- Kanehisa, Minoru and Susumu Goto (2000). “KEGG: kyoto encyclopedia of genes and genomes”. In: *Nucleic acids research* 28.1, pp. 27–30.
- Kempfer, Rieke and Ana Pombo (2019). “Methods for mapping 3D chromosome architecture”. In: *Nature Reviews Genetics*, pp. 1–20.
- Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg (2013). “TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions”. In: *Genome biology* 14.4, R36.
- Kim, Somi, Nam-Kyung Yu, and Bong-Kiun Kaang (2015). “CTCF as a multifunctional protein in genome regulation and gene expression”. In: *Experimental & molecular medicine* 47.6, e166–e166.
- Kim, Tae-Kyung et al. (2010). “Widespread transcription at neuronal activity-regulated enhancers”. In: *Nature* 465.7295, pp. 182–187.
- Kim, Yoori, Zhubing Shi, Hongshan Zhang, Ilya J Finkelstein, and Hongtao Yu (2019). “Human cohesin compacts DNA by loop extrusion”. In: *Science* 366.6471, pp. 1345–1349.
- Koch, Frederic et al. (2011). “Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters”. In: *Nature structural & molecular biology* 18.8, p. 956.
- Kojic, Aleksandar, Ana Cuadrado, Magali De Koninck, Daniel Giménez-Llorente, Miriam Rodriguez-Corsino, Gonzalo Gómez-López, François Le Dily, Marc A Marti-Renom, and Ana Losada (2018). “Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization”. In: *Nature structural & molecular biology* 25.6, pp. 496–504.
- Kon, Ayana et al. (2013). “Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms”. In: *Nature genetics* 45.10, pp. 1232–1237.
- Kouzine, Fedor et al. (2013). “Global regulation of promoter melting in naive lymphocytes”. In: *Cell* 153.5, pp. 988–999.
- Kovtonyuk, Larisa V, Kristin Fritsch, Xiaomin Feng, Markus G Manz, and Hitoshi Takizawa (2016). “Inflamm-aging of hematopoiesis, hematopoietic stem cells, and the bone marrow microenvironment”. In: *Frontiers in immunology* 7, p. 502.
- Kowalczyk, Monika S et al. (2012). “Intragenic enhancers act as alternative promoters”. In: *Molecular cell* 45.4, pp. 447–458.

- La Manno, Gioele et al. (2018). “RNA velocity of single cells”. In: *Nature* 560.7719, pp. 494–498.
- Lai, Fan, Ulf A Orom, Matteo Cesarini, Malte Beringer, Dylan J Taatjes, Gerd A Blobel, and Ramin Shiekhattar (2013). “Activating RNAs associate with Mediator to enhance chromatin architecture and transcription”. In: *Nature* 494.7438, pp. 497–501.
- Langer-Safer, Pennina R, Michael Levine, and David C Ward (1982). “Immunological method for mapping genes on Drosophila polytene chromosomes”. In: *Proceedings of the National Academy of Sciences* 79.14, pp. 4381–4385.
- Larson, Daniel R, Christoph Fritzsche, Liang Sun, Xiuhai Meng, David S Lawrence, and Robert H Singer (2013). “Direct observation of frequency modulated transcription in single cells using light activation”. In: *Elife* 2, e00750.
- Larsson, Anton JM et al. (2019). “Genomic encoding of transcriptional burst kinetics”. In: *Nature* 565.7738, pp. 251–254.
- Law, Yun Zhi, Jean-Daniel Bancal, Valerio Scarani, et al. (2014). “Quantum randomness extraction for various levels of characterization of the devices”. In: *Journal of Physics A: Mathematical and Theoretical* 47.42, p. 424028.
- Lawrence, Moyra, Sylvain Daujat, and Robert Schneider (2016). “Lateral thinking: how histone modifications regulate gene expression”. In: *Trends in Genetics* 32.1, pp. 42–56.
- Lazar, Nathan H, Kimberly A Nevonen, Brendan O’Connell, Christine McCann, Rachel J O’Neill, Richard E Green, Thomas J Meyer, Mariam Okhovat, and Lucia Carbone (2018). “Epigenetic maintenance of topological domains in the highly rearranged gibbon genome”. In: *Genome research* 28.7, pp. 983–997.
- Lee, ChangHwan, Heaji Shin, and Judith Kimble (2019). “Dynamics of Notch-dependent transcriptional bursting in its native context”. In: *Developmental cell* 50.4, pp. 426–435.
