

Role of CLZ1 in maintenance of epigenetic state

Emma Rachael Stewart

PhD

University of York

Biology

September 2019

Abstract

The ability to transcriptionally repress genes is essential for cell survival. X chromosome inactivation is a widely used model for studying establishment and maintenance of facultative heterochromatin. This process requires Xist lncRNA and recruitment of chromatin modifiers including polycomb repressive complexes (PRC1/2), which lay down the classical epigenetic marks H2AK119Ub1 and H3K27me3. The nuclear matrix (NM) is the insoluble nuclear substructure which remains following extraction of soluble proteins, lipids, chromatin and, in some protocols, RNA. This thesis focuses on CIZ1, a NM protein known for its role in DNA replication. A CIZ1-null mouse model has shown CIZ1 loss results in fully penetrant female-specific lymphoproliferative disorder, and CIZ1-null primary embryonic fibroblasts (PEFs) display loss of Xist localisation and epigenetic mark enrichment from the inactive X chromosome (Xi).

I have optimised a protocol capable of profiling how proteins are anchored within the nucleus and generated a model of NM structure at the Xi. I demonstrate that CIZ1 is normally anchored via attachment to RNA but becomes part of the core protein NM during Xi replication, coincident with Xi relocation from the nuclear periphery to the nucleolus and back. Compromised Xi relocation in CIZ1-null PEFs is accompanied by increased solubility of PRC2 catalytic subunit EZH2, and genome-wide deregulation of polycomb-regulated genes. Compromised Xi relocation, solubilisation of EZH2 and polycomb target deregulation are also observed in WT cells following adaption to long term culture, suggesting this CIZ1-dependent process is typically compromised in cell lines. CIZ1 loss also appears to lead to DREAM complex target deregulation and expression of unannotated transcripts in activated CIZ1-null lymphocytes. The data are consistent with CIZ1 having a role in how enzyme and template meet, likely contributing to the maintenance of epigenetic landscape in PEFs, and offers a pathway that could explain the widespread links between CIZ1 and cancer.

Table of Contents

Abstract	2
Table of Contents	3
List of Tables	9
List of Figures	10
List of Appendices	13
Acknowledgements	15
Declaration	16
1.0 Introduction	17
1.1 X Chromosome Inactivation	18
1.1.1 Initiation	20
1.1.2 Spreading	20
1.1.3 Maintenance	21
1.1.4 <i>Xist</i> structure.....	22
1.1.5 Polycomb repressive complex recruitment.....	22
1.1.6 Essential roles of XCI factors	25
1.2 The Nuclear Matrix	27
1.2.1 DNA Replication factories.....	27
1.2.2 Nuclear matrix proteins involved in XCI.....	28
1.2.3 Destabilisation in cancer and cell lines	29
1.3 Cell Cycle	29
1.3.1 Growth	29

1.3.2	Quiescence.....	30
1.4	Cip1-interacting zinc finger protein 1 (CIZ1).....	31
1.4.1	Role of CIZ1 in DNA replication.....	31
1.4.2	CIZ1 and disease	32
1.4.3	Phenotypes of CIZ1 null PEFs	33
1.5	Project Aims.....	34
2.0	Materials and Methods	35
2.1	Cell Culture	35
2.1.1	Cell Isolation	35
2.1.2	Cell culture.....	36
2.1.3	Derivation of cell lines.....	36
2.1.4	Ethics.....	37
2.2	Sub-nuclear fractionation	38
2.2.1	Extraction for analysis by western blot	38
2.2.2	Extraction for analysis by immunofluorescence microscopy	39
2.3	Antibodies and detection protocols	39
2.3.1	Western Blot.....	39
2.3.2	Immunofluorescence Microscopy	40
2.3.3	Proximity Ligation Assay (PLA)	41
2.3.4	S phase labelling	41
2.4	Imaging.....	42
2.4.1	Image Capture	42
2.4.2	Image Processing.....	42

2.5	Statistical Analysis	43
2.6	Transcriptome analysis	43
2.6.1	RNA Isolation.....	43
2.6.2	Sample Preparation and Sequencing.....	44
2.6.3	Transcriptome Assembly	44
2.6.4	Data Analysis.....	45
2.6.5	Data storage and availability	45
3.0	Visualisation of masked epitopes after nuclear matrix extraction....	46
3.1	Introduction.....	46
3.2	Aims	47
3.3	Experimental Design	47
3.4	Results and Discussion	48
3.4.1	Nuclear matrix interactions	48
3.4.2	Nucleus Wide Superficial and Buried Populations	50
3.4.3	Revealing masked epitopes at the Xi	53
3.4.4	Summary model of nuclear matrix proteins at the Xi.....	56
3.4.5	Detection by western blot	57
3.4.6	CIZ1 isoform expression upon induction of pluripotency.....	59
3.5	Conclusions	61
4.0	Effect of CIZ1 deletion on gene expression in fibroblasts.....	62
4.1	Introduction.....	62
4.2	Aims	63

4.3	Experimental Design	63
4.4	Results and Discussion	65
4.4.1	Genome wide effects of CIZ1 loss and culture adaption	65
4.4.2	Loss of CIZ1 has no preferential effect on gene expression from the X chromosome	68
4.4.3	Loss of CIZ1 results in PRC target deregulation genome wide	74
4.5	Conclusions	78
5.0	Compromised epigenetic maintenance in CIZ1 null PEFs and cell lines	79
5.1	Introduction.....	79
5.2	Aims	82
5.3	Results and Discussion	83
5.3.1	CIZ1 at replicating Xi is resistant to extraction with RNase	83
5.3.2	Nuclear Myosin 1 as a possible component of chromatin movement machinery	88
5.3.3	Isoform preference of different EZH2 antibodies	90
5.3.4	EZH2p is resistant to extraction.....	95
5.3.5	EZH2 and CIZ1 are in close proximity in primary WT cells	98
5.3.6	Re-establishment of epigenetic marks in culture adapted cells.....	100
5.3.7	Proposed model	102
5.4	Conclusions	104
6.0	Loss of CIZ1 in young adult splenic lymphocytes.....	105
6.1	Introduction.....	105
6.2	Aims	106

6.3	Experimental Design	107
6.4	Results and Discussion	108
6.4.1	Generation of RNA	108
6.4.2	Effects of CIZ1 loss on gene expression from the X Chromosome	109
6.4.3	Genome wide effects of CIZ1 loss.....	111
6.4.4	Identification of activation gene sets for further analysis	116
6.4.5	Deregulation of cell cycle genes in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes	119
6.4.6	Deregulation of mitotic checkpoint genes in CIZ1 null spleens	124
6.4.7	Additional avenues investigated	130
6.4.7.1	<i>TCF21</i>	130
6.4.7.2	<i>Apoptosis</i>	130
6.4.7.3	<i>FOXP3</i>	132
6.4.7.4	<i>Deregulation of TCF21 targets and genes bound by FOXP3 in CIZ1 null spleens compared to WT spleens.....</i>	135
6.4.8	Summary of avenues for investigation	139
6.4.8.1	<i>Unannotated transcripts</i>	139
6.4.8.2	<i>Mitosis</i>	139
6.4.8.3	<i>Limitations</i>	139
6.5	Conclusions	140
7.0	Discussion and Future Directions.....	141
7.1	How is the nuclear matrix organized at the Xi?	141
7.2	CIZ1-dependent chromosomal relocation.....	142
7.3	Deregulation of chromatin remodelling complex targets.....	143

7.4	Expression of unannotated transcripts.....	144
7.5	A role for CIZ1 in chromatin remodelling?.....	145
7.6	Limitations of using cell lines	146
7.6.1	Differences between primary cells and ‘normal’ cell lines	146
7.6.2	Suitability of cell lines when investigating the role of CIZ1	146
7.6.3	Implications.....	147
7.6.4	Developing a better model for mechanism of movement	148
7.7	Limitations of splenocyte analysis	149
7.8	Conclusions	150
Appendix A		151
Appendix B		165
Appendix C		182
Appendix D		197
Appendix E		211
Appendix F		216
Appendix G		219
Appendix H		221
List of Abbreviations		230
References		233

List of Tables

Table 1.1	A list of the subunits in canonical PRC1 and PRC2 and their functions (Simon, 2009)	23
Table 2.1	List of primary cell strains and cell lines	37
Table 2.2	Primary antibodies used for western blot and immunofluorescence.	40
Table 4.1	List of X-linked CIZ1 dependent genes and their known roles ..	68
Table 4.2	PRC related gene sets from GSEA curated gene sets	74
Table 4.3	PRC related gene sets from GSEA oncogenic signatures	75
Table 6.1	The number of gene sets which appear in both B and T cell top 20 outputs of GSEA analysis for all four categories.	119
Table 6.2	The 21 CIZ1-dependent genes identified in (Figure 6.16D)....	138

List of Figures

Figure 1.1	Schematic of imprinted and random XCI.....	19
Figure 1.2	Repeat structure of Xist lncRNA.....	22
Figure 1.3	Models of PRC recruitment to the Xi.....	26
Figure 1.4	Cell cycle.....	30
Figure 1.5	Schematic of mouse CIZ1 protein.....	31
Figure 1.6	Phenotypes of CIZ1 null PEFs.....	34
Figure 3.1	Overview of the nuclear matrix extraction protocol.....	47
Figure 3.2	RNA-bound CIZ1 at the inactive X chromosome is cross-linkable to the core protein NM (Ridings-Figueroa et al., 2017).....	49
Figure 3.3	CIZ1 Antibody Map.....	50
Figure 3.4	Reveal of previously masked epitopes following chromatin removal (Swarts et al., 2018).....	53
Figure 3.5	Nuclear matrix extraction profile of YY1 and SAF-A (Ridings-Figueroa et al., 2017, Stewart and Coverley, 2018).....	54
Figure 3.6	Line intensity scans on unextracted and extracted nuclei (Stewart and Coverley, 2018).....	55
Figure 3.7	Model of Nuclear Matrix Structure (Ridings-Figueroa et al., 2017).....	56
Figure 3.8	Schematic of nuclear matrix extraction for detection by western blot.....	57
Figure 3.9	RNase inhibitor VRC prevents H3 extraction.....	58
Figure 3.10	Changes in CIZ1 splice variant expression upon induction of pluripotency.....	60

Figure 4.1	Schematic of RNA isolation from WT and CIZ1 null cells	64
Figure 4.2	The effect of loss of CIZ1 and culture adaption on gene expression	67
Figure 4.3	Effect of loss of CIZ1 on X-linked gene expression.....	69
Figure 4.4	Expression of components of X chromosome inactivation.....	71
Figure 4.5	Isoforms of polycomb repressive complex subunits.....	74
Figure 4.6	Loss of CIZ1 and culture adaption results in PRC target deregulation genome wide	76
Figure 4.7	Downregulation of polycomb repressive complex targets in primary embryonic fibroblasts	78
Figure 5.1	CIZ1-dependent Xi relocation during its replication.....	81
Figure 5.2	Xi-CIZ1 is RNase resistant in a subpopulation of cells.....	83
Figure 5.3	The inactive X chromosome replicated during mid-S phase	85
Figure 5.4	CIZ1 at replicating Xi is resistant to extraction with RNase.....	87
Figure 5.5	Downregulation of Nuclear Myosin 1 and actins following adaption to long term culture.....	89
Figure 5.6	EZH2 isoforms, and their relative antigenicity.....	91
Figure 5.7	Differential isoform detection with EZH2 antibodies.....	94
Figure 5.8	Loss of CIZ1 drives changes in EZH2 binding	97
Figure 5.9	Proximity Ligation Assay between CIZ1 and EZH2.....	99
Figure 5.10	Reinstatement of epigenetic marks in culture adapted cells ...	102
Figure 5.11	Model of step down of epigenetic control through adaption....	103
Figure 6.1	Recruitment of CIZ1 and Xist to the Xi upon activation.....	106

Figure 6.2	Schematic of RNA isolation from lymphocytes and spleens ...	107
Figure 6.3	Analysis of lymphocyte RNA quality	108
Figure 6.4	Effect of CIZ1 loss on expression from the X chromosome	110
Figure 6.5	CIZ1 loss results in up-regulation of immunity related genes in lymphocytes in the absence of activation signals.....	113
Figure 6.6	Effect of activation on WT and CIZ1 null lymphocytes	115
Figure 6.7	Identification of activation gene sets for further analysis.....	117
Figure 6.8	Unannotated transcripts in gene activation sets.....	118
Figure 6.9	Inappropriate upregulation and failure to upregulate cell cycle and DNA damage response genes upon lymphocyte activation	120
Figure 6.10	Upregulation of cell cycle and DNA damage response genes upon lymphocyte activation	122
Figure 6.11	CIZ1 dependent deregulation of DREAM targets.....	126
Figure 6.12	CIZ1 dependent deregulation of DREAM targets.....	128
Figure 6.13	Gene set enrichment analysis of X-linked CIZ1-dependent genes in diseased spleens.....	129
Figure 6.14	Targets of TCF21 and genes induced during apoptosis	132
Figure 6.15	Deregulation of genes bound by FOXP3 but not targets of FOXP3.....	134
Figure 6.16	CIZ1 dependent deregulation TCF21 targets and genes bound by FOXP3.....	137

List of Appendices

Appendix A

PDF of peer-reviewed publication

Ridings-Figueroa R, **Stewart ER**, Nesterova TB, Coker H, Pintacuda G, Godwin J, et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* 2017.

Appendix B

PDF of peer-reviewed publication

Swarts DRA, **Stewart ER**, Higgins GS, Coverley D. CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours. *Cell Cycle.* 2018;17(18):2268-83.

Appendix C

PDF of peer-reviewed publication

Stewart ER, Coverley D. Visualization of Hidden Epitopes at the Inactive X Chromosome. *Methods Mol Biol.* 2018;1861:103-12.

Appendix D

PDF of peer-reviewed publication

Stewart ER, Turner RML, Newling K, Ridings-Figueroa R, Scott V, Ashton PD, et al. Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture. *Nat Commun.* 2019;10(1):460.

Appendix E

Example python scripts used to perform bioinformatic analysis.

Appendix F

Sequences of EZH2 splice variants 'TCONS_00153598' and 'TCONS_00153599'.

Appendix G

Table of X-linked genes significantly changed in CIZ1 null spleens compared to WT, relating to (**Chapter 6**).

Appendix H

A series of tables, labelled 1-8, containing top 20 outputs from gene set enrichment analysis (GSEA). Relates to **Chapter 6**.

Acknowledgements

I would like to thank the BBSRC for funding this PhD, my supervisors Dawn Coverley and Bob White for giving me the opportunity to carry out this work. I would also like to thank my TAP panel, Paul Genever and Michael Plevin for their advice and support throughout my PhD.

Thank you to the Coverley lab, past a present. A special thanks to Dawn for your endless support and guidance over the past 4 years, encouraging the amazing opportunities I have had, and prompt feedback on the many bits of writing I have thrown your way! In particular, I would like to thank Justin for all of his advice and helpful comments; and Robert, Rebeca, Lindsey, Victoria, Dorian and Barbora for their contributions to my research. Alongside Louisa, Grace, Gaby, Sajad and Ernesto you have all made the lab an enjoyable place to be. (And a special extra shout out to Gaby for helping me move to a new house on possibly the hottest day ever!)

Thank you to everyone who has collaborated with me, particularly Katherine Newling, Sally James, Peter Ashton, Neil Brockdorff and James Hewitson; and the countless staff members at the University of York who have been incredibly supportive. I am also grateful to Sue, Jim, Adi, Owen, Adam, Andy, Christine, Amy and all the students at York St John University for making my PIPs such a wonderful experience.

My PhD would not have been the same without my friends for the gradshares, quizzes and boardgames, in particular Lotte, Caroline, Jack, Alex, Phil, Iulia, Nathaniel, Tess and James. Thank you, Becky and Kate, for the gin and curry nights, and Holly and Laura for the endless laughs and awful TV!

Finally, I would like to especially thank my family. Thank you in particular to my parents, sisters and grandparents for your love, pride, encouragement and making me the person I am today. And of course, my partner and thesis writing buddy David. Thank you for all your love and always believing in me, you've been my absolute rock. We've had some amazing adventures these past 4 years and I can't wait for all the ones to come!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author, unless otherwise explicitly stated. My sole contributions to this work have not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Text, tables and figures contained in Chapters 2, 3, 4 and 5 are published in peer-reviewed journals on which I am primary or contributing author. These are referenced at the beginning of each Chapter. My contributions to these publications have been reproduced in full. Where necessary to communicate the scientific story, I have described data from these papers generated by other authors, and explicitly credited the work to the particular author.

The peer-reviewed publications I have contributed to are:

Ridings-Figueroa R, **Stewart ER**, Nesterova TB, Coker H, Pintacuda G, Godwin J, et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* 2017.

Swarts DRA, **Stewart ER**, Higgins GS, Coverley D. CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours. *Cell Cycle.* 2018;17(18):2268-83.

Stewart ER, Coverley D. Visualization of Hidden Epitopes at the Inactive X Chromosome. *Methods Mol Biol.* 2018;1861:103-12.