- Lee, Tong Ihn and Richard A Young (2013). “Transcriptional regulation and its misregulation in disease”. In: *Cell* 152.6, pp. 1237–1251.
- Lenive, Oleg, Paul DW Kirk, and Michael PH Stumpf (2016). “Inferring extrinsic noise from single-cell gene expression data using approximate Bayesian computation”. In: *BMC systems biology* 10.1, pp. 1–17.
- Li, Congxin, François Cesbron, Michael Oehler, Michael Brunner, and Thomas Höfer (2018). “Frequency modulation of transcriptional bursting enables sensitive and rapid gene regulation”. In: *Cell systems* 6.4, pp. 409–423.

- Li, Wenbo et al. (2013). “Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation”. In: *Nature* 498.7455, pp. 516–520.
- Li, Yan et al. (2020). “The structural basis for cohesin–CTCF-anchored loops”. In: *Nature* 578.7795, pp. 472–476.
- Liao, Yang, Gordon K Smyth, and Wei Shi (2013). “The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote”. In: *Nucleic acids research* 41.10, e108–e108.
- (2014). “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features”. In: *Bioinformatics* 30.7, pp. 923–930.
- Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo (2015). “The molecular signatures database hallmark gene set collection”. In: *Cell systems* 1.6, pp. 417–425.
- Lieberman-Aiden, Erez et al. (2009). “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome”. In: *Science (New York, N.Y.)* ISSN: 0036-8075. DOI: 10.1126/science.1181369.
- Liu, Miao, Yanfang Lu, Bing Yang, Yanbo Chen, Jonathan SD Radda, Mengwei Hu, Samuel G Katz, and Siyuan Wang (2020). “Multiplexed imaging of nucleome architectures in single cells of mammalian tissue”. In: *Nature Communications* 11.1, pp. 1–14.
- Liu, Mingli, Shanchun Guo, and Jonathan K Stiles (2011). “The emerging role of CXCL10 in cancer”. In: *Oncology letters* 2.4, pp. 583–589.
- Love, Michael I, Wolfgang Huber, and Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome biology* 15.12, p. 550.
- Lucas, Joseph S, Yaojun Zhang, Olga K Dudko, and Cornelis Murre (2014). “3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions”. In: *Cell* 158.2, pp. 339–352.
- Lupiáñez, Darío G et al. (2015). “Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions”. In: *Cell* 161.5, pp. 1012–1025.
- Luppino, Jennifer M, Daniel S Park, Son C Nguyen, Yemin Lan, Zhuxuan Xu, Rebecca Yunker, and Eric F Joyce (2020). “Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes”. In: *Nature Genetics*, pp. 1–9.
- Marco-Sola, Santiago, Michael Sammeth, Roderic Guigó, and Paolo Ribeca (2012). “The GEM mapper: fast, accurate and versatile alignment by filtration”. In: *Nature methods* 9.12, p. 1185.

- Martinez-Jimenez, Celia Pilar et al. (2017). “Aging increases cell-to-cell transcriptional variability upon immune stimulation”. In: *Science* 355.6332, pp. 1433–1436.
- Mateo, Leslie J, Sedona E Murphy, Antonina Hafner, Isaac S Cinquini, Carly A Walker, and Alistair N Boettiger (2019). “Visualizing DNA folding and RNA in embryos at single-cell resolution”. In: *Nature* 568.7750, pp. 49–54.
- McCarthy, Davis J, Yunshun Chen, and Gordon K Smyth (2012). “Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation”. In: *Nucleic acids research* 40.10, pp. 4288–4297.
- McGeachy, Mandy J, Yi Chen, Cristina M Tato, Arian Laurence, Barbara Joyce-Shaikh, Wendy M Blumenschein, Terrill K McClanahan, John J O’shea, and Daniel J Cua (2009). “The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17–producing effector T helper cells in vivo”. In: *Nature immunology* 10.3, p. 314.
- Melgar, Michael F, Francis S Collins, and Praveen Sethupathy (2011). “Discovery of active enhancers through bidirectional expression of short transcripts”. In: *Genome biology* 12.11, R113.
- Michaelis, Christine, Rafal Ciosk, and Kim Nasmyth (1997). “Cohesins: chromosomal proteins that prevent premature separation of sister chromatids”. In: *Cell* 91.1, pp. 35–45.
- Mikhaylichenko, Olga, Vladyslav Bondarenko, Dermot Harnett, Ignacio E Schor, Matilda Males, Rebecca R Viales, and Eileen EM Furlong (2018). “The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription”. In: *Genes & development* 32.1, pp. 42–57.
- Molina, Nacho, David M Suter, Rosamaria Cannavo, Benjamin Zoller, Ivana Gotic, and Félix Naef (2013). “Stimulus-induced modulation of transcriptional bursting in a single mammalian gene”. In: *Proceedings of the National Academy of Sciences* 110.51, pp. 20563–20568.
- Mootha, Vamsi K et al. (2003). “Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria”. In: *Cell* 115.5, pp. 629–640.
- Morrish, Fionnuala and David Hockenbery (2014). “MYC and mitochondrial biogenesis”. In: *Cold Spring Harbor perspectives in medicine* 4.5, a014225.
- Mueller, Florian et al. (2013). “FISH-quant: automatic counting of transcripts in 3D FISH images”. In: *Nature methods* 10.4, pp. 277–278.

- Mullenders, Jasper et al. (2015). “Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms”. In: *Journal of Experimental Medicine* 212.11, pp. 1833–1850.
- Müller, Stefan, Michaela Neusser, and Johannes Wienberg (2002). “Towards unlimited colors for fluorescence in-situ hybridization (FISH)”. In: *Chromosome Research* 10.3, pp. 223–232.
- Mumbach, Maxwell R, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf, and Howard Y Chang (2016). “HiChIP: efficient and sensitive analysis of protein-directed genome architecture”. In: *Nature methods* 13.11, pp. 919–922.
- Mumbach, Maxwell R et al. (2019). “HiChIRP reveals RNA-associated chromosome conformation”. In: *Nature methods* 16.6, pp. 489–492.
- Munsky, Brian, Gregor Neuert, and Alexander Van Oudenaarden (2012). “Using gene expression noise to understand gene regulation”. In: *Science* 336.6078, pp. 183–187.
- Nagano, Takashi, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D Laue, Amos Tanay, and Peter Fraser (2013). “Single-cell Hi-C reveals cell-to-cell variability in chromosome structure”. In: *Nature* 502.7469, pp. 59–64.
- Nagano, Takashi, Yaniv Lubling, Eitan Yaffe, Steven W Wingett, Wendy Dean, Amos Tanay, and Peter Fraser (2015). “Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell”. In: *Nature protocols* 10.12, p. 1986.
- Nasmyth, Kim (2001). “Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis”. In: *Annual Review of Genetics*. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.35.102401.091334.
- Nicolas, Damien, Benjamin Zoller, David M Suter, and Felix Naef (2018). “Modulation of transcriptional burst frequency by histone acetylation”. In: *Proceedings of the National Academy of Sciences* 115.27, pp. 7153–7158.
- Nora, Elphege P., Anton Goloborodko, Anne Laure Valton, Johan H. Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A. Mirny, and Benoit G. Bruneau (2017). “Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization”. In: *Cell*. ISSN: 10974172. DOI: 10.1016/j.cell.2017.05.004.
- Nora, Elphège P et al. (2012). “Spatial partitioning of the regulatory landscape of the X-inactivation centre”. In: *Nature* 485.7398, pp. 381–385.

- Nuebler, Johannes, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A Mirny (2018). “Chromatin organization by an interplay of loop extrusion and compartmental segregation”. In: *Proceedings of the National Academy of Sciences* 115.29, E6697–E6706.
- Ochiai, Hiroshi et al. (2019). “Genome-wide analysis of transcriptional bursting-induced noise in mammalian cells”. In: *bioRxiv*, p. 736207.
- Oppmann, Birgit et al. (2000). “Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12”. In: *Immunity* 13.5, pp. 715–725.
- Osterwalder, Marco et al. (2018). “Enhancer redundancy provides phenotypic robustness in mammalian development”. In: *Nature* 554.7691, pp. 239–243.
- Ostuni, Renato et al. (2013). “Latent enhancers activated by stimulation in differentiated cells”. In: *Cell* 152.1-2, pp. 157–171.
- Paige, Jeremy S, Karen Y Wu, and Samie R Jaffrey (2011). “RNA mimics of green fluorescent protein”. In: *Science* 333.6042, pp. 642–646.