Stewart ER, Turner RML, Newling K, Ridings-Figueroa R, Scott V, Ashton PD, et al. Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture. *Nat Commun.* 2019;10(1):460.

1.0 Introduction

A cell's ability to transcriptionally repress gene expression is a vital requirement for survival. For example, differentiating embryonic stem (ES) cells need to shut down pluripotency genes and prevent transcription of genes related to other cell lineages (for example, a cell differentiating down a mesoderm lineage should not express genes relating to the ectoderm lineage) (Miyamoto et al., 2015). This process can include multiple factors such as expression of repressive long non-coding RNAs (lncRNAs) and recruitment of chromatin modifying enzymes which lay down repressive epigenetic marks (Fernandes et al., 2019, Miyamoto et al., 2015). These processes result in production of facultative heterochromatin, making the gene to be 'shut down' inaccessible to RNA polymerases. Crucially, these changes need to be maintained through subsequent cell divisions to prevent aberrant gene reactivation which could lead to cell death, lineage deviation, or cellular transformation (reviewed in (Delgado and Leon, 2006)).

X chromosome inactivation has become a widely used model for studying genome-wide transcriptional silencing, as well as the regulatory role of long noncoding RNAs and their significance.

This thesis focuses on Cip1-interacting zinc finger protein 1 (CIZ1), a DNA replication protein with numerous links to disease, and its role in maintenance of the cells epigenetic landscape. My research provides mechanistic insight into how epigenetic marks are maintained through cell division and how this process is corrupted in cells adapted to long term culture.

1.1 X Chromosome Inactivation

X chromosome inactivation (XCI) is a mammalian dosage compensation mechanism that prevents double dosage of genes located on the X chromosome in XX females compared to XY males (Lyon, 1961). During XCI one of the X chromosomes in female cells is transcriptionally silenced and condenses into a Barr body (Barr and Bertram, 1949) (becoming the inactive X chromosome, Xi) while the other remains transcriptionally active (becoming the active X chromosome, Xa) (Payer and Lee, 2008).

There are two types of X inactivation in eutherian mammals (**Figure 1.1**). The first is imprinted XCI, in which the paternal X chromosome is always inactivated (Okamoto et al., 2004). This is the only form of XCI in marsupial mammals (Sharman, 1971). In eutherian mammals this occurs in early embryogenesis before implantation and the paternal X remains inactivated in all extra-embryonic tissues (Takagi and Sasaki, 1975, Okamoto et al., 2004). At the blastula stage, the paternal X chromosome in the inner cell mass (ICM), the cells that go on to become the embryo, is reactivated and later followed by random XCI (Okamoto et al., 2004). During random XCI either the maternal or paternal X chromosome can be inactivated and the same X chromosome will not necessarily be inactivated in every cell (Okamoto et al., 2004, Lyon, 1961).

It is important to note that not every single gene in the Xi is switched off. There are a small portion of genes that 'escape' the inactivation and continue to be expressed from the otherwise transcriptionally repressed Xi (Berletch et al., 2011). Escapees seem to be less common in mice (~3%) than in humans (~15%). Humans also have an additional 10% of genes, which escape in certain cell and tissue-types (Yang et al., 2010, Berletch et al., 2015). Though there appears to be partial conservation, with many of the escape genes overlapping in mice and humans, there are significant differences (Yang et al., 2010).

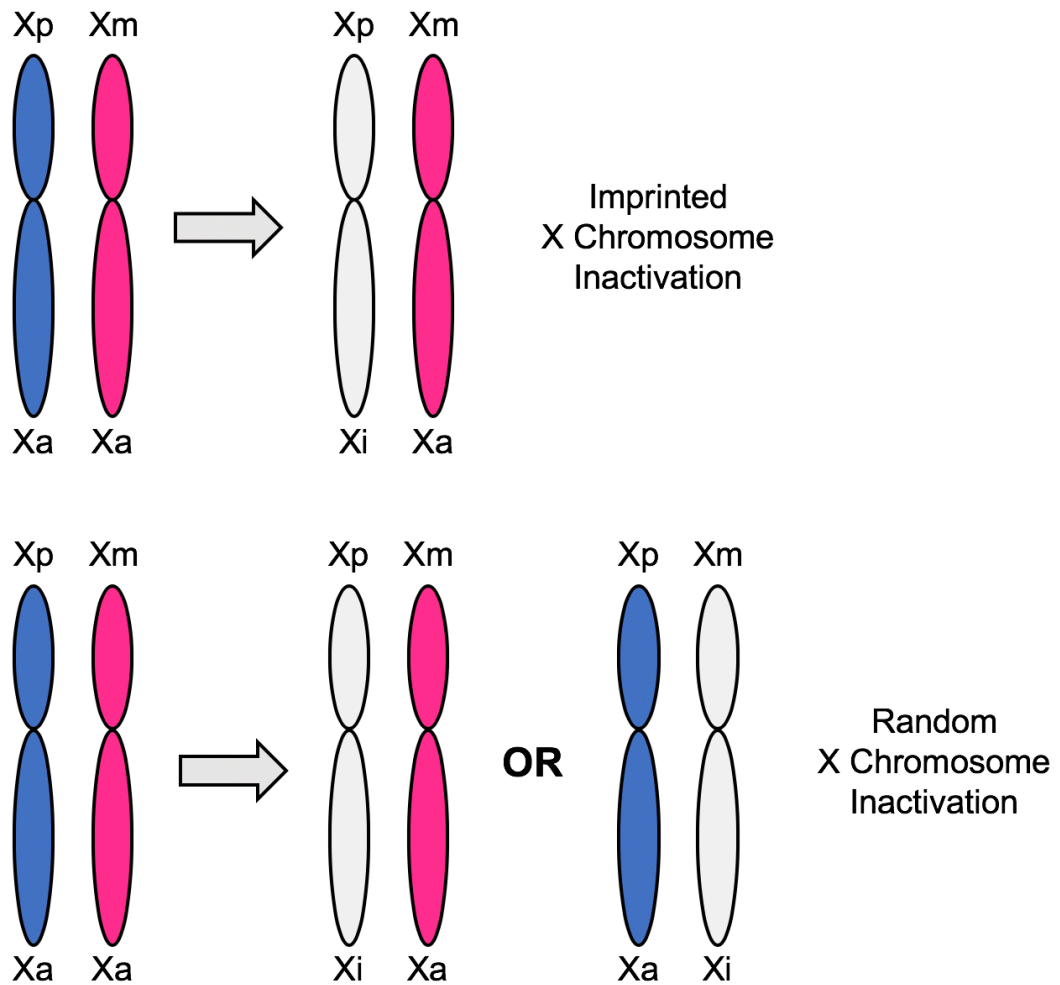


Figure 1.1 Schematic of imprinted and random XCI

(Upper) Schematic representing transcriptional repression (grey shading) of the paternal X chromosome (Xp) in imprinted XCI. (Lower) In random XCI either the maternal (Xm) or paternal (Xp) X chromosome will be transcriptionally silenced at approximately equal frequencies (Okamoto et al., 2004).

1.1.1 Initiation

Initiation of XCI is controlled by a region of the X chromosome called the X inactivation centre (XIC). During differentiation, XCI is initiated by up-regulation of *Xist* (X-inactive specific transcript), a 17kb long non-coding RNA (lncRNA) (Brown et al., 1992).

Prior to XCI, *Xist* expression is suppressed by numerous factors including expression of another non-coding RNA, *Tsix*, which suppresses *Xist* transcription through antisense transcription (Lee et al., 1999). Pluripotency factors such as OCT4 (octamer-binding transcription factor 4) negatively regulate *Xist* by binding to the *Xist/Tsix* locus and repressing *Xist* expression while promoting *Tsix* expression (Donohoe et al., 2009). Upon differentiation the *Xist*-suppressing pluripotency factors are downregulated and expression of *Xist* activators such as RNF12 (Ring Finger Protein 12) are upregulated (Jonkers et al., 2009), eventually resulting in *Xist* expression on the future Xi (Augui et al., 2011).

If a cell has an abnormal number of X chromosomes such as in the chromosomal disorders XXX females (triple-X-syndrome) and XXY males (Klinefelter syndrome), extra X chromosomes can also be inactivated to leave only a single active X (Xa) (Monkhorst et al., 2008, Monkhorst et al., 2009). However, this appears to be mosaic (Monkhorst et al., 2009). Though there are many hypotheses relating to how a cell 'counts' the number of X chromosomes it has (Starmer and Magnuson, 2009, Lyon, 1971) (Gartler and Riggs, 1983), and how a cell then chooses which to inactivate (Augui et al., 2007), the mechanism is still unproven.

1.1.2 Spreading

Once *Xist* expression has been upregulated on the future Xi, it must then spread in-cis to cover the entire X chromosome in order for transcriptional silencing to occur. The mechanism by which *Xist* spreads along the X chromosome is a subject of intense investigation but still remains unclear. One proposed mechanism suggests long interspersed repetitive sequence elements (LINES), which are enriched on the X chromosome compared to

autosomes, function as a 'booster', increasing *Xist* spreading efficacy (Lyon, 2000). This is supported by the observation that ectopic expression of *Xist* on autosomes, results in less efficient chromosome-wide spreading and silencing. However ectopic *Xist* is still capable of localised inactivation, and the density of LINES on the X chromosome are highest near the XIC and lowest near 'escape' genes (Hoge et al., 2003).

Multiple RNA-binding proteins and histone modifying complexes have also been reported to have a role in XCI. Transcription factor YY1 (Yin and Yang Protein 1) and SAF-A (Scaffold-Attachment Factor A, also known as Heterogenous Nuclear Ribonucleoprotein U, hnRNP U) have both been reported as having a role in *Xist* localization (Jeon and Lee, 2011, Hasegawa et al., 2010) and the polycomb repressive complexes PRC1 and PRC2 are required for the deposition of the H2AK119Ub1 and H3K27me3 repressive epigenetic marks respectively (reviewed in (Brockdorff, 2017)).

1.1.3 Maintenance

Following spreading, one X chromosome is now in an inactivated state which requires maintenance through subsequent cell divisions. At this stage *Xist* (although still expressed) is no longer necessary, as conditional *Xist* deletion does not result in reactivation of Xi gene expression (Csankovszki et al., 1999).

Temporary relocation of the Xi from the nuclear lamina to the peri-nucleolar ring during mid-to-late S phase has been implicated in maintenance of epigenetic mark enrichment on Xi chromatin. Loss of *Xist* or knockdown of another X-linked lncRNA *Firre* (Functional Intergenic Repeating RNA Element) results in loss of peri-nucleolar localisation and loss of H3K27me3 (Zhang et al., 2007, Yang et al., 2015). While deletion of *Xist* led to single-gene reactivations, knockdown of *Firre* did not (Yang et al., 2015, Zhang et al., 2007). However, knockdown of *Firre* did lead to a, possibly compensatory, increase in expression of X silencing linked chromatin modifiers (Yang et al., 2015). So far, only these lncRNAs have been implicated in in relocation of Xi.

1.1.4 *Xist* structure

Xist (X-inactive specific transcript), is a 17kb lncRNA containing several repeat regions labelled A-F (**Figure 1.2**) (Brown et al., 1992). Though the function of many of these regions has been characterised, the mechanism behind the function often remains to be elucidated. A-repeats are known to be essential for spreading and silencing, but the mechanism remains unclear (Moindrot et al., 2015). Repeats B and F, collectively known as the XN region, are thought to be involved in recruitment of PRC2 (da Rocha et al., 2014). C-repeats are thought to be involved in the association of *Xist* with the X chromosome (Sado and Brockdorff, 2013). The role of the E-repeats is also unclear but recent papers have shown that exon 7, which contains the E-repeats, is crucial for *Xist* retention at its site of expression on Xi (Yamada et al., 2015). It has recently been revealed that the D-repeats have a role in regulating *Xist* expression but again, the mechanism is unknown (Lv et al., 2016).

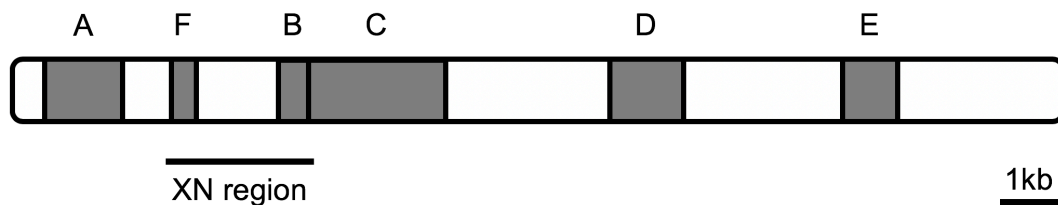


Figure 1.2 Repeat structure of *Xist* lncRNA

1.1.5 Polycomb repressive complex recruitment

Polycomb repressive complexes (PRC) are histone modifying complexes, which lay down two classical epigenetic marks typically associated with the Xi. PRC1 mono-ubiquitinates lysine 119 on histone 2A (H2AK119Ub1) and PRC2 tri-methylates lysine 27 of histone 3 (H3K27me3) (Margueron and Reinberg, 2011). There are multiple variants of PRC1 and PRC2 depending which of the numerous possible subunits they consist of (**Table 1.1**) (Simon, 2009). Canonical PRC1 and PRC2 is the name given to the complexes formed by the main subunits (Simon, 2009, Schwartz and Pirrotta, 2013). *Xist* is required for recruitment of these complexes to the Xi, however the mechanism behind the

recruitment, and the order in which the complexes are recruited, is still highly debated (Kohlmaier et al., 2004).

Table 1.1 A list of the subunits in canonical PRC1 and PRC2 and their functions (Simon, 2009)

Complex	Core Subunit	Function	Variants
PRC1	RING	E3 Ubiquitin Ligase	RING1A
			RING1B
	PCGF	Co-factor for E3 Ubiquitin Ligase	PCGF2 (MEL18)
			PCGF4 (BMI1)
	PHC	Unknown	PHC1 (EDR1)
			PHC2 (EDR2)
			PHC3 (EDR3)
	CBX	Chromodomain recognises H3K27me3	CBX2
			CBX4
			CBX6
			CBX7
			CBX8
	PRC2	EZH	H3K27 methyltransferase
EZH2			
EED		Support EZH	EED
SUZ12		Support EZH	SUZ12
RBAP		Histone Binding	RBBP4 (RBAP48)
			RBBP7 (RBAP46)

PRC2 recruitment to the Xi is least well understood with multiple conflicting publications. There are several reports that the catalytic subunit of PRC2, EZH2 (Enhancer of Zeste 2), binds *Xist* A repeats either directly (Zhao et al., 2008) or mediated through post-translational modifications such as phosphorylation (Kaneko et al., 2010) or cofactors such as Jarid2 (Jumonji And A-T rich Interaction Domain Containing 2, through binding of *Xist* XN region) (da Rocha et al., 2014) and SPEN (SMART/HDAC1-Associated Repressor Protein, unknown mechanism through the *Xist* A repeats) (McHugh et al., 2015, Moindrot et al., 2015, Monfort et al., 2015). Evidence that *Xist* lacking the A repeats is still capable of PRC2 recruitment contradicts the hypotheses of direct EZH2 binding and mediation through SPEN (da Rocha et al., 2014, Wutz et al., 2002). Additionally, no canonical PRC2 subunits were identified in proteomic studies of the *Xist* interactome (Chu et al., 2015, McHugh et al., 2015). Mediation through Jarid2 has also been challenged by observations that deletion of the region of Jarid2 hypothesised to interact with *Xist* XN region had no effect on PRC2 recruitment (da Rocha et al., 2014, Kaneko et al., 2014). However, as these hypotheses require different regions of *Xist*, it is possible that more than one of these mechanisms are correct and can compensate in the absence of the other.

Recruitment of PRC1 to the Xi was originally hypothesised to occur based on the classical model of the CBX (Chromobox protein) subunit of canonical PRC1 recognising H3K27me3 already deposited by PRC2 (Cao et al., 2002), therefore generating the model that PRC2 recruitment must precede PRC1. However, this was opposed by the discovery of non-canonical PRC1, in which substitution of the CBX subunit for RYBP (RING1 and YY1 binding protein) allows H2AK119Ub1 deposition at the Xi in the absence of H3K27me3 (Tavares et al., 2012, Schoeftner et al., 2006). Identification of non-canonical PRC1 subunits including RYBP and RING1 (Ring Finger Protein 1) in proteomic studies of the *Xist* interactome (Chu et al., 2015) supports a hypothesis of direct interaction between non-canonical PRC1 and *Xist*.

The identification of a PRC2-independent mechanism of PRC1 recruitment throws into question the hypothesis that PRC2 must initiate the polycomb

cascade. In fact, further reports demonstrating PRC2 can recognise H2AK119Ub1 suggests that PRC1 could in fact precede PRC2 (Kalb et al., 2014, Cooper et al., 2014, Blackledge et al., 2014), leading to two competing models (**Figure 1.3**).

1.1.6 Essential roles of XCI factors

Knocking-out many factors involved in X inactivation is lethal during embryonic development. This includes *Xist* (Senner et al., 2011, Marahrens et al., 1997), RNF12 (Shin et al., 2010), EZH2 (O'Carroll et al., 2001), EED (Faust et al., 1998), SUZ12 (Pasini et al., 2004) and RING (Voncken et al., 2003). Though some PRC component knockouts are viable, they can have severe developmental defects, such as in the case of BMI1 (a Polycomb Group Ring Finger (PCGF) protein) knockout, which results in skeletal and neurological abnormalities and severe haematopoietic defects (Richly et al., 2011, van der Lugt et al., 1994). A common phenotype of loss of PRC subunits is failure to differentiate (Richly et al., 2011). Overexpression of polycomb repressive complex subunits is also detrimental and has been linked with numerous cancers. This includes breast cancer (Yoo and Hennighausen, 2012), lung cancer (Sato et al., 2013) and lymphoma (Raaphorst, 2005). This demonstrates the critical role of the factors involved in X inactivation.

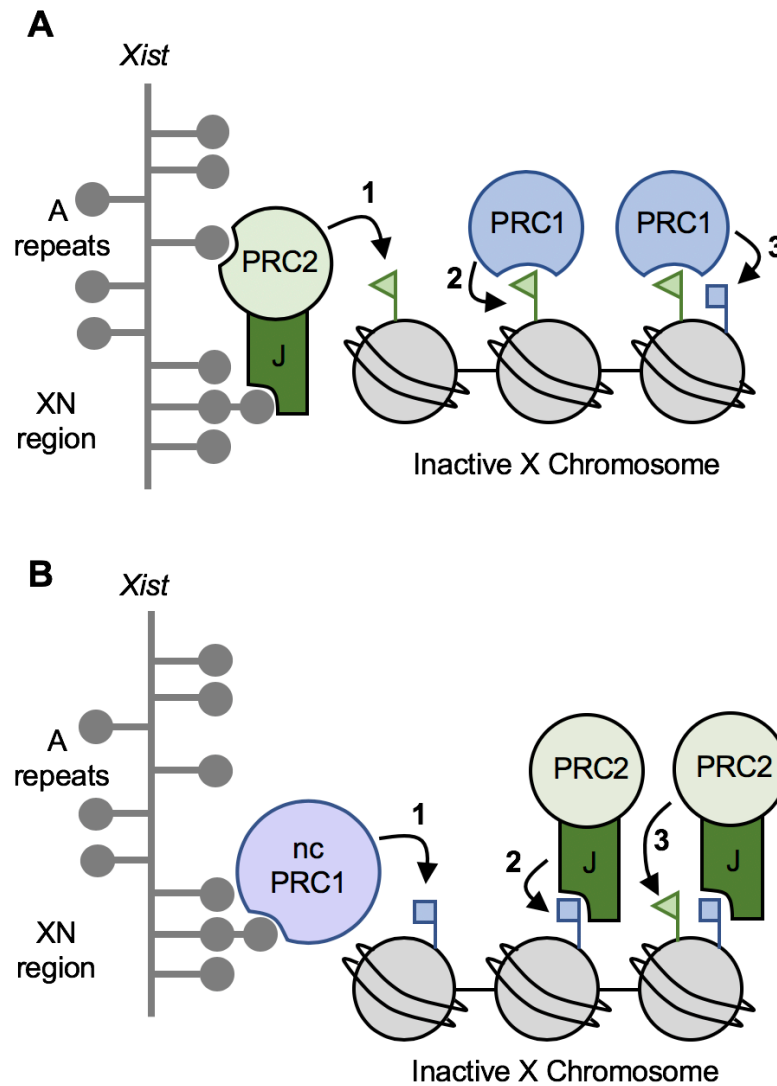


Figure 1.3 Models of PRC recruitment to the Xi

(A) Classical hierarchical model (Cao et al., 2002) showing (1) PRC2 initiates the polycomb cascade through interaction with *Xist* A repeats and/or interaction with *Jarid2* (J), which binds the XN region of *Xist*. (2) PRC2 deposits H3K27me3 (green flag) on the Xi, which is detected by the CBX subunit of canonical PRC1. (3) This results in PRC1 recruitment and deposition of H2AK119Ub1 (blue flag). (B) Reverse hierarchical model (Brockdorff, 2017) whereby (1) non-canonical PRC1 initiates the polycomb cascade by interaction with *Xist* at the XN region and deposition of H2AK119Ub1 (blue flag) at the Xi. (2) PRC2 is recruited by PRC1-mediated H2AK119Ub1, possibly through external mediators such as *Jarid2* (J, which contains a ubiquitin interaction motif (Cooper et al., 2016)), and (3) deposits H3K27me3 (green flag) at the Xi. This could be followed by further canonical PRC1 recruitment (not shown).

1.2 The Nuclear Matrix

Many of the proteins involved in XCI are sometimes referred to as part of the Nuclear Matrix (NM), including YY1 and SAF-A (Jeon and Lee, 2011, Helbig and Fackelmayer, 2003). In this thesis, I define the nuclear matrix biochemically, as the insoluble nuclear substructure that remains following extraction of soluble proteins, lipids, and chromatin. I further define the RNA-independent nuclear matrix as the protein fraction that remains after removal of all of these components, in addition to removal of RNA by enzymatic digestion (Verheijen et al., 1988).

First used in 1974, the term nuclear matrix was used to describe the insoluble fraction of the nucleus, consisting of RNA and protein, which remained following extraction with 2M NaCl (Berezney and Coffey, 1974). Electron microscopy of extracted nuclei revealed a network of 10 nm filaments; however the method was criticised as high levels of salt promotes protein aggregation (Capco et al., 1982, Wan et al., 1999, Verheijen et al., 1988). This led to development of a new gentler method of extraction, instead using 0.5 M NaCl and DNase I to digest the chromatin (Capco et al., 1982). This is the method I have refined in this thesis to reveal the RNA-protein nuclear matrix, and further refined to reveal the core protein nuclear matrix.

DNA replication (Jackson and Cook, 1986), DNA repair (Qiao et al., 2001), transcription (Jackson and Cook, 1985) and RNA splicing (Zeitlin et al., 1987) are all previously reported to occur in association with the nuclear matrix.

1.2.1 DNA Replication factories

Chromatin is periodically attached to the nuclear matrix at scaffold/matrix attachment regions (S/MARs), generating chromatin loop structures, which can be measured (Wilson and Coverley, 2017). During DNA replication, these loops are individual replicons (Buongiorno-Nardelli et al., 1982, Lemaitre et al., 2005). Enrichment of origins of replication and DNA replication factors at the base of chromatin loops led to these regions being described as 'DNA replication factories', where DNA replication is initiated and elongated

(Lagarkova et al., 1998, Munkley et al., 2011, Dijkwel et al., 1991, Hesketh et al., 2015). Replication factories, within which the required factors are delivered and immobilized, gives the cell spatial and temporal control over replication (reviewed in (Wilson and Coverley, 2013)).

1.2.2 Nuclear matrix proteins involved in XCI

SAF-A, also known as hnRNP U, is a nuclear matrix protein for which gene knock-out is embryonic lethal (Roshon and Ruley, 2005), and knock-down of the protein in an siRNA screen results in loss of *Xist* localisation and loss of H3K27me3 from the Xi (Hasegawa et al., 2010). SAF-A interacts directly with *Xist* via its RNA-binding domain, the RGG box, however interaction of its DNA binding domain (SAF-box) with scaffold/matrix attachment regions (S/MARs) on target chromatin is also required for *Xist* localisation (Hasegawa et al., 2010). Recent literature has implied that SAF-A binds *Xist* at exons 1 and 7, with truncation of exon 7 impairing the interaction (Yamada et al., 2015). It is possible that SAF-A at the Xi is differentially post-translationally modified, as many antibodies are not capable of recognising SAF-A at the Xi, but can be seen enriched at the Xi with other antibodies or when GFP-SAF-A is transfected into cells (Smeets et al., 2014).

YY1, also known as Nuclear Matrix Protein 1 (NMP-1) (Guo et al., 1995), is a transcription factor anchored to the nuclear matrix via its C-terminus, with roles in transcriptional repression and nuclear trafficking (McNeil et al., 1998, Bushmeyer and Atchison, 1998). YY1 has both DNA and RNA binding motifs and interacts with *Xist* via *Xist* C repeats, and potentially behaves as an 'adaptor' to tether *Xist* to its target chromatin, requiring both DNA and RNA interactions to tether *Xist* onto the Xi (Jeon and Lee, 2011, Makhlouf et al., 2014). Loss of YY1 has been shown to prevent the upregulation of *Xist* seen in the initiation and maintenance stages of XCI (Makhlouf et al., 2014). These data suggest possible roles for YY1 in controlling the cis spreading of *Xist* and regulation of its levels.

1.2.3 Destabilisation in cancer and cell lines

Multiple lines of evidence suggest that 'plastic' cells, such as stem cells, and cancer cells have a compromised or immature nuclear matrix (Khanuja et al., 1993, Varma and Mishra, 2011, Munkley et al., 2011). This includes an increase in chromatin loop sizes in cancer cells, suggesting fewer attachments to a nuclear matrix structure (Wilson and Coverley, 2017) and solubilisation of nuclear matrix proteins in cell lines derived from cancers (Khanuja et al., 1993, Munkley et al., 2011). As the nuclear matrix appears to be a regulatory structure, its degradation in cancers is not wholly unexpected. This has led to studies proposing nuclear matrix proteins as possible cancer biomarkers (Davido and Getzenberg, 2000).

1.3 Cell Cycle

1.3.1 Growth

In order for cells to accurately divide and multiply they must execute the sequence of events that make up the cell cycle, controlled by cyclins and cyclin-dependent kinases (CDKs) (reviewed in (Suryadinata et al., 2010)). The cell cycle can be split into four core stages (**Figure 1.4**), growth 1 (G1), synthesis (S), growth 2 (G2) and mitosis (M). G1, S and G2 are collectively known as 'Interphase'.

During G1, the cell typically increases in size and synthesises the components required to progress through into DNA synthesis, such as nucleotides and histones. In order to progress through to S phase, the cell must overcome the barrier of the restriction point (Pardee, 1974), R, after which the cell is committed to the next cell cycle. Following passage through R, the cell enters S and, if undamaged, faithfully replicates its entire genome. Initiation of DNA replication is dependent on sequential interaction of CDK2 with cyclin E followed by cyclin A (Coverley et al., 2002).

Once replicated the cell enters G2, and then M phase when it divides to produce two identical daughter cells, each with a complete copy of the original

genome. The new daughter cells exit mitosis into G1, and the cycle starts again (Norbury and Nurse, 1992).

1.3.2 Quiescence

Most cells in the body exist in a non-proliferative state, having exited the cell cycle. This state is called quiescence, or G0, which is normally entered during G1 (Martin and Stein, 1976, Smith and Martin, 1973). Failure to maintain quiescence, leading to superfluous cell cycling, is thought to be involved in tumourigenesis (Fiore et al., 2018). In a laboratory setting, quiescence can be induced through several mechanisms including contact inhibition (growing cells to high density), or deprivation of growth factors (Gos et al., 2005). Entry into and exit from quiescence involves large scale chromatin remodeling events (Mehta et al., 2010) and requires the dimerization partner, RB-like, E2F and multi-vulval class B (DREAM) complex, which represses pro-cell cycle genes to in order for quiescence to be maintained (Sadasivam and DeCaprio, 2013).

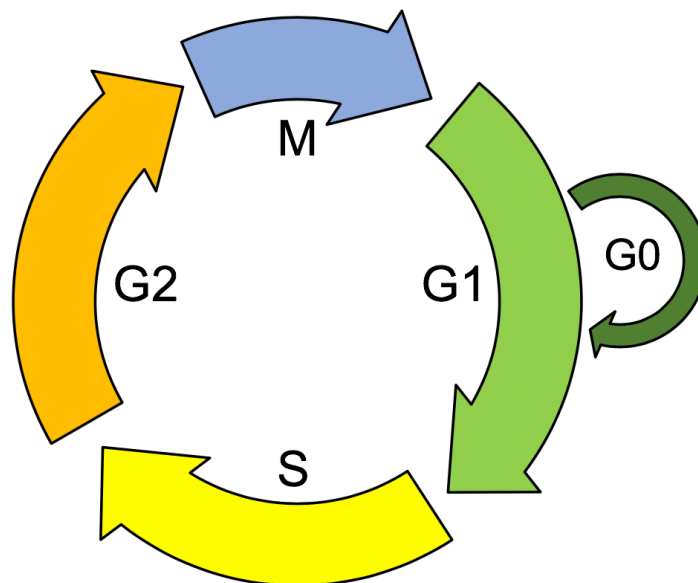


Figure 1.4 Cell cycle

Schematic showing the stages of cell cycle. Growth 1 (G1), Synthesis (S) and Growth 2 (G2) are collectively known as 'Interphase'. Cells in G1 can choose to exit cell cycle and become quiescent (G0). Quiescent cells can re-enter cell cycle into G1.

1.4 Cip1-interacting zinc finger protein 1 (CIZ1)

Human Cip1-interacting zinc finger protein (CIZ1) is a nuclear matrix protein encoded by 17 exons that are extensively alternatively spliced, with over 22 variants identified (Rahman et al., 2010). The protein can be split into two functional regions, with the N-terminus playing a role in DNA replication (Copeland et al., 2010, Coverley et al., 2005, Copeland et al., 2015), and the C-terminus in anchorage to the nuclear matrix (Ainscough et al., 2007). In this thesis, I primarily work on mouse CIZ1 (**Figure 1.5**), which is also encoded by 17 exons and shares 69% amino acid sequence identity with its human orthologue.

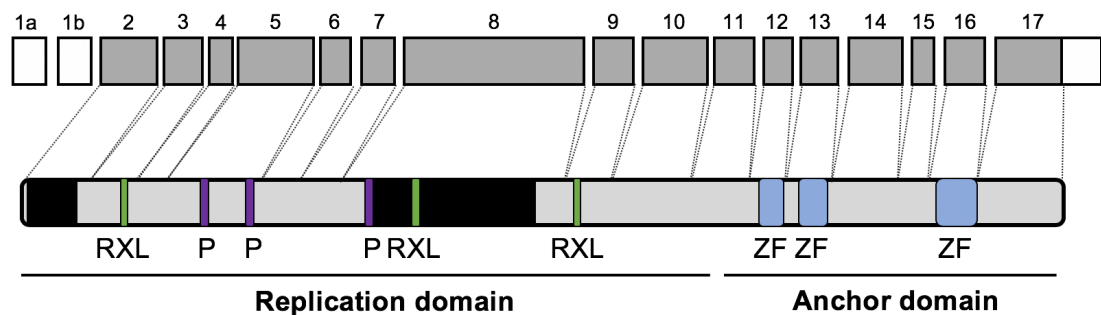


Figure 1.5 Schematic of mouse CIZ1 protein

Above, exon structure of CIZ1 mRNA. Translated regions are shaded in grey. Below, protein structure. The N terminus of the protein, also known as the replication domain, contains multiple cyclin binding motifs (RXL, green), functional phosphorylation sites (P, purple) and polyglutamine repeats (black shading). The C terminus of the protein, also known as the anchor domain, contains multiple zinc-fingers (ZF, blue).

1.4.1 Role of CIZ1 in DNA replication

CIZ1 is involved in the initiation of mammalian DNA replication (Coverley et al., 2005), and CIZ1 has been shown to directly bind to cell cycle controlling proteins such as p21 (Mitsui et al., 1999), CDK2, cyclin E and cyclin A (Copeland et al., 2010). During mid-G1 levels of cyclin E rise and it interacts with specific binding motifs in the N-terminus of CIZ1. As G1 progresses, levels of cyclin A rise and cyclin A displaces cyclin E and interacts with CIZ1 itself (Copeland et al., 2010). This lead to the proposal that CIZ1 might act as a

scaffold to support the localisation of cyclins E and A (Copeland et al., 2010). Moreover cyclin-dependent phosphorylation of CIZ1 limits this interaction so that it can only occur under specific conditions, thought to be late G1/early S phase (Copeland et al., 2015). The C-terminal domain mediated attachment to the nuclear matrix allows CIZ1 to be anchored ,and so act as a spatially restricted scaffold that could deliver cyclins sequentially to replication factories on the nuclear matrix, fine tuning control of DNA replication in a spatial and temporal manner (Ainscough et al., 2007).

1.4.2 CIZ1 and disease

Upregulation of CIZ1 has been heavily linked to human cancers, including those of the prostate, gallbladder, colon and liver (Yin et al., 2013, Wang et al., 2014, Zhang et al., 2015, Liu et al., 2015, Lei et al., 2016, Wu et al., 2016). Specific mutations in exon 7 of CIZ1 has been linked with the familial neurological disorder cervical dystonia (Xiao et al., 2012).