- Pardue, Mary Lou and Joseph G. Gall (1969). “MOLECULAR HYBRIDIZATION OF RADIOACTIVE DNA TO THE DNA OF CYTOLOGICAL PREPARATIONS”. In: *Proceedings of the National Academy of Sciences* 64.2, pp. 600–604. ISSN: 0027-8424. DOI: 10.1073/pnas.64.2.600. eprint: <https://www.pnas.org/content/64/2/600.full.pdf>. URL: <https://www.pnas.org/content/64/2/600>.
- Pardue, Mary Lou and Joseph G Gall (1970). “Chromosomal localization of mouse satellite DNA”. In: *Science* 168.3937, pp. 1356–1358.
- Parelho, Vania et al. (2008). “Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms”. In: *Cell*. ISSN: 00928674. DOI: 10.1016/j.cell.2008.01.011.
- Park, Beom Seok, Dong Hyun Song, Ho Min Kim, Byong-Seok Choi, Hayyoung Lee, and Jie-Oh Lee (2009). “The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex”. In: *Nature* 458.7242, pp. 1191–1195.
- Peccoud, Jean and Bernard Ycart (1995). “Markovian modeling of gene-product synthesis”. In: *Theoretical population biology* 48.2, pp. 222–234.
- Peinado, Héctor et al. (2012). “Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET”. In: *Nature medicine* 18.6, pp. 883–891.
- Pennacchio, Len A, Wendy Bickmore, Ann Dean, Marcelo A Nobrega, and Gill Bejerano (2013). “Enhancers: five essential questions”. In: *Nature Reviews Genetics* 14.4, pp. 288–295.

- Pennacchio, Len A et al. (2006). “In vivo enhancer analysis of human conserved non-coding sequences”. In: *Nature* 444.7118, pp. 499–502.
- Perry, Michael W, Alistair N Boettiger, and Michael Levine (2011). “Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo”. In: *Proceedings of the National Academy of Sciences* 108.33, pp. 13570–13575.
- Phipson, Belinda, Stanley Lee, Ian J Majewski, Warren S Alexander, and Gordon K Smyth (2016). “Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression”. In: *The annals of applied statistics* 10.2, p. 946.
- Pietras, Eric M (2017). “Inflammation: a key regulator of hematopoietic stem cell fate in health and disease”. In: *Blood, The Journal of the American Society of Hematology* 130.15, pp. 1693–1698.
- Poltorak, Alexander et al. (1998). “Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene”. In: *Science* 282.5396, pp. 2085–2088.
- Quinodoz, Sofia A et al. (2018). “Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus”. In: *Cell* 174.3, pp. 744–757.
- Rahl, Peter B, Charles Y Lin, Amy C Seila, Ryan A Flynn, Scott McCuine, Christopher B Burge, Phillip A Sharp, and Richard A Young (2010). “c-Myc regulates transcriptional pause release”. In: *Cell* 141.3, pp. 432–445.
- Rahnamoun, Homa, Jihoon Lee, Zhengxi Sun, Hanbin Lu, Kristen M Ramsey, Elizabeth A Komives, and Shannon M Lauberth (2018). “RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation”. In: *Nature structural & molecular biology* 25.8, pp. 687–697.
- Raj, Arjun, Charles S. Peskin, Daniel Tranchina, Diana Y. Vargas, and Sanjay Tyagi (2006a). “Stochastic mRNA synthesis in mammalian cells”. In: *PLoS Biology*. ISSN: 15457885. DOI: 10.1371/journal.pbio.0040309.
- Raj, Arjun, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi (2006b). “Stochastic mRNA synthesis in mammalian cells”. In: *PLoS biology* 4.10, e309.
- Raj, Arjun and Alexander Van Oudenaarden (2008). “Nature, nurture, or chance: stochastic gene expression and its consequences”. In: *Cell* 135.2, pp. 216–226.
- Raj, Arjun et al. (2008). “Nature, nurture, or chance: stochastic gene expression and its consequences.” In: *Cell*. ISSN: 1097-4172. DOI: 10.1016/j.cell.2008.09.050.

- Rand, Ulfert et al. (2012). “Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response”. In: *Molecular systems biology* 8.1, p. 584.
- Rao, Suhas S. P. et al. (2014). “A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping”. In: *Cell*. ISSN: 10974172. DOI: 10.1016/j.cell.2014.11.021.
- Rao, Suhas S. P. et al. (2017). “Cohesin loss eliminates all loop domains”. In: *Cell* 171.2, pp. 305–320.