Alternative splicing of CIZ1 also features heavily in links with disease. Exclusion of exon 4 due to intronic mutation has been identified in human Ewing's tumour cell lines (Rahman et al., 2007) and expression of a CIZ1 variant excluding exon 8 is more highly expressed in brain tissue from Alzheimer's patients than brain tissue from healthy individuals (Dahmcke et al., 2008). Interestingly both of these exclusions were reported to result in disturbed sub-nuclear distribution of CIZ1, preventing CIZ1 nuclear matrix anchorage when expressed ectopically in mouse cell lines (Dahmcke et al., 2008, Rahman et al., 2007).

Similarly, F-variant CIZ1 has been linked with early stage common solid tumours (Swarts et al., 2018). F-variant results from another exon 8 alternative splicing event, resulting in a truncated form of CIZ1 following a short alternative reading frame (ARF) immediately following the splice site (Rahman et al., 2010), thereby removing the nuclear matrix anchor domain completely.

B-variant, arising from alternative splicing in exon 14 at the C-terminal end of CIZ1, is present in the blood plasma of patients with early stage lung cancer and is being developed as a biomarker, however whether this splicing event

is a driver or consequence of the cancer is unknown (Coverley et al., 2016, Higgins et al., 2012).

A CIZ1 knock out mouse model, targeting exon 5, revealed CIZ1 is not essential for development, as mice are viable and appear normal. However, this study found CIZ1 null mice were susceptible to developing leukaemia after retroviral insertional mutagenesis (Nishibe et al., 2013). Recently, generation of CIZ1 null mice using a gene trap following exon 1 by Justin Ainscough also generated viable mice which appear normal. However, 100% of female mice exhibited a lymphoproliferative disorder phenotype in adulthood (Ridings-Figueroa et al., 2017). My work in this thesis uses primary embryonic fibroblasts (PEFs), lymphocytes and spleen tissue derived from this second mouse model and contributed to the paper that describes this animal model (Ridings-Figueroa et al., 2017).

1.4.3 Phenotypes of CIZ1 null PEFs

Previous work in the Coverley lab analysing multiple female cell strains revealed that CIZ1 is strongly enriched at the Xi (Ridings-Figueroa et al., 2017) as well as smaller foci throughout the nucleus (Ainscough et al., 2007). However, whether CIZ1 anchorage at the Xi is the same as in the rest of the nucleus was unknown. The presence of CIZ1 at the Xi is in agreement with previous proteomic analysis of the *Xist* interactome, which identified CIZ1 as a possible *Xist* binding partner in post-differentiation ES cells (Chu et al., 2015). Direct interaction between CIZ1 and *Xist* could occur through the zinc-fingers present in the C terminus of the protein (**Figure 1.5**).

Analysis of female CIZ1 null PEFs revealed that absence of CIZ1 resulted in delocalisation of *Xist* and loss of enrichment of H2AK119Ub1 and H3K27me3 from the Xi (**Figure 1.6**, (Ridings-Figueroa et al., 2017)). However, if CIZ1 was essential in *Xist* localisation and recruitment of PRC, it would be expected that XCI would fail in CIZ1 null cells and therefore CIZ1 loss would result in a strong embryonic phenotype, which is not what is seen. The answer to this may lie in the observation that CIZ1 binding to *Xist* was specifically detected in the post-differentiation state, and therefore CIZ1 may not be required for the initiation

and spreading phases of XCI and is instead involved in the maintenance phase.

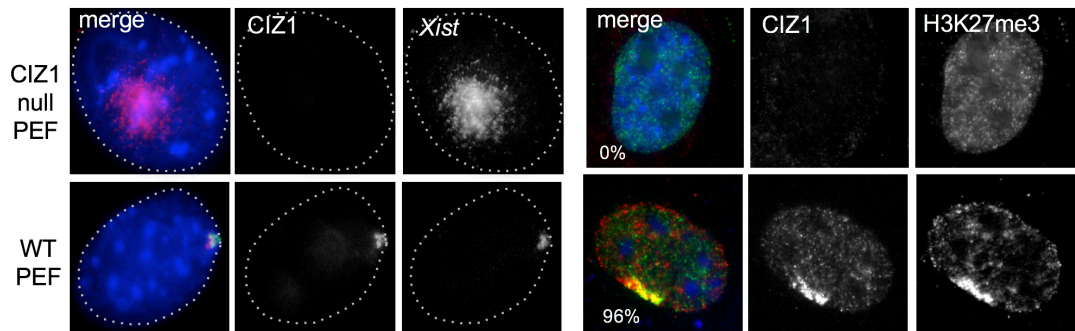


Figure 1.6 Phenotypes of CIZ1 null PEFs

Figure taken from (Ridings-Figueroa et al., 2017), showing loss of Xist localization and H3K27me3 enrichment from the Xi in CIZ1 null PEFs.

1.5 Project Aims

The goal of this thesis is to shed light on the role of CIZ1 in transcriptional silencing, using the inactive X chromosome as a model. A further goal was to investigate the cause of the female-specific lymphoproliferative disorder phenotype. Specifically, my lines of investigation aimed to address the following questions:

- 1) Is CIZ1 at the inactive X chromosome part of the nuclear matrix?
- 2) Does loss of CIZ1 have an effect on gene expression in cultured primary embryonic fibroblasts?
- 3) Are CIZ1 null cell lines a suitable model for functional investigations?
- 4) Does loss of CIZ1 have an effect on lymphocyte activation? If so, could these effects identify avenues of investigation into the cause of the female-specific lymphoproliferative disorder?

The most unexpected outcome is mechanistic insight into how template and chromatin modifying enzymes meet (specifically PRC2 subunit EZH2 and Xi chromatin) described in (Chapter 5).

2.0 Materials and Methods

Text and tables contained within this chapter are my contributions to an item of published work, which has been reproduced in full here (in **Appendix D**) (Stewart et al., 2019).

2.1 Cell Culture

2.1.1 Cell Isolation

CIZ1 null mice were generated by Justin Ainscough from C57BL/6 ES clone IST13830B6 (TIGM) harbouring a neomycin resistance gene trap inserted downstream of exon 1. Absence of *Ciz1*/CIZ1 in homozygous progeny was confirmed by quantitative polymerase chain reaction (qPCR), immunofluorescence and immunoblot (Ridings-Figueroa et al., 2017). Mouse Primary Embryonic Fibroblasts (PEFs) were derived from individual embryos at day 13-14 of gestation by Justin Ainscough at the University of Leeds.

Embryos were dissected in ice-cold phosphate buffered saline (PBS) to remove the head and visceral organs. The resulting carcass was disrupted by multiple passages through an 18G needle in cell culture media (DMEM GlutaMAX (Gibco) supplemented with 10% Foetal Bovine Serum (FBS, PAA) and 1% Pen/Strep/Glutamine (Gibco)). Each embryo was plated into a single well of a 6-well dish containing 2 ml of cell culture media, and cells were allowed to settle for 16-24 hours at 37°C, 5% CO₂ before washing with PBS to remove residual tissue, followed by replacement of 2 ml cell culture media. The attached cells were incubated for a further 2 days before growing colonies were passaged at a 1:5 ratio to form a uniform monolayer. These cells were designated passage (p) 1 and were frozen in cell freezing media (10% dimethylsulfoxide (DMSO), 10% FBS, 80% cell culture media) and stored in liquid nitrogen. Frozen p1 cells were transported to the University of York on dry ice and stored in liquid nitrogen until use.

2.1.2 Cell culture

Frozen stocks of PEFs and culture derived cell lines were thawed, diluted 10-fold with cell culture media (DMEM GlutaMAX (Gibco) supplemented with 10% FBS (PAA) and 1% Pen/Strep/Glutamine (Gibco)) and plated at high density (approximately 90% confluency). Cells were allowed to settle for 8-16 hours at 37°C, 5% CO₂ before media was replaced with fresh cell culture media. All primary and culture adapted mouse embryonic fibroblasts were cultured in DMEM GlutaMAX (Gibco) supplemented with 10% FBS (PAA) and 1% Pen/Strep/Glutamine (Gibco). During rapid proliferation phase cells were grown to 90% confluence then passaged 1:3 every 2-3 days. Populations of cells cultured up to passage (p) 5 were considered early passage and designated here as primary cells. Experiments were not performed after passage 5, unless to generate culture adapted lines. For inducible cells harbouring transactivator and responder transgenes, addition of doxycycline to media (5-10 µg/ml) was used to induce N-terminal Green Fluorescent Protein (GFP) tagged full-length CIZ1 (GFP-CIZ1) expression within 6 hours. Female 3T3 cell line D001 (Coverley et al., 2002) (a kind gift from Stephen Downes) were grown in the same media. Female breast cancer cell line MCF-7 were grown in MEM supplemented with 10% FBS (Sigma), 1% Pen/Strep/Glutamine (Gibco) and 0.01 mg/ml human recombinant insulin. All cell lines used are given in **Table 2.1**.

2.1.3 Derivation of cell lines

To generate culture adapted lines, beyond passage 5, cells were supplemented with fresh media every 3 days and passaged at 90% confluence (typically once per week as proliferation rate slowed) without dilution (passage ratio 1:1). By passage 10-12, individual colonies of intermediate state cells emerged. These colonies were dissociated and re-plated as a whole (1:1) until rapidly proliferating culture adapted populations emerged by passage 20 and were designated as culture adapted lines. These were routinely grown to 90% confluence then passaged as for primary cells. The E14.2 and E14.4 culture adapted lines were generated by me. The E13.1, E13.8, E13.15 and E13.17 culture adapted cell lines were generated by Justin Ainscough.

Table 2.1 List of primary cell strains and cell lines

During my PhD I have cultured all of the primary cell strains and culture adapted cell lines in this table. The cell lines from which I isolated RNA for RNA-sequencing (**Chapter 2.6.1**) are shaded grey.

Name	Sex	Type	CIZ1 genotype	Reference
E13.1	Female	Primary embryonic fibroblast	+/+	(Ridings-Figueroa et al., 2017)
E13.8	Female	Primary embryonic fibroblast	+/+	(Ridings-Figueroa et al., 2017)
E14.4	Female	Primary embryonic fibroblast	+/+	(Ridings-Figueroa et al., 2017)
E13.15	Female	Primary embryonic fibroblast	-/-	(Ridings-Figueroa et al., 2017)
E13.17	Female	Primary embryonic fibroblast	-/- with inducible CIZ1 transgene	(Ridings-Figueroa et al., 2017)
E14.2	Female	Primary embryonic fibroblast	-/-	(Ridings-Figueroa et al., 2017)
E14.19	Female	Primary embryonic fibroblast	-/- with inducible CIZ1 transgene	(Stewart et al., 2019)
E13.1	Female	Adapted fibroblast cell line	+/+	(Stewart et al., 2019)
E13.8	Female	Adapted fibroblast cell line	+/+	(Stewart et al., 2019)
E14.4	Female	Adapted fibroblast cell line	+/+	(Stewart et al., 2019)
E13.15	Female	Adapted fibroblast cell line	-/-	(Stewart et al., 2019)
E13.17	Female	Adapted fibroblast cell line	-/- with inducible CIZ1 transgene	(Stewart et al., 2019)
E14.2	Female	Adapted fibroblast cell line	-/-	(Stewart et al., 2019)
D001 3T3	Female	Adapted murine fibroblast	+/+	(Coverley et al., 2002)
MCF-7	Female	Human breast cancer cell line	+/+	(Swarts et al., 2018)

2.1.4 Ethics

All work with animal models is compliant with UK ethical regulations. Breeding and genetic modification of mice was carried out under UK Home Office license and with approval of the Animal Welfare and Ethical Review Body at the University of Leeds. Analysis on cells and tissues derived from these mice was carried out with approval of the Animal Welfare and Ethical Review Body at the University of York.

2.2 Sub-nuclear fractionation

All buffers were made in RNase free water. All reagents were from Sigma unless otherwise stated. All samples are named by the last treatment in the series that they received. Extraction for analysis by western blot (Wilson et al., 2016b) and extraction for analysis by immunofluorescence microscopy (Stewart and Coverley, 2018) was performed using the same buffers.

2.2.1 Extraction for analysis by western blot

Plates of adherent cells were washed twice in ice cold 1X PBS then twice in ice cold cytoskeletal buffer (CSK: 10 mM 1,4-piperazinediethanesulfonic acid (PIPES)/KOH pH 6.8, 100 mM sodium chloride (NaCl), 300 mM sucrose, 1 mM ethylene-bis(oxyethylenitrilo)tetraacetic acid (EGTA), 1 mM magnesium chloride ($MgCl_2$), 1 mM dithiothreitol (DTT), 1 cOmplete™ Protease Inhibitor Cocktail per 50 ml) and drained at a shallow angle on ice for 2 minutes. Cells were scrape harvested and supplemented with 2 mM phenylmethylsulfonyl fluoride (PMSF) before addition of Triton-X-100 to 0.1% and, to samples not subsequently receiving RNase, vanadyl ribonucleoside complex (VRC) to 2.5 mM (NEB). Samples were mixed by pipetting, and after 1 minute on ice, centrifuged at 1000 xg. The soluble fraction from this stage was designated detergent (det.) supernatant (S, sample i). The pellet fraction was either denatured for analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; det. P, ii) or resuspended in 0.1% Triton-X-100 and 0.5 M NaCl in CSK and incubated on ice for 1 minute before a second centrifugation at 1000 xg (salt S, iii, and P, iv). Pellets to be further extracted were washed with digestion buffer (40 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris)/HCl, 10 mM NaCl, 6 mM $MgCl_2$, 1mM calcium chloride ($CaCl_2$), pH 7.9, supplemented 1/500 with RNaseOUT for RNase free samples), before resuspension in fresh digestion buffer. Samples were then incubated at 30°C for 1 hour with gentle agitation in the presence of DNase 1 (Roche 04716728001) at 0.3 U/ μ l, RNase (Roche 11119915001) 0.5 U/ μ l, both DNase1 and RNase, or no enzyme (mock). Before final centrifugation, reactions were supplemented to 0.5 M NaCl for 5 min, then separated to yield

pellets (P, vi and viii) and supernatants (S, v and vii). Pellets were resuspended in 1X denaturing buffer (2% SDS, 15% glycerol, 1.7% betamercaptoethanol, 75 mM Tris pH 6.8 with bromophenol blue), and supernatants supplemented with 4X denaturing buffer, then heated to 95°C with repeated vortexing to shear remaining nucleic acid.

2.2.2 Extraction for analysis by immunofluorescence microscopy

Cells grown on 13 mm coverslips were incubated with CSK supplemented with 0.1% Triton-X-100 and 2.5 mM VRC (CSK-D) for 1 minute. Following CSK-D removal, coverslips were either fixed with 4% paraformaldehyde (PFA; detergent sample) or incubated with CSK-D with 0.5 M NaCl (CSK-DS) for 1 minute. For coverslips to be treated with RNase, VRC was left out from this step onwards. Coverslips were washed twice for 1 minute with digestion buffer (supplemented with 2.5 mM VRC for non-RNase samples) before incubation for 1 hour at 37°C in digestion buffer with 2.5 mM VRC (High Salt sample), digestion buffer supplemented with 2.5 mM VRC and DNase1 at 0.45 U/μl (DNase sample) or digestion buffer supplemented with RNase at 0.75 U/μl (RNase sample). Coverslips were incubated for 1 minute with CSK-DS (supplemented with 2.5 mM VRC for non-RNase samples) before fixation with 4% PFA.

2.3 Antibodies and detection protocols

2.3.1 Western Blot

For western blots, adherent cells were either extracted as above or, after washing in ice cold PBS, scrape harvested into denaturing buffer with fresh 2 mM PMSF to generate whole cell lysates. Samples were heated to 95°C and separated through 4-15% gradient gels (Bio-Rad) then transferred onto nitrocellulose membranes using the iBlot system (Invitrogen). Blots were blocked in either 5% low fat milk in PBS with 0.1% Tween-20, or 5% bovine serum albumin (BSA) in tris buffered saline (TBS) with 0.05% Tween-20, depending on antibody, then incubated with primary antibodies (**Table 2.2**) overnight at 4°C with gentle agitation. Blots were washed and probed with

horseradish peroxidase (HRP)-conjugated anti-species secondary antibodies for 1 hour at room temperature (Jackson Immunochemicals 115-035-174 and 211-032-171), and imaged using EZ-ECL (Biological Industries) with a Syngene PXi chemiluminescence imaging system. Band intensities were quantified using GeneSys 4.03.05.0 software.

2.3.2 Immunofluorescence Microscopy

For immunofluorescence (IF) microscopy, cells grown on glass coverslips were washed with PBS before fixation in 4% PFA, or where indicated bathed in CSK with 0.1% Triton-X-100 for 1 minute prior to fixation. After fixation cells were rinsed twice with PBS then incubated for 15 minutes in BSA Antibody buffer (0.02% SDS, 0.1% Triton-X-100, 10 mg/ml nuclease-free BSA in PBS), followed by 2 hours at 37°C with primary antibody (in BSA antibody buffer). After three washes in the same buffer, anti-species secondary antibodies (Alexa Fluor 488 or 568) were applied for one hour, followed by three further washes, and mounting in VectorShield with 4',6-diamidino-2-phenylindole (DAPI; Vector Labs). All antibodies including dilutions are listed in (**Table 2.2**).