- Risso, Davide, John Ngai, Terence P Speed, and Sandrine Dudoit (2014). “Normalization of RNA-seq data using factor analysis of control genes or samples”. In: *Nature biotechnology* 32.9, pp. 896–902.
- Ritchie, Matthew E, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies”. In: *Nucleic acids research* 43.7, e47–e47.
- Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth (2010). “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26.1, pp. 139–140.
- Robinson, Mark D and Alicia Oshlack (2010). “A scaling normalization method for differential expression analysis of RNA-seq data”. In: *Genome biology* 11.3, R25.
- Rooney, John P, Ian T Ryde, Laurie H Sanders, Evan H Howlett, Meryl D Colton, Kaylyn E Germ, Greg D Mayer, J Timothy Greenamyre, and Joel N Meyer (2015). “PCR based determination of mitochondrial DNA copy number in multiple species”. In: pp. 23–38.
- Rudan, Matteo Vietri, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T Odom, Amos Tanay, and Suzana Hadjur (2015). “Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture”. In: *Cell reports* 10.8, pp. 1297–1309.
- Rudkin, George T and BD Stollar (1977). “High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence”. In: *Nature* 265.5593, pp. 472–473.
- Sainsbury, Sarah, Carrie Bernecky, and Patrick Cramer (2015). “Structural basis of transcription initiation by RNA polymerase II”. In: *Nature reviews Molecular cell biology* 16.3, pp. 129–143.
- Salmon, Jessica et al. (2020). “Epigenetic reprogramming of plasmacytoid dendritic cells drives type I interferon-dependent differentiation of acute myeloid leukemias for therapeutic bene-

- fit”. In: *bioRxiv*. DOI: 10.1101/2020.08.23.235499. eprint: <https://www.biorxiv.org/content/early/2020/08/24/2020.08.23.235499.full.pdf>. URL: <https://www.biorxiv.org/content/early/2020/08/24/2020.08.23.235499>.
- Sanborn, Adrian L. et al. (2015a). “Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes”. In: *Proceedings of the National Academy of Sciences*. ISSN: 0027-8424. DOI: 10.1073/pnas.1518552112.
- Sanborn, Adrian L et al. (2015b). “Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes”. In: *Proceedings of the National Academy of Sciences* 112.47, E6456–E6465.
- Santamaría, Rodrigo and Philippe Pierre (2012). “Voronto: mapper for expression data to ontologies”. In: *Bioinformatics* 28.17, pp. 2281–2282.
- Sartorelli, Vittorio and Shannon M Lauberth (2020). “Enhancer RNAs are an important regulatory layer of the epigenome”. In: *Nature Structural & Molecular Biology*, pp. 1–8.
- Schaaf, Cheri A, Hojoong Kwak, Amanda Koenig, Ziva Misulovin, David W Gohara, Audrey Watson, Yanjiao Zhou, John T Lis, and Dale Dorsett (2013). “Genome-wide control of RNA polymerase II activity by cohesin”. In: *PLoS genetics* 9.3, e1003382.
- Schardin, Margit, Thomas Cremer, HD Hager, and M Lang (1985). “Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories”. In: *Human genetics* 71.4, pp. 281–287.
- Schaukowitch, Katie, Jae-Yeol Joo, Xihui Liu, Jonathan K Watts, Carlos Martinez, and Tae-Kyung Kim (2014). “Enhancer RNA facilitates NELF release from immediate early genes”. In: *Molecular cell* 56.1, pp. 29–42.
- Schmid, Bianca et al. (Dec. 2016). “Live Cell Analysis and Mathematical Modeling Identify Determinants of Attenuation of Dengue Virus 2’-O-Methylation Mutant”. In: *PLOS Pathogens* 11.12, pp. 1–36. DOI: 10.1371/journal.ppat.1005345. URL: <https://doi.org/10.1371/journal.ppat.1005345>.
- Schneider, William M, Meike Dittmann Chevillotte, and Charles M Rice (2014). “Interferon-stimulated genes: a complex web of host defenses”. In: *Annual review of immunology* 32, pp. 513–545.
- Schoenfelder, Stefan and Peter Fraser (2019). “Long-range enhancer–promoter contacts in gene expression control”. In: *Nature Reviews Genetics*, p. 1.

- Schoenfelder, Stefan et al. (2010). “Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells”. In: *Nature genetics* 42.1, p. 53.
- Schwarzer, Wibke et al. (2017). “Two independent modes of chromatin organization revealed by cohesin removal”. In: *Nature* 551.7678, pp. 51–56.
- Seibler, Jost et al. (2003). “Rapid generation of inducible mouse mutants”. In: *Nucleic acids research* 31.4, e12–e12.
- Seitan, Vlad C et al. (2011). “A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation”. In: *Nature* 476.7361, p. 467.
- Seitan, Vlad C. et al. (2013). “Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments”. In: *Genome Research*. ISSN: 10889051. DOI: 10.1101/gr.161620.113.
- Senecal, Adrien, Brian Munsky, Florence Proux, Nathalie Ly, Floriane E Braye, Christophe Zimmer, Florian Mueller, and Xavier Darzacq (2014). “Transcription factors modulate c-Fos transcriptional bursts”. In: *Cell reports* 8.1, pp. 75–83.
- Setty, Manu et al. (2016). “Wishbone identifies bifurcating developmental trajectories from single-cell data”. In: *Nature biotechnology* 34.6, pp. 637–645.
- Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli (2012). “Three-dimensional folding and functional organization principles of the Drosophila genome”. In: *Cell* 148.3, pp. 458–472.
- Shah, Sheel et al. (2018). “Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH”. In: *Cell* 174.2, pp. 363–376.
- Shalek, Alex K. et al. (2013). “Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells”. In: *Nature* 498.7453, pp. 236–240.
- Shalek, Alex K. et al. (2014). “Single-cell RNA-seq reveals dynamic paracrine control of cellular variation”. In: *Nature*. ISSN: 14764687. DOI: 10.1038/nature13437.
- Shao, Shipeng, Weiwei Zhang, Huan Hu, Boxin Xue, Jinshan Qin, Chaoying Sun, Yuao Sun, Wensheng Wei, and Yujie Sun (2016). “Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system”. In: *Nucleic acids research* 44.9, e86–e86.
- Shen, Li (2014). “GeneOverlap: An R package to test and visualize gene overlaps”. In: *R Package*.

- Shi, Zhubing, Haishan Gao, Xiao-chen Bai, and Hongtao Yu (2020). “Cryo-EM structure of the human cohesin-NIPBL-DNA complex”. In: *Science*.
- Sigova, Alla A et al. (2015). “Transcription factor trapping by RNA in gene regulatory elements”. In: *Science* 350.6263, pp. 978–981.
- Singer, Robert H and David C Ward (1982). “Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog”. In: *Proceedings of the National Academy of Sciences* 79.23, pp. 7331–7335.
- Singh, Abhyudai, Brandon Razooky, Chris D Cox, Michael L Simpson, and Leor S Weinberger (2010). “Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression”. In: *Biophysical journal* 98.8, pp. L32–L34.
- Snijder, Berend and Lucas Pelkmans (2011). “Origins of regulated cell-to-cell variability”. In: *Nature reviews Molecular cell biology* 12.2, pp. 119–125.
- Soutourina, Julie (2018). “Transcription regulation by the Mediator complex”. In: *Nature reviews Molecular cell biology* 19.4, p. 262.
- Spielmann, Malte, Darío G Lupiáñez, and Stefan Mundlos (2018). “Structural variation in the 3D genome”. In: *Nature Reviews Genetics* 19.7, pp. 453–467.
- Srinivasan, Madhusudhan et al. (2018). “The cohesin ring uses its hinge to organize DNA using non-topological as well as topological mechanisms”. In: *Cell* 173.6, pp. 1508–1519.
- Stavreva, Diana A et al. (2019). “Transcriptional Bursting and Co-bursting Regulation by Steroid Hormone Release Pattern and Transcription Factor Mobility”. In: *Molecular cell* 75.6, pp. 1161–1177.
- Stoeger, Thomas, Nico Battich, Markus D Herrmann, Yauhen Yakimovich, and Lucas Pelkmans (2015). “Computer vision for image-based transcriptomics”. In: *Methods* 85, pp. 44–53.
- Stoeger, Thomas, Nico Battich, and Lucas Pelkmans (2016). “Passive noise filtering by cellular compartmentalization”. In: *Cell* 164.6, pp. 1151–1161.
- Su, Jun-Han, Pu Zheng, Seon S Kinrot, Bogdan Bintu, and Xiaowei Zhuang (2020). “Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin”. In: *Cell*.