Table 2.2 Primary antibodies used for western blot and immunofluorescence.

Antibody	Concentration		Western Blocking Buffer	Company/Reference
	Western	IF		
CIZ1	1:1000	1:1000	PBS	1794(Coverley et al., 2005)
CIZ1		1:20		hC221a
CIZ1		1:50		Bella
CIZ1		1:50		F-variant
Fibrillarin		1:100		ab4566, Abcam
H2AK119Ub1		1:10000		D27C4, CST
H3K27me3		1:2000		ab6002, Abcam
H3K27me3		1:2000		CST
EZH2	1:1000	1:250	TBS	D2C9, CST
EZH2	1:1000		TBS	ab191080, Abcam
EZH2	1:1000	1:100	TBS	STJ112944, St John's Laboratory
H3	1:10000		PBS	ab1791, Abcam
SAF-A		1:50		ab10297, Abcam
YY1		1:50		SC7341, Santa Cruz
PTBP		1:1000		ab5642, Abcam

2.3.3 Proximity Ligation Assay (PLA)

Primary WT, primary CIZ1 null and culture adapted WT murine embryonic fibroblast cells were grown on glass coverslips and washed with 0.1% Triton-X-100 containing PBS prior to fixation with 4% PFA for 15 minutes. Coverslips were blocked and incubated with primary antibodies (**Table 2.2**, same concentration as IF) before washing with Duolink® In Situ Wash Buffer A and the proximity ligation assay was performed as detailed in Duolink® PLA fluorescence protocol (Sigma), using species specific secondary antibody probes and green detection reagents. In brief, following standard primary antibody incubation as detailed in antibodies and detection protocols (**Chapter 2.3.2** and **Table 2.2**), coverslips were washed twice for 5 minutes with Wash Buffer A and incubated with species-specific PLUS and MINUS PLA probes in the provided antibody diluent for 1 hour at 37°C. Coverslips were washed twice for 5 minutes with Wash Buffer A before incubation for 30 minutes at 37°C with 1X Ligation Buffer supplemented with DNA Ligase. Coverslips were washed twice for 5 minutes with Wash Buffer A followed by incubation for 100 minutes at 37°C in Amplification buffer supplemented with DNA polymerase. Coverslips were washed twice for 10 minutes with Wash Buffer B before a final 1 minute wash in 0.01X Wash Buffer B and mounting in Vector Shield with DAPI (Vector Labs).

2.3.4 S phase labelling

5-ethynyl-2'-deoxyuridine (EdU, 10µM) was added to adherent cells on coverslips at ~70% confluence for a 30 minute pulse period under standard growth conditions. For pulse/chase experiments, coverslips were transferred to warm PBS then into fresh media (without EdU) for 30-60 minutes. To visualize newly synthesized DNA, coverslips were washed briefly in CSK with 0.1% Triton-X-100, or subjected to extraction up to the RNase treatment step before fixation with 4% PFA for 15 minutes. Coverslips were then washed in PBS and EdU detected using the Click-iT® EdU Alexa Fluor® 594 kit (ThermoFisher), as recommended. Briefly, coverslips were blocked with 3% BSA before incubation in a light-proof humidified chamber with Click-iT® reaction cocktail for 60 minutes. For dual staining (e.g. CIZ1 or fibrillarin),

coverslips were first incubated with primary antibody as described under antibodies and detection protocols (**Table 2.2**) before blocking with 3% BSA. Anti-species secondary antibody diluted in Click-iT® reaction buffer was included in the EdU detection step. Coverslips were then washed and mounted using VectorShield with DAPI (Vector Labs).

2.4 Imaging

2.4.1 Image Capture

Fluorescence images were captured using a Zeiss Axiovert 200M fitted with a 63X/1.40 Plan-Apochromat objective and Zeiss filter sets 2, 10, 15 (G365 FT395 LP420, BP450-490 FT510 BP515-565, BP546/12 FT580 LP590), using AxioCam 506 mono and Axiovision image acquisition software (SE64 release 4.9.1).

2.4.2 Image Processing

Where changes in fluorescence intensity are quantified across an extraction series, coverslips were imaged as a set with all images for each filter set captured with the same exposure time. Images were saved at 1499 by 1205 pixels in tagged image file format for downstream analysis. For EZH2, at least 50 individual nuclei from unmodified images were quantified for each cell population and each extraction condition, using Fiji (Schindelin et al., 2012). Where indicated, EZH2 levels across an extraction series were normalized to the detergent-treated sample (given an arbitrary value of 100) to enable comparison between cell lines. For presentation, images were enhanced using Adobe Photoshop CS4 or Affinity Photo 1.5.2, maintaining identical manipulations across extraction series so that image intensities reflect actual relationships (unless specifically stated otherwise). All quantification of image intensity was carried out prior to manipulation. For presentation of images illustrating positional, rather than intensity information, images were not necessarily modified identically.

2.5 Statistical Analysis

Experiments were designed to use the minimum number of animals while achieving statistically valid data and include two types of analysis. For multivariate data (lists of genes or proteins that are changed in one genotype compared to another) three biological replicates (independent PEF lines) enables calculation of average log 2 fold change and T-test p-values, as well as false detection rate (FDR)-adjusted p-value for each data point (gene), and is the minimum required for differential expression analysis using Cuffdiff. Randomization is not appropriate, and PEF lines are matched as closely as possible. Unless indicated, data is represented as means with standard error of the mean (SEM).

2.6 Transcriptome analysis

2.6.1 RNA Isolation

Primary (before passage 5) and culture-adapted derivative cell lines of murine embryonic fibroblasts 13.1, 13.8 and 14.4 (female WT), and 13.15, 13.17, 14.2 and 14.19 (female CIZ1 null) were grown to 80% confluence. RNA was extracted with TRIzol (Ambion 15596-026) following manufacturer's instructions. Briefly, adherent cells grown to 80% confluence were washed twice with PBS, drained on a shallow angle for 2 minutes and excess PBS removed. 1mL of TRIzol was added per 28 cm² and incubated for 3-5 minutes at room temperature with periodic agitation. Lysates were collected in clean eppendorff tubes. Chloroform was added to a ratio of 1:5 and lysates were shaken vigorously for 15 seconds before incubation at room temperature for 3 minutes, followed by centrifugation at 12,000 xg for 15 minutes. The aqueous phase was transferred into a clean eppendorff and mixed with equal volume of isopropanol through gentle inversion and incubated at room temperature for 10 minutes, followed by a 10 minute centrifugation at 12,000 xg. Supernatant was removed, and the RNA pellet washed with an equal volume of 75% ethanol to the volume of TRIzol used to extract the RNA. Sample was centrifuged for 5 minutes at 12,000 xg and the supernatant removed. RNA pellet was resuspended in nuclease free water. Isolated RNA was then treated

with DNase (Roche 04716728001), before quality analysis by agarose gel, NanoDrop spectrophotometer and Agilent 2100 Bioanalyzer.

2.6.2 Sample Preparation and Sequencing

Libraries were prepared by Sally James in the University of York Genomics Facility with NEBNext® Ultra™ RNA library Prep Kit for Illumina®, and enriched for mRNA using NEBNext Poly(A) mRNA Magnetic Isolation Module, which is optimized for production of libraries with 250-400 bp inserts. Enriched mRNA was fragmented by heating to 95°C for 12 minutes, cDNA synthesised from random primers, followed by end repair, dA-tailing, adaptor ligation and PCR enrichment. Libraries were sequenced at the Leeds Institute for Molecular Medicine (LIMM) using Illumina HiSeq 3000 system, using paired-end sequencing to generate ~50 million reads per sample.

2.6.3 Transcriptome Assembly

Transcript assembly was performed by Katherine Newling. Sequence reads were trimmed to remove any adapter sequences using Cutadapt version 1.8.3 (Martin, 2012) then aligned to version GRCm38 of the mouse genome using HISAT2 (Kim et al., 2015). Transcriptomes were assembled and gene expression quantified using the Tuxedo pipeline (version 2.2.1) (Trapnell et al., 2012), from which predicted splice variant assemblies were extracted, including those that map to the EZH2 locus 6:47530040-47595351 and other PRC subunits. Average expression of each splice variant was calculated from the 3 independent lines for WT and CIZ1 null primary and culture adapted cells. Variants arising from sequencing of the opposite strand were removed. Total expression of the transcript was calculated for each cell type along with the percentage of each splice variant. Splice variants contributing less than 5% of the total in all cell types were removed from the analysis to simplify visual representation. Cufflinks was used to assemble transcriptomes for each sample using the GTF annotation file for the GRCm38 mouse genome (*C57BL/6*), followed by Cuffmerge to merge individual sample transcriptomes. Quantification, normalisation and differential expression was carried out using Cuffquant, Cuffnorm, and Cuffdiff, respectively.

2.6.4 Data Analysis

Gene Set Enrichment Analysis (Subramanian et al., 2005, Mootha et al., 2003) was performed in Python 3.6 using one-sided Fisher's Exact tests as part of the SciPy library (v.0.19.0). False discovery rate (FDR) was controlled using the Benjamini-Hochberg method in the StatsModels library (v.0.8.0), to generate q values. As this correction is based on p-values and different comparisons have different numbers of p-significant expression changes, the threshold at which q-significance is reached will vary between comparisons. Volcano plots were generated in Excel. Heat maps were generated using Spyder (v.3.1.4), accessed via Anaconda Distribution (v.1.6.2), using the pandas, seaborn and matplotlib modules. Transcription units which did not have a numerical value for $\log_2(\text{fold change})$ due to mean expression of 0 in one condition were manually removed before generating the plots. Individual fragment counts per kilobase per million (FPKM) were extracted for replicate cell lines to calculate means and SEM. Examples of the custom code used to perform Gene Set Enrichment Analysis and generate heatmaps are included as (**Appendix E**).

2.6.5 Data storage and availability

Transcriptome data is available at GEO repository under accession code GSE122235

[\[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122235\]](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122235)

3.0 Visualisation of masked epitopes after nuclear matrix extraction

Text and figures contained within this chapter are an amalgamation of my contributions to three items of published work which have been reproduced here (**Appendix A, B, C**) (Ridings-Figueroa et al., 2017, Swarts et al., 2018, Stewart and Coverley, 2018).

3.1 Introduction

The nuclear matrix (NM) is biochemically defined as the insoluble nuclear substructure that remains following extraction of soluble proteins, lipids, chromatin and, in some protocols, RNA (Verheijen et al., 1988).

Enrichment of nuclear matrix proteins that are known to be involved in X chromosome inactivation (XCI) is sometimes only seen in the location of the inactive X chromosome (Xi) when using fluorescently tagged exogenously expressed forms. This is true for hnRNP U (SAF-A), a NM protein that binds to both RNA and DNA that is required for XCI and mitotic spindle organisation (Roshon and Ruley, 2005, Hasegawa et al., 2010, Ma et al., 2011). Endogenous SAF-A is not apparent within the Xi territory in fixed whole cells, but is revealed if cells are treated to remove chromatin and associated proteins before fixation (Ridings-Figueroa et al., 2017).

This suggests epitope masking is a particular issue for nuclear matrix proteins, and tells us that the fraction of the protein that we can see following typical immunofluorescence strategies may not be the fraction of the protein that is carrying out the function we want to study.

This chapter describes the application of a nuclear matrix extraction protocol which allows unmasking of previously undetected epitopes both nucleus-wide and at the Xi. It also details refinements that I developed to enable accurate distinction between RNA-dependent and RNA-independent nuclear matrix proteins.

3.2 Aims

The aims of these experiments were to ask the following questions:

- Is CIZ1 part of the nuclear matrix at the inactive X chromosome?
- Are CIZ1 splice variants differentially attached to the nuclear matrix?
- Are different splice variants of CIZ1 expressed depending on the differentiation state of the cell?

3.3 Experimental Design

Cells were serially extracted as described in (**Chapter 2.2, Figure 3.1**) followed by immunodetection by either immunofluorescence (**Chapter 2.3.2**) or western blot (**Chapter 2.3.1**). Where detection by immunofluorescence was used, images from the same channel across the series were taken at the same exposures to allow direct comparison of intensity. Where detection was performed by western blot equal volumes were loaded in all lanes and H3 was used as both a loading control and to evaluate efficacy of the DNase I treatment to release histone into the soluble fraction.



Figure 3.1 Overview of the nuclear matrix extraction protocol

Schematic of the NM extraction protocol (Ridings-Figueroa et al., 2017). Unextracted cells are serially treated to remove soluble proteins and lipids with detergent containing buffers, loosely attached proteins with high salt containing buffers and chromatin or RNA bound factors with nucleases, to leave behind the NM. Absence of RNase reveals the RNA-protein NM, and its inclusion the core protein NM (Stewart and Coverley, 2018).

3.4 Results and Discussion

3.4.1 Nuclear matrix interactions

Prior work in the Coverley group demonstrated that nuclear matrix (NM) protein CIZ1 is present at the inactive X chromosome (Xi) through co-localisation with *Xist* and H3K27me3 (Ridings-Figueroa et al., 2017). However, whether CIZ1 at the Xi territory is part of the NM remained unknown.

NM extraction can be used to determine what entities proteins are attached to in different locations within the nucleus. NM extraction followed by immunostaining for CIZ1 (antibody 1794) in female mouse 3T3 cells (**Figure 3.2A,B**) shows that, while the nucleus-wide speckles are resistant to all extraction methods, CIZ1 at the Xi (Xi-CIZ1) is sensitive to extraction with RNase in all except 3% of cells. This identifies Xi-CIZ1 as part of the RNA-protein NM, consistent with its dependence on, and functional interaction with the lncRNA *Xist* (Ridings-Figueroa et al., 2017, Sunwoo et al., 2017). However, nucleus-wide CIZ1 foci are part of the core protein NM because they are not removed along with RNA.

As an addition to the protocol, cells can be treated with the protein-protein specific cross-linker DTSP prior to extraction. Its specificity permits the required extraction of nucleotides while cross-linking proteins in close proximity to each other. Following pre-treatment with DTSP, Xi-CIZ1 is now retained following treatment with RNase in almost all cells (**Figure 3.2C**). Thus, though Xi-CIZ1 is bound to RNA, it is also in close proximity to a protein fraction that we refer to as the core protein NM.

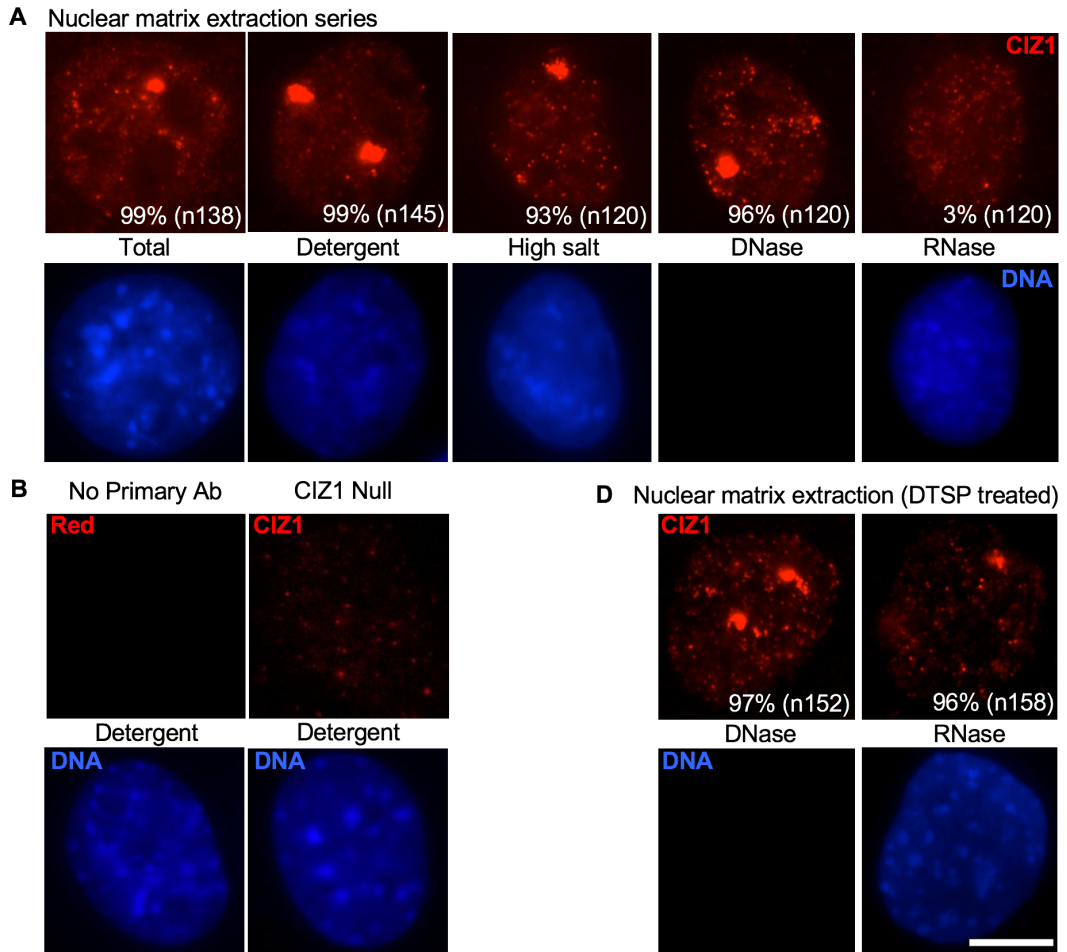


Figure 3.2 RNA-bound CIZ1 at the inactive X chromosome is cross-linkable to the core protein NM (Ridings-Figueroa et al., 2017)

(A) Images show CIZ1 (red, 1794 antibody) after serial extraction of 3T3 cells. The proportion of cells with discrete CIZ1-Xi domains is indicated ($n > 100$ for each condition), some with two domains, indicating that a proportion of cells is tetraploids with two Xis. DNA (blue) shows the extent of nuclease treatment, scale bar 10 microns. (B) Negative controls. Left, as in (A) Detergent sample without incubation with CIZ1 primary antibody (1794). Right, as in (A) Detergent sample, but for CIZ1 null culture adapted cells. (C) As in (A) but with prior protein-protein cross-linking with DTSP.