- Subramanian, Aravind et al. (2005). “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles”. In: *Proceedings of the National Academy of Sciences* 102.43, pp. 15545–15550.
- Sutamewagul, Gerek and Carlos E Vigil (2018). “Clinical use of FLT3 inhibitors in acute myeloid leukemia”. In: *OncoTargets and therapy* 11, p. 7041.

- Suter, David M, Nacho Molina, David Gatfield, Kim Schneider, Ueli Schibler, and Felix Naef (2011). “Mammalian genes are transcribed with widely different bursting kinetics”. In: *Science* 332.6028, pp. 472–474.
- Szabo, Quentin, Frédéric Bantignies, and Giacomo Cavalli (2019). “Principles of genome folding into topologically associating domains”. In: *Science advances* 5.4, eaaw1668.
- Tan, Longzhi, Dong Xing, Chi-Han Chang, Heng Li, and X Sunney Xie (2018). “Three-dimensional genome structures of single diploid human cells”. In: *Science* 361.6405, pp. 924–928.
- Tay, Savaş, Jacob J Hughey, Timothy K Lee, Tomasz Lipniacki, Stephen R Quake, and Markus W Covert (2010). “Single-cell NF- κ B dynamics reveal digital activation and analogue information processing”. In: *Nature* 466.7303, pp. 267–271.
- Thiecke, Michiel J et al. (2020). “Cohesin-dependent and independent mechanisms support chromosomal contacts between promoters and enhancers”. In: *bioRxiv*.
- Thol, Felicitas et al. (2014). “Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications”. In: *Blood, The Journal of the American Society of Hematology* 123.6, pp. 914–920.
- Thomas, Karen E, Carole L Galligan, Raj Deonarain Newman, Eleanor N Fish, and Stefanie N Vogel (2006). “Contribution of interferon- β to the murine macrophage response to the Toll-like receptor 4 agonist, lipopolysaccharide”. In: *Journal of Biological Chemistry* 281.41, pp. 31119–31130.
- Trapnell, Cole et al. (2014). “The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells”. In: *Nature biotechnology* 32.4, p. 381.
- Tsai, Pei-Fang et al. (2018). “A muscle-specific enhancer RNA mediates cohesin recruitment and regulates transcription in trans”. In: *Molecular cell* 71.1, pp. 129–141.
- Tunnacliffe, Edward and Jonathan R Chubb (2020). “What Is a Transcriptional Burst?” In: *Trends in Genetics*.
- Tyner, Jeffrey W et al. (2018). “Functional genomic landscape of acute myeloid leukaemia”. In: *Nature* 562.7728, pp. 526–531.
- Uhlmann, Frank, Dominik Wernic, Marc-André Poupart, Eugene V Koonin, and Kim Nasmyth (2000). “Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast”. In: *Cell* 103.3, pp. 375–386.

- Vallejos, Catalina A, John C Marioni, and Sylvia Richardson (2015). “BASiCS: Bayesian analysis of single-cell sequencing data”. In: *PLoS Comput Biol* 11.6, e1004333.
- Van Der Lelij, Petra et al. (2017). “Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts”. In: *Elife* 6, e26980.
- Varol, Chen, Alexander Mildner, and Steffen Jung (2015). “Macrophages: development and tissue specialization”. In: *Annual review of immunology* 33, pp. 643–675.
- Viny, Aaron D et al. (2015). “Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis”. In: *Journal of Experimental Medicine* 212.11, pp. 1819–1832.
- Viny, Aaron D et al. (2019). “Stag1 and Stag2 regulate cell fate decisions in hematopoiesis through non-redundant topological control”. In: *bioRxiv*, p. 581868.
- Visel, Axel et al. (2009). “ChIP-seq accurately predicts tissue-specific activity of enhancers”. In: *Nature* 457.7231, pp. 854–858.
- Vos, Seychelle M, Lucas Farnung, Henning Urlaub, and Patrick Cramer (2018). “Structure of paused transcription complex Pol II–DSIF–NELF”. In: *Nature* 560.7720, pp. 601–606.
- Walters, Mark C, Steven Fiering, Jeff Eidemiller, Wendy Magis, Mark Groudine, and DI Martin (1995). “Enhancers increase the probability but not the level of gene expression”. In: *Proceedings of the National Academy of Sciences* 92.15, pp. 7125–7129.
- Wang, Hao et al. (2012). “Widespread plasticity in CTCF occupancy linked to DNA methylation”. In: *Genome research* 22.9, pp. 1680–1688.