3.4.2 Nucleus Wide Superficial and Buried Populations

To identify whether CIZ1 splice variants are differentially attached to the NM, further analysis was performed on human MCF-7 cells. These are a human breast cancer cell line which express CIZ1 F-variant, a human and disease specific splice variant (Swarts et al., 2018). This splice variant arises from an alternative splicing event in exon 8 resulting in a unique alternative reading frame (ARF), allowing generation of a variant specific antibody.

NM extraction was carried out on MCF-7 cells followed by immunodetection by immunofluorescence against three different CIZ1 epitopes (**Figure 3.4A**). Bella and 1794 recognise epitopes that could be found in many CIZ1 splice variants, while the F-variant antibody only recognises that specific splice variant.

The average staining intensity per antibody per nucleus was measured for each stage of the extraction process. Intensity measures were made using on unmodified images which were taken at the same exposure across a series (**Figure 3.4B**). The same process was performed on images of the Hoechst DNA stained nuclei.

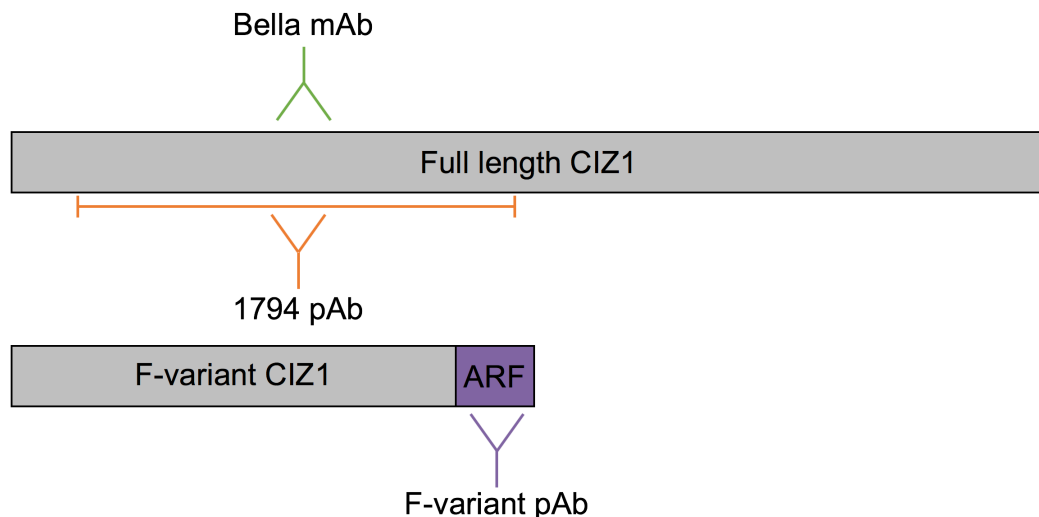


Figure 3.3 CIZ1 Antibody Map

Map of full length CIZ1 and a CIZ1 splice variant known as F-variant (Swarts et al., 2018), including the location of the epitope recognised by three different CIZ1 antibodies.

Significant decrease in DNA intensity following extraction with either DNase alone or both DNase and RNase demonstrates effective removal of chromatin (**Figure 3.4B**). Conversely, intensity of CIZ1 staining under the same condition significantly increased with all antibodies ($p < 0.001$, t-test), though to varying levels. While the fluorescence intensity of 1794 doubles following DNase treatment compared to following detergent treatment, the fluorescence intensity of the Bella antibody increases over 16-fold. Interestingly, this increase is significantly reduced when cells are treated with RNase in addition to DNase, implying that the population masked by chromatin is almost entirely RNA bound.

This implies that although CIZ1 is in general masked by chromatin, the epitope the Bella antibody recognises is masked to a much greater extent than the epitope recognised by 1794. This could be because Bella is a monoclonal antibody raised specifically against the exon7/8 junction, while 1794 is a polyclonal antibody raised against a large portion of the N terminus of the protein. Therefore we could expect 1794 can recognise a larger proportion of both the CIZ1 protein itself and more splice variants.

Alternatively, in the context of RNA-protein NM bound CIZ1, the exon 7/8 boundary could be closer to chromatin than other N terminal sequences. Both of these could result in the lower Bella antibody signal intensity prior to chromatin removal, while 1794 is able to recognise and bind to other less buried epitopes throughout.

Though the difference in fluorescence intensity following DNase treatment is the largest, there are additional features worth noting. For example, while 1794 and Bella both display an increase in fluorescence following high salt extraction, F-variant exhibits a signal drop ($p < 0.001$, t-test). This suggests there is a sub-population of F-variant protein that is loosely bound rather than incorporated in to the nuclear matrix fractions.

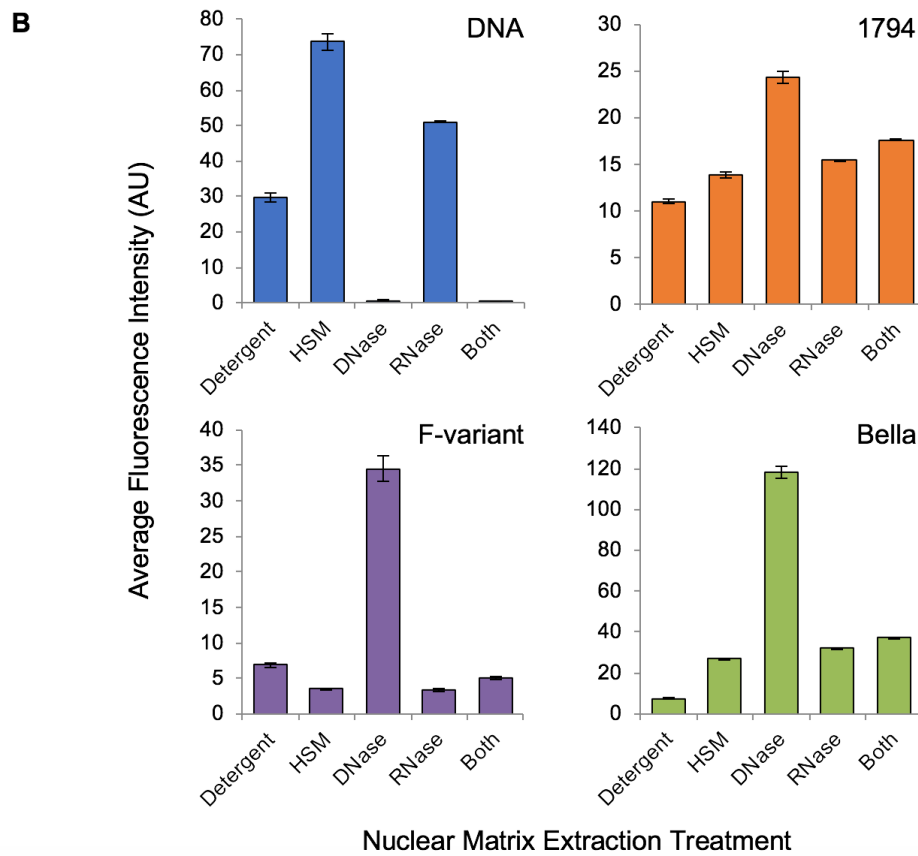
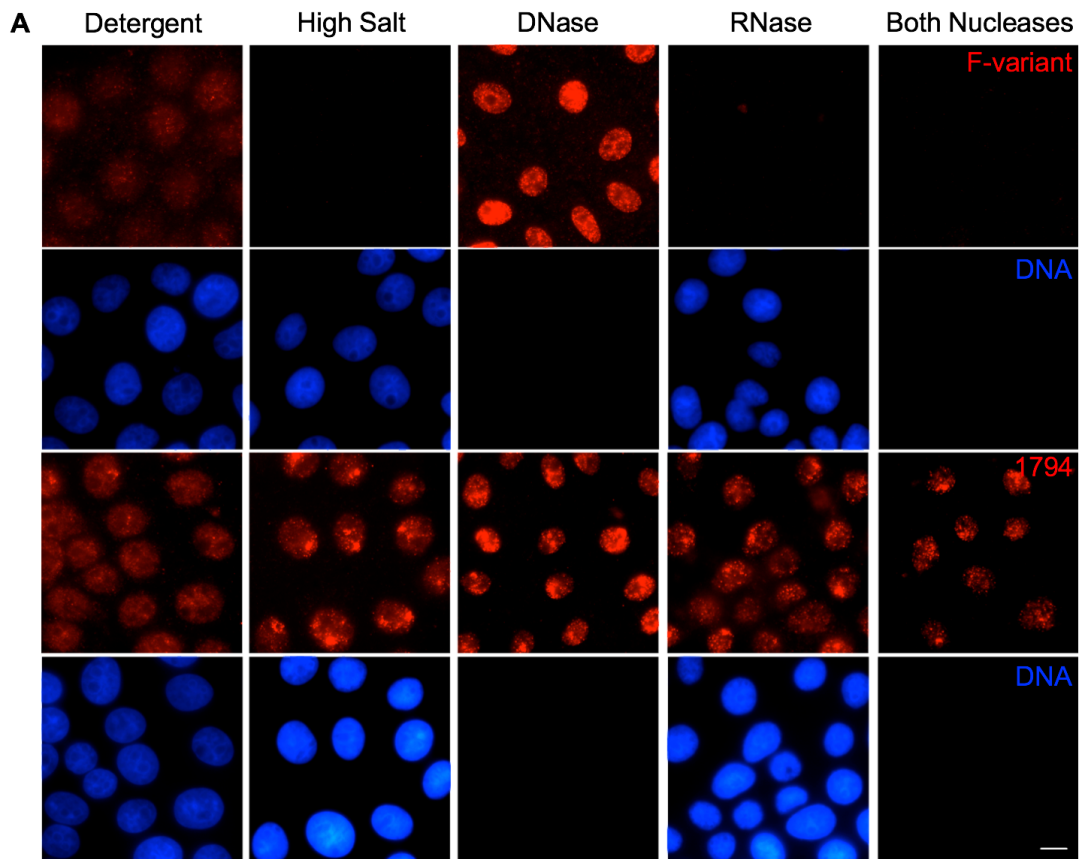


Figure 3.4 Reveal of previously masked epitopes following chromatin removal (Swarts et al., 2018)

(A) Example images of a nuclear matrix extraction in MCF-7 cells following immunostaining for CIZ1 in red (F-variant upper, 1794 lower) and DNA in blue, scale bar 10 microns. (B) Quantification of average fluorescence intensity (arbitrary units, AU) of three CIZ1 antibodies and DNA during nuclear matrix extraction, error bars SEM.

3.4.3 Revealing masked epitopes at the Xi

As this protocol can reveal nucleus-wide masked epitopes of NM proteins, whether this protocol also reveals the epitopes of NM proteins known to be masked at the Xi was investigated (Helbig and Fackelmayer, 2003). NM extraction was carried out on mouse embryonic fibroblasts (MEFs) followed by immunostaining for two NM proteins known to be involved in XCI, SAF-A (also known as hnRNP U) (Hasegawa et al., 2010) and YY1 (Jeon and Lee, 2011) (**Figure 3.5**).

As seen for CIZ1, signal intensity increased following removal of chromatin but there was no obvious enrichment at the Xi territory with either antibody at any stage of NM extraction (**Figure 3.5**). However there were obvious differences between the two proteins. Following detergent treatment, YY1 signal increased, likely due to increased antibody accessibility due to membrane disruption (**Figure 3.5A**). Conversely, 22% of nuclei displayed a bright SAF-A signal in unextracted cells, but this was removed following detergent wash, implying this population is highly soluble (**Figure 3.5B**).

Line intensity scans of nuclei co-immunostained for CIZ1 to reveal Xi, and for SAF-A revealed that SAF-A actually appears to be excluded from the Xi territory in the absence of any extraction, but is no longer excluded following removal of chromatin (**Figure 3.6A**). The same analysis on another NM protein PTBP also revealed lower signal intensity at the location of the Xi in the absence of extraction, but appears enriched at the Xi following chromatin removal (**Figure 3.6B**). Therefore this protocol is allowing visualisation of NM proteins in the location of the inactive X chromosome.

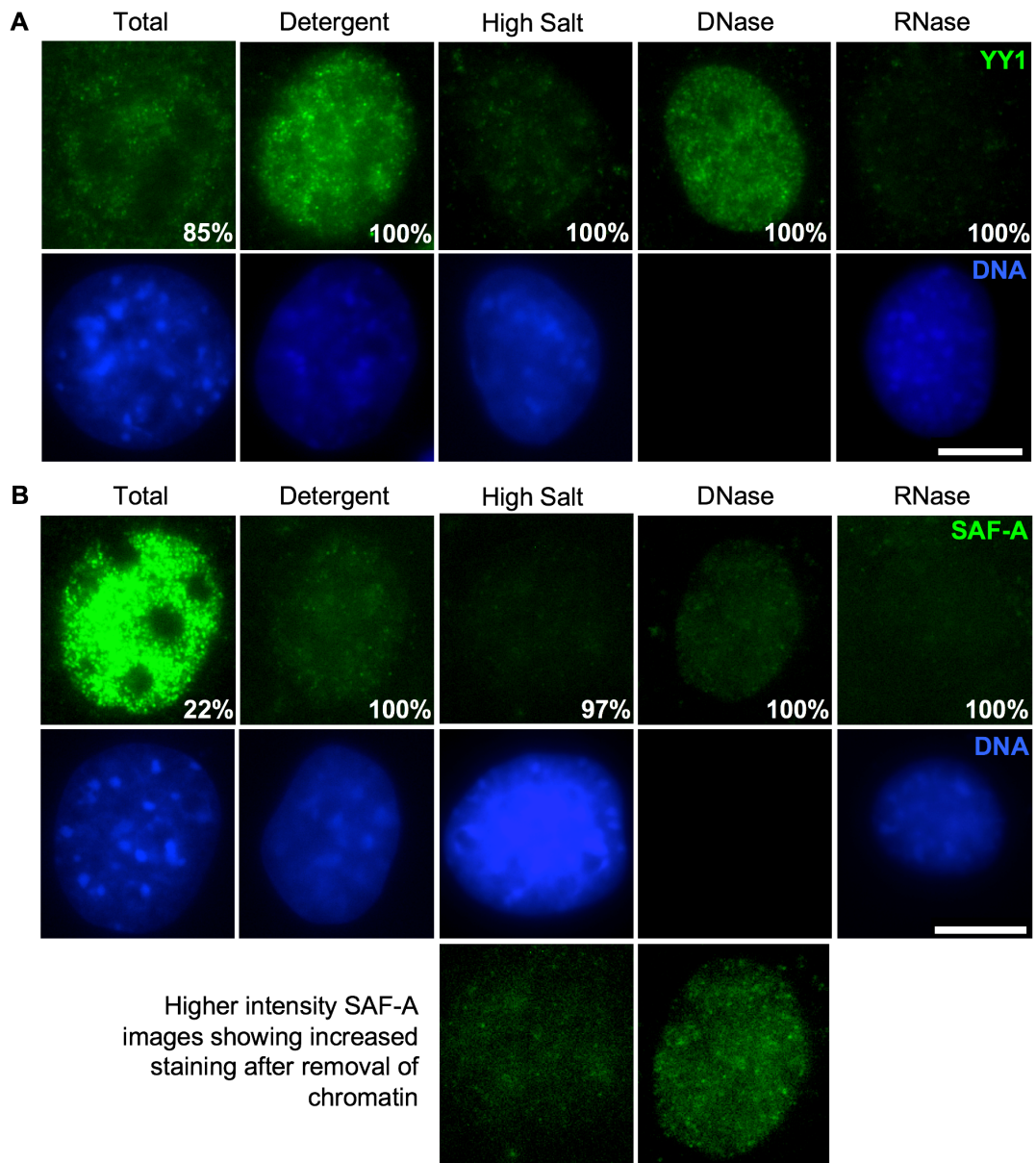


Figure 3.5 Nuclear matrix extraction profile of YY1 and SAF-A (Ridings-Figueroa et al., 2017, Stewart and Coverley, 2018)

(A) Images show YY1 (green) after serial extraction of MEFs. The proportion of cells which exhibit the same staining pattern is indicated ($n > 100$ for each condition). DNA (blue) shows the extent of nuclease treatment, scale bar 10 microns. (B) As in (A) but immunostaining for SAF-A (green). All images of the same channel were taken at the same exposure and modified identically unless otherwise stated.

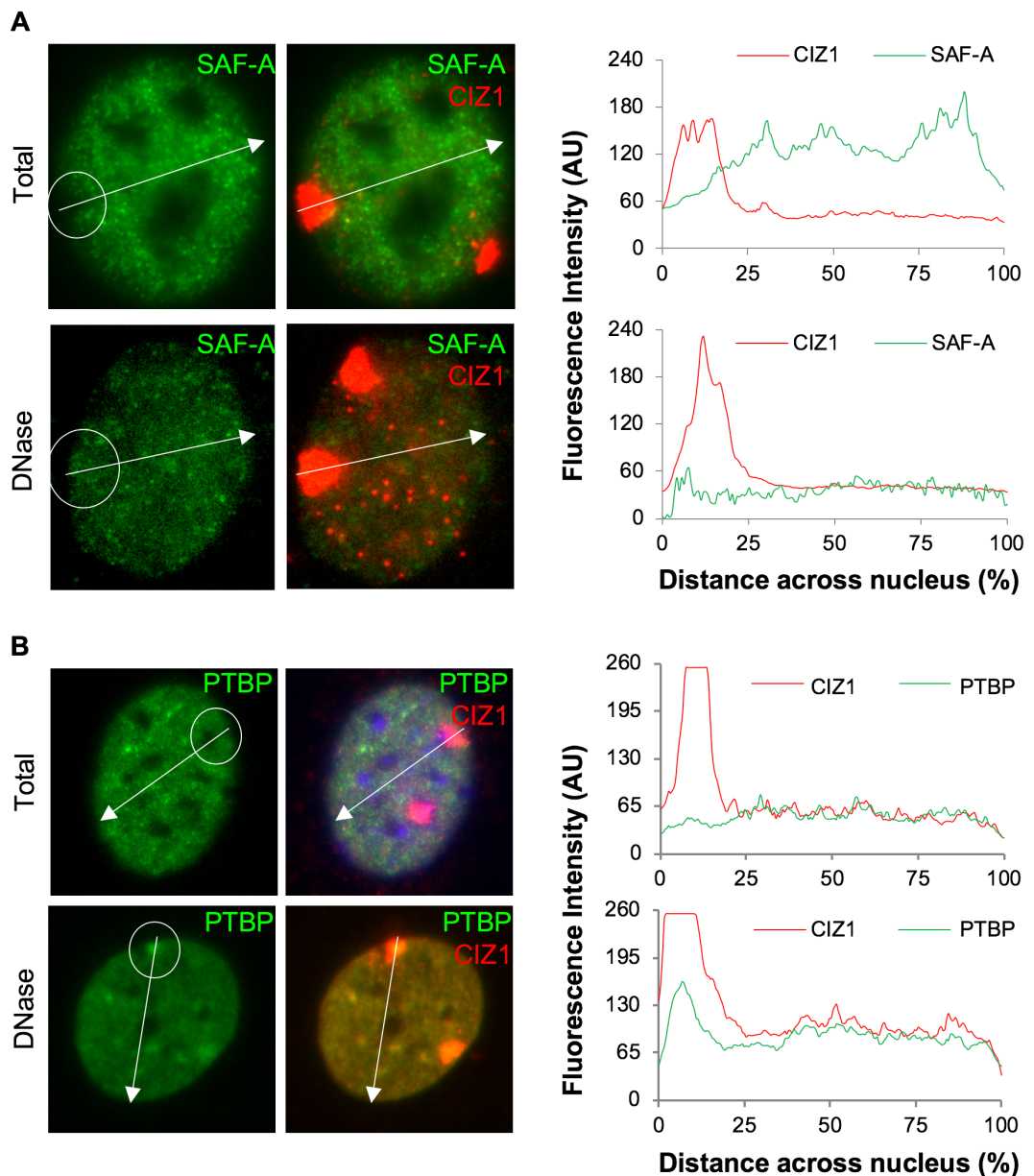


Figure 3.6 *Line intensity scans on unextracted and extracted nuclei* (Stewart and Coverley, 2018)

(A) Left, example nuclei co-immunostained for SAF-A (green) and CIZ1 (1794, red) without any extraction (total) and following chromatin removal (DNase). Right, line intensity scan of fluorescence intensity of each channel along the indicated white arrow of the nuclei on the left. (B) As in (A) but immunostaining for PTBP (green). Line intensity scan was performed on unmodified images taken at the same exposure, $n > 50$. Figure images are examples not modified identically. PTBP immunofluorescence staining was carried out by undergraduate student Barbora Kriukelyte, under my supervision.

3.4.4 Summary model of nuclear matrix proteins at the Xi

Taken together, the data can be assimilated into a model of the nuclear matrix (**Figure 3.7**) (Ridings-Figueroa et al., 2017). The core protein NM consisting of proteins such as CIZ1 is in close proximity to the RNA-protein NM, consisting of RNase sensitive proteins such as SAF-A, YY1, PTBP and Xi-CIZ1, and the RNA they interact with such as *Xist*. Also in close proximity is chromatin, which masks NM protein epitopes and can be removed with this extraction protocol to allow antibody accessibility.

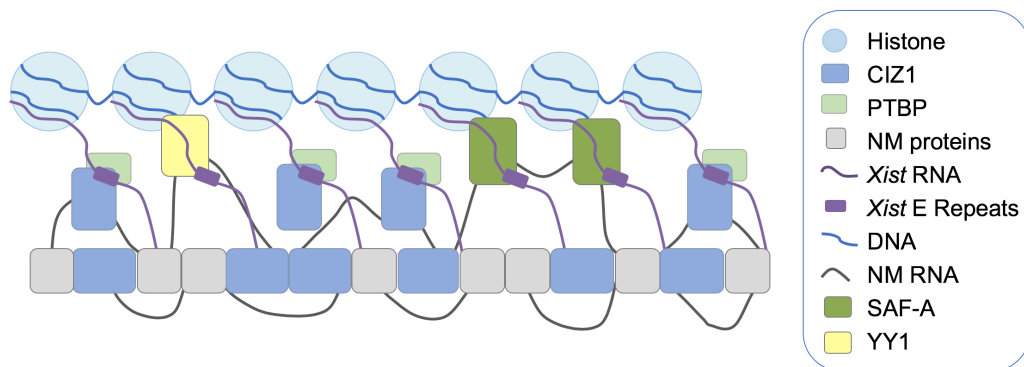


Figure 3.7 Model of Nuclear Matrix Structure (Ridings-Figueroa et al., 2017)

Model showing two populations of CIZ1 (dark blue) in the nuclear matrix. RNA-dependent CIZ1 (**Figure 3.2**) interacts with *Xist* via the E repeats (purple), and RNA-independent CIZ1 is part of the core nuclear matrix, along with other nuclear matrix proteins (grey). SAF-A, YY1 and PTBP1 are also part of the RNA-dependent nuclear matrix (dark green, yellow and light green respectively; **Figures 3.5** and **3.6**).

3.4.5 Detection by western blot

As an alternative to analysis by immunofluorescence, the same protocol can be performed on cells in suspension to produce cell lysates which can be analysed by western blot (**Figure 3.8**) (Wilson et al., 2016a). This can be useful as extracted proteins are released into the soluble fraction which can be kept and visualised to confirm extraction. Additionally, analysis by western can allow differentiation between variants of a protein if they give visibly different bands, and reveal their possible differential extractability.

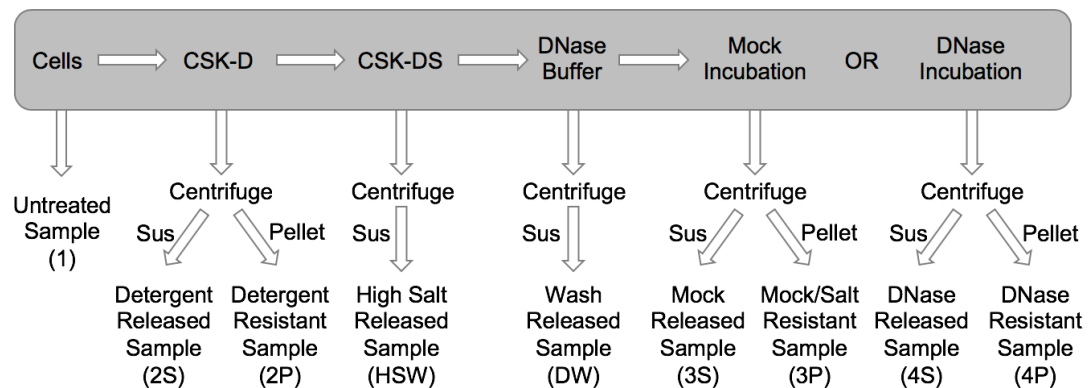


Figure 3.8 Schematic of nuclear matrix extraction for detection by western blot

Schematic showing the different stages of the nuclear matrix extraction protocol and the fractions produced.

Successful chromatin removal can be assessed by looking for release of histone H3 into the soluble fraction following DNase treatment. Though the protocol and buffers for detection by western blot are largely the same as for analysis by immunofluorescence, some small changes are required. For example, the RNase inhibitor vanadyl ribonucleoside complex (VRC) used in the immunofluorescence protocol prevents H3 extraction when performing the extraction in suspension (**Figure 3.9**). Therefore a different RNase inhibitor, RNase OUT is used.

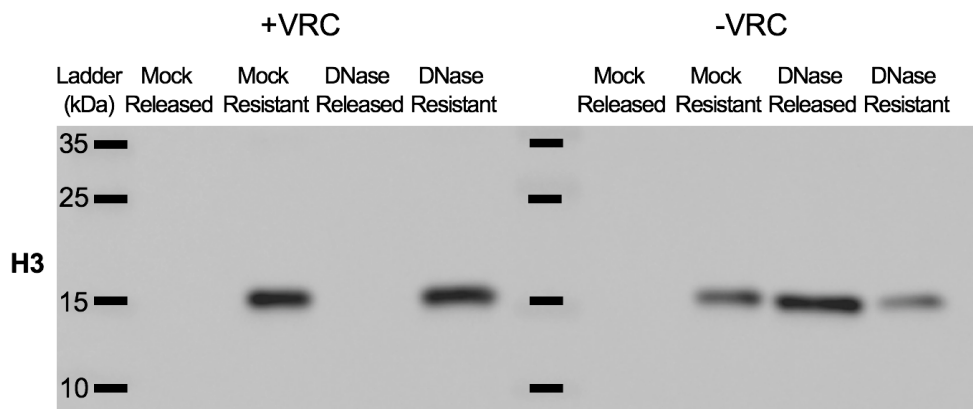


Figure 3.9 RNase inhibitor VRC prevents H3 extraction

Nuclear matrix extraction followed by detection by western blot with an antibody against histone H3, performed on D3T3 cells in the presence and absence of Vanadyl Ribonucleoside Complex (VRC). Upon successful digestion with DNase I H3 is released from the digested DNA becoming soluble, and is therefore detected primarily in the DNase released sample. However, in the presence of VRC, H3 is not released from the DNase resistant sample into the DNase released sample, implying VRC inhibits enzymatic digestion of DNA in this version of the nuclear matrix extraction protocol.

3.4.6 CIZ1 isoform expression upon induction of pluripotency

It has previously been reported that the NM is compromised or immature in ES cells and cancers (Khanuja et al., 1993, Varma and Mishra, 2011). Previous work in the Coverley lab identified that cyclin E, a binding partner of CIZ1, is not anchored to the NM in murine ES cells, but is recruited over the course of differentiation (Munkley et al., 2011). However, CIZ1 anchorage over differentiation was not investigated. As CIZ1 is responsible for recruitment of cyclin E into the nuclear matrix (Copeland et al., 2010) we hypothesised that CIZ1 would also no longer be attached to the nuclear matrix in ES cells, possibly due to expression of CIZ1 splice variants excluding regions required for nuclear matrix binding.

As part of a preliminary collaborative study with the Tada lab (Kyoto University), I performed nuclear matrix extraction on TIG1 cells (human female lung fibroblasts), and derived intermediately reprogrammed stem cells (iRS) and induced pluripotent stem (iPS) cells (Tada et al., 2014). These cells are of particular interest as the process of reprogramming human cells into iPS cells is incredibly inefficient, typically less than 1% (reviewed in (Stadtfield and Hochedlinger, 2010)). In addition, once iPS cells have been generated, dissociation into single cell cultures results in apoptosis. Therefore, investigations which require large amounts of cells and many useful molecular tools which requires clonal expansion from a single cell are not possible. These intermediately reprogrammed (iRS) cells can be expanded from single cell cultures, and reprogram with high efficiency into iPS cells when grown to high-density, allowing generation of much larger and purer cultures of iPS cells than previously possible (Tada et al., 2014).

NM extraction with analysis by western blot on this series of reprogramming cells allows investigation into whether CIZ1 is no longer attached to a NM structure following induction of pluripotency and whether any changes in anchorage coincide with a change in CIZ1 isoform expression.

Total samples reveal multiple CIZ1 bands, as expected as CIZ1 is known to be heavily alternatively spliced (Rahman et al., 2010). The identity of most of these is not known and is an active area of investigation by others in the Coverley lab. CIZ1 splice variant expression is similar in iRS and TIG1 cells, however there are notable changes following transition into the iPS state (**Figure 3.10**). Following induction of pluripotency, a detergent resistant 140 KDa band is absent. Interestingly this is larger than full-length CIZ1, which has an expected MW of approximately 100 KDa. Recent work in the Coverley lab suggests that this could be a multimer of smaller CIZ1 variants or tight complex with another interaction partner. Also associated with differentiation was a band at approximately 50 KDa, however this band was solubilised by detergent treatment.

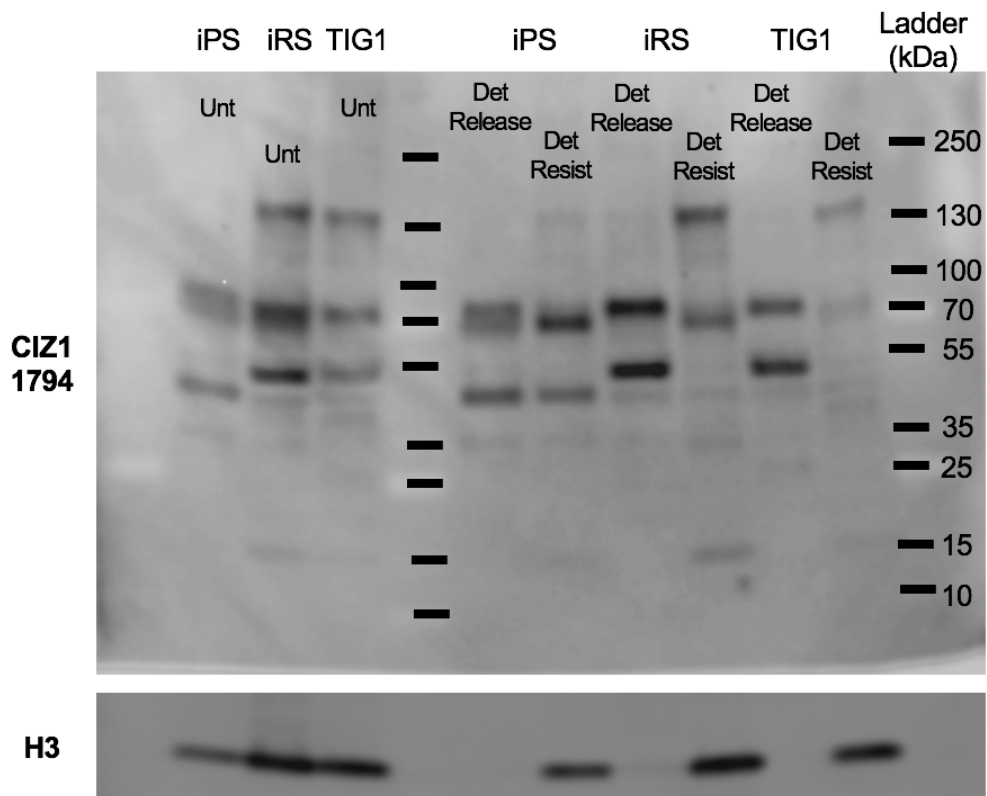


Figure 3.10 *Changes in CIZ1 splice variant expression upon induction of pluripotency*

Nuclear matrix extraction followed by detection by western blot was performed on TIG1 cells and derived iRS and iPS cells with antibodies against CIZ1 (1794) and H3. Sample names as referred to in (Figure 3.8).