- Wang, Siyuan, Jun-Han Su, Brian J Beliveau, Bogdan Bintu, Jeffrey R Moffitt, Chao-ting Wu, and Xiaowei Zhuang (2016a). “Spatial organization of chromatin domains and compartments in single chromosomes”. In: *Science* 353.6299, pp. 598–602.
- Wang, Siyuan, Jun-Han Su, Feng Zhang, and Xiaowei Zhuang (2016b). “An RNA-aptamer-based two-color CRISPR labeling system”. In: *Scientific reports* 6.1, pp. 1–7.
- Weintraub, Abraham S et al. (2017). “YY1 is a structural regulator of enhancer-promoter loops”. In: *Cell* 171.7, pp. 1573–1588.
- Wendt, Kerstin S. et al. (2008). “Cohesin mediates transcriptional insulation by CCCTC-binding factor”. In: *Nature*. ISSN: 14764687. DOI: 10.1038/nature06634.
- Whyte, Warren A, David A Orlando, Denes Hnisz, Brian J Abraham, Charles Y Lin, Michael H Kagey, Peter B Rahl, Tong Ihn Lee, and Richard A Young (2013). “Master transcription factors and mediator establish super-enhancers at key cell identity genes”. In: *Cell* 153.2, pp. 307–319.

- Will, Anja J. et al. (2017). “Composition and dosage of a multipartite enhancer cluster control developmental expression of *Ihh* (Indian hedgehog)”. In: *Nature genetics* 49.10, pp. 1539–1545.
- Winter, Georg E et al. (2017). “BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment”. In: *Molecular cell* 67.1, pp. 5–18.
- Wit, Elzo de et al. (2015). “CTCF binding polarity determines chromatin looping”. In: *Molecular cell* 60.4, pp. 676–684.
- Wutz, Gordana et al. (2017). “Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins”. In: *The EMBO journal* 36.24, pp. 3573–3599.
- Xue, Qiong, Yao Lu, Markus R Eisele, Endah S Sulistijo, Nafeesa Khan, Rong Fan, and Kathryn Miller-Jensen (2015). “Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation”. In: *Science signaling* 8.381, ra59–ra59.
- Yang, Zhiyuan, Jasper HN Yik, Ruichuan Chen, Nanhai He, Moon Kyoo Jang, Keiko Ozato, and Qiang Zhou (2005). “Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4”. In: *Molecular cell* 19.4, pp. 535–545.
- Yokoshi, Moe, Kazuma Segawa, and Takashi Fukaya (2020). “Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication”. In: *Molecular Cell*. ISSN: 1097-2765. DOI: <https://doi.org/10.1016/j.molcel.2020.02.007>. URL: <http://www.sciencedirect.com/science/article/pii/S1097276520300770>.
- Yoon, Young Soo, Sangkyun Jeong, Qi Rong, Kye-Yoon Park, Jae Hoon Chung, and Karl Pfeifer (2007). “Analysis of the H19ICR insulator”. In: *Molecular and cellular biology* 27.9, pp. 3499–3510.
- Yu, Wen-Mei, Xia Liu, Jinhua Shen, Olga Jovanovic, Elena E Pohl, Stanton L Gerson, Toren Finkel, Hal E Broxmeyer, and Cheng-Kui Qu (2013). “Metabolic regulation by the mitochondrial phosphatase PTPMT1 is required for hematopoietic stem cell differentiation”. In: *Cell stem cell* 12.1, pp. 62–74.
- Yuen, Kobe C, Brian D Slaughter, and Jennifer L Gerton (2017). “Condensin II is anchored by TFIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters”. In: *Science advances* 3.6, e1700191.

- Zenklusen, Daniel, Daniel R Larson, and Robert H Singer (2008). “Single-RNA counting reveals alternative modes of gene expression in yeast”. In: *Nature structural & molecular biology* 15.12, p. 1263.
- Zhang, Martin JinYE, Vasilis Ntranos, and David Tse (2020). “Determining sequencing depth in a single-cell RNA-seq experiment”. In: *Nature communications* 11.1, pp. 1–11.
- Zhao, Zhihu et al. (2006). “Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions”. In: *Nature genetics* 38.11, pp. 1341–1347.
- Zheng, Meizhen et al. (2019). “Multiplex chromatin interactions with single-molecule precision”. In: *Nature* 566.7745, pp. 558–562.