Associated with pluripotency were bands at 68 KDa and 45 KDa, which were partially solubilised with detergent in iPS cells. The 68 KDa band was also seen in iRS cells, but at lower levels and was not solubilised, suggesting this splice variant is upregulated as pluripotency is induced, and solubilisation could occur due to saturation of binding sites. Present in all samples is a soluble 72 KDa band.

Overall, it is clear CIZ1 variant expression changes upon induction of pluripotency. The 140 KDa band is a band of interest as a candidate variant present at the Xi as this variant is resistant to extraction and is not present in pluripotent cells where the Xi is reactivated (Maherali et al., 2007).

3.5 Conclusions

I have two optimised nuclear matrix extraction methods which can be used to profile binding of proteins within the nucleus. These extraction methods have allowed me to determine that CIZ1 is part of the RNA-dependent nuclear matrix at the Xi. I have also determined that CIZ1 splice variants are differentially attached to the nuclear matrix. For example F-variant CIZ1 is less capable of incorporation into the RNA-independent nuclear matrix than other CIZ1 splice variants detected by the 1794 antibody. The increase in fluorescence signal intensity seen for many nuclear matrix protein antibodies demonstrates that nuclear matrix extraction reveals epitopes which were previously obscured by chromatin. In addition, the RNase sensitivity of these chromatin masked epitopes suggests the RNA-protein nuclear matrix is in extremely close proximity to chromatin. Overall, the ability of these extraction processes to reveal previously buried epitopes both at the inactive X chromosome and nucleus-wide, demonstrates that the fraction of protein seen in unextracted cells may not be the most relevant fraction or the one performing the function I want to study. This highlights the importance of using these protocols in my subsequent experiments. Finally, I have determined that different CIZ1 isoforms are expressed depending on differentiation state, however further profiling is required to identify the actual splice variants involved is required.

4.0 Effect of CIZ1 deletion on gene expression in fibroblasts

Text and figures contained within this chapter are an amalgamation of my contributions to two pieces of published work, which have been reproduced in full here (**Appendix A and D**) (Ridings-Figueroa et al., 2017, Stewart et al., 2019).

4.1 Introduction

Recently, work in the Coverley lab has shown that female CIZ1 null mice develop fully penetrant lymphoproliferative disorder in adulthood, and primary embryonic fibroblasts (PEFs) derived from CIZ1 null embryos exhibit loss of *Xist* localisation and loss of enrichment of H3K27me3 and H2AK119Ub1 at the inactive X chromosome (Xi) (Ridings-Figueroa et al., 2017). However the role of CIZ1, the mechanism of how loss of CIZ1 leads to these phenotypes, and the effect of CIZ1 loss on gene expression, particularly from the X chromosome, have not been investigated.

Cell lines are commonly used tools when carrying out *in vitro* biological investigations as they are highly proliferative and can be grown for extended periods of time with almost limitless expansion capabilities. However, issues with cell line misidentification and investigations into the effect of long term culture on cells has led to debate over the physiological relevance of data obtained (Pan et al., 2009, Lorsch et al., 2014).

This chapter analyses RNA-sequencing data from primary WT and CIZ1 null PEFs, and culture adapted cell lines derived from these primary populations, to identify candidate pathways affected by CIZ1 loss, to inform further *in vitro* experimentation into the mechanistic roles of CIZ1. In addition this chapter investigates whether findings from primary cells match those from culture adapted derived cell lines, to establish whether they are an appropriate model to use in subsequent experimentation.

4.2 Aims

The aims of these experiments were to generate samples and carry out RNA sequencing on WT and CIZ1 null mouse embryonic fibroblasts and:

- Identify genes and/or pathways deregulated upon loss of CIZ1
- Identify whether CIZ1 loss results in reactivation of the inactive X chromosome
- Investigate whether CIZ1 dependent changes identified in primary mouse embryonic fibroblasts are also evident in derived culture adapted cell lines

4.3 Experimental Design

Primary embryonic fibroblasts (PEFs) were isolated from three WT and three CIZ1 null female mouse embryos by Justin Ainscough, as described in (**Chapter 2.1.1**). RNA was extracted from all PEF lines between 3 and 5 passages by Rebeca Ridings Figueroa. To expand and complement this data set, all six lines were further passaged through crisis to derive a culture adapted mouse embryonic fibroblast cell line (MEFs) from each of the six PEF lines. RNA was extracted from all MEF lines following at least 20 passages in culture (**Figure 4.1**). In addition, RNA was also extracted from CIZ1 null PEFs in which full-length CIZ1 expression had been induced for 24 hours as described in (**Chapter 2.1.2**).

RNA quality for both sets of samples was analysed at the University of York Genomics Facility and libraries generated by Sally James as described (**Chapter 2.6.2**), for sequencing at the Leeds Institute for Molecular Medicine (LIMM) using the Illumina HiSeq 3000 system.

Sequence reads were processed by Katherine Newling as stated in (**Chapter 2.6.3**) to assemble a transcriptome from which gene expression differences between genotypes and primary and culture adapted cells could be investigated.

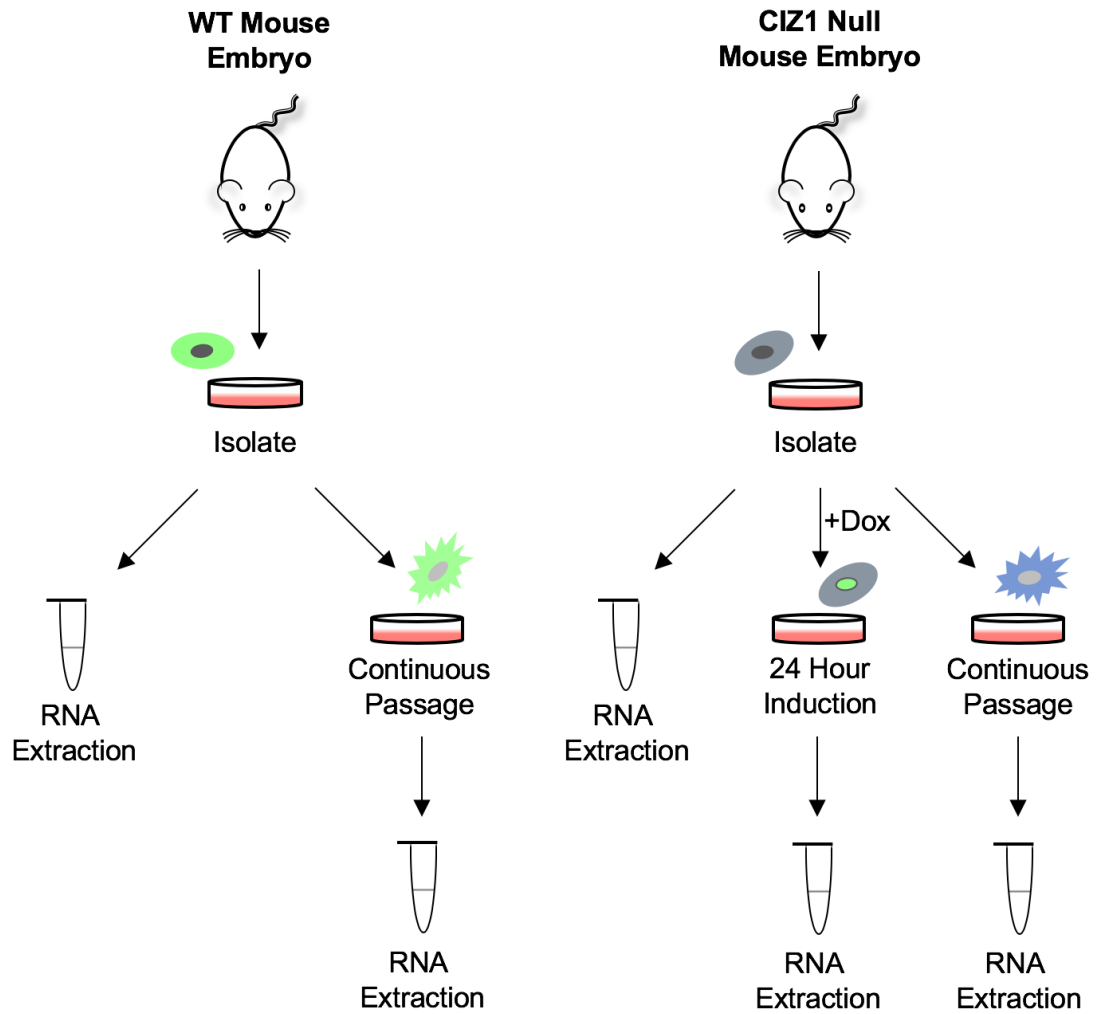


Figure 4.1 Schematic of RNA isolation from WT and CIZ1 null cells

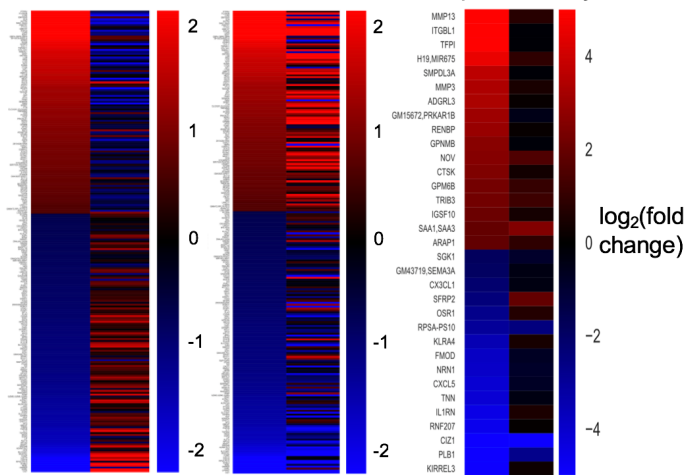
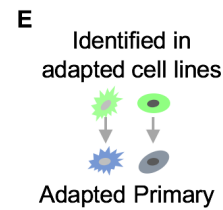
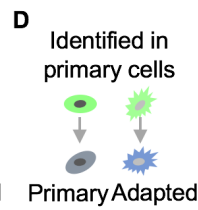
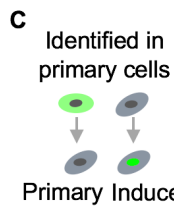
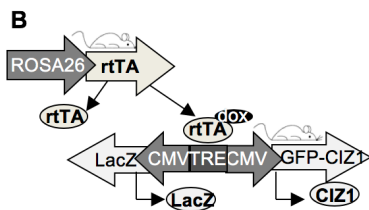
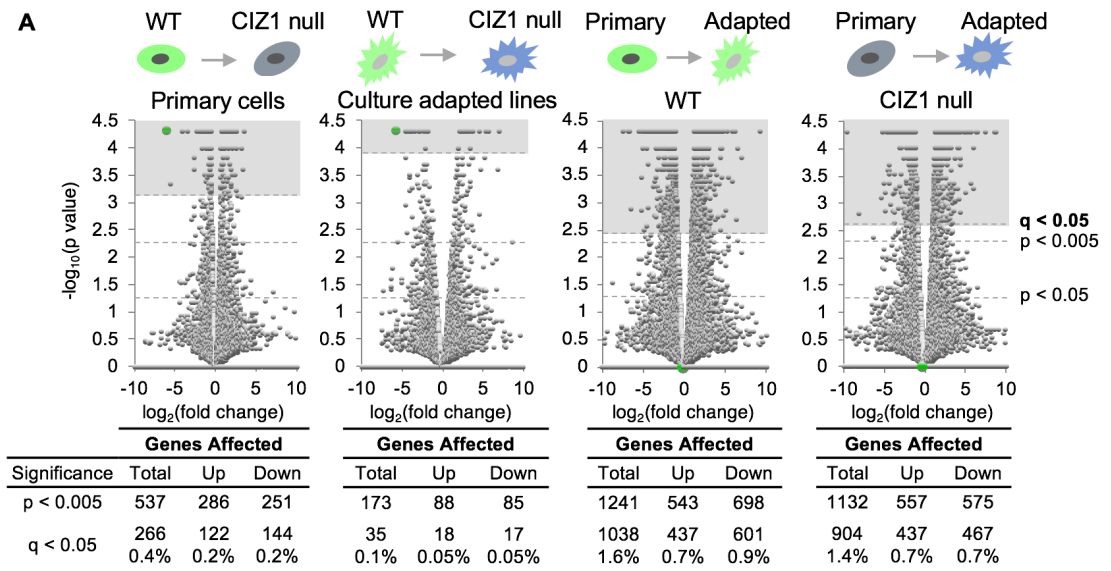
RNA was isolated from WT and CIZ1 primary cells before passage 5. To induce full-length GFP tagged CIZ1 expression in CIZ1 null cells, doxycycline was added for 24 hours prior to RNA isolation. RNA was also isolated from cells of both genotypes which had adapted to long term culture (>20 passages).

4.4 Results and Discussion

4.4.1 Genome wide effects of CIZ1 loss and culture adaption

Loss of CIZ1 affects a relatively small number of genes, with only 266 genes being identified as CIZ1-dependent ($q < 0.05$) in PEFs (**Figure 4.2A**). Within these a similar proportion of genes are up- and down- regulated. Re-expression of full-length CIZ1 reverses these changes back towards WT levels in 75% of cases within 24 hours (**Figure 4.2B-C**), confirming their CIZ1 dependency. However, only 35 genes were identified as CIZ1-dependent in the derived culture-adapted (MEF) lines (**Figure 4.2A**), and only 12 of these genes overlap with the 266 genes identified in PEFs (**Figure 4.2D-F**).

Culture adaption has a much greater effect on gene expression than loss of CIZ1, with over 1038 genes affected in the WT state (**Figure 4.2A**). The significant overlap between the genes affected by culture adaption and the 266 CIZ1-dependent genes identified in PEFs (39% $q < 0.05$, 56% $p < 0.05$) explains the lack of coherency in CIZ1 dependent genes between PEFs and derived MEFs and argues that adaption to culture involves corruption of the process in which CIZ1 normally functions.



CIZ1-dependent genes

F

266 CIZ1-dependent genes in primary cells

12 CIZ1-dependent genes in cell lines

35 CIZ1-dependent genes in cell lines

Figure 4.2 The effect of loss of CIZ1 and culture adaption on gene expression

(A) Volcano plots showing mean fold change in transcript level (FPKM) against significance (P -value, calculated by cuffdiff) for all 65530 annotated transcription units assigned to mouse genome assembly GRCm38, derived from whole transcriptome RNA-seq of three primary WT (13.1 p4, 13.8 p4, 14.4 p4) embryonic fibroblast populations and culture-adapted cell lines derived from the same primary cells, and three primary CIZ1 null (13.15 p3, 13.17 p3, 14.2 p4) embryonic fibroblast populations and culture-adapted cell lines derived from the same primary cells. CIZ1 is indicated in green. The populations of cells being compared are indicated above each plot. Below, the number of transcription units affected at the indicated significance thresholds.

(B) Schematic of transgenes used to create doxycycline-inducible expression of full-length GFP-CIZ1 in CIZ1 null mice and derived cells (generated by Justin Ainscough (Ridings-Figueroa et al., 2017)).

(C) Heat map showing 266 transcription units ($q < 0.05$, cuffdiff) affected by the loss of CIZ1 in primary cells, organised by fold-change. Right column, mean fold change in the same transcription units in CIZ1 null cells (13.17 p3, 14.19p1), 24 h after induction of full-length CIZ1.

(D) Heat map showing fold-change in 266 CIZ1-dependent transcription units ($q < 0.05$, cuffdiff) in primary cells (left), compared to fold-change in the same genes in culture-adapted derivatives, upon loss of CIZ1 (right).

(E) Heat map showing the 35 transcription units that are CIZ1-dependent in adapted cells ($q < 0.05$, cuffdiff) (left), and fold change in these transcription units in primary cells (right).

(F) Venn diagram showing overlap between transcription units that are differentially expressed upon loss of CIZ1 in primary and derived lines. Heat maps are organised by fold-change from up (red) to down (blue).

4.4.2 Loss of CIZ1 has no preferential effect on gene expression from the X chromosome

As CIZ1 is highly enriched at the inactive X chromosome and its loss results in a female specific phenotype, changes in gene expression from the X chromosome was investigated. There was no wholesale effect of loss of CIZ1 on gene expression from the X chromosome (**Figure 4.3**). In fact, the proportion of CIZ1-dependent genes which are X-linked (3.4%) was no higher than that for autosomes, with only 9 genes significantly changed (**Table 4.1**). It cannot be determined from these data whether the altered gene expression is due to changes in expression from the inactive X chromosome or the active X chromosome. Therefore, it was concluded that there is no evidence that loss of CIZ1 results in inactive X chromosome reactivation.

Table 4.1 *List of X-linked CIZ1 dependent genes and their known roles*

Gene Name	Locus	Known Roles
AGTR2	X:21484543-21489164	Receptor for angiotensin II. Mediates apoptosis (Lehtonen et al., 1999). Mutations associated with X-linked cognitive disability (Vervoort et al., 2002).
FHL1	X:56731786-56793641	Contains highly conserved zinc-fingers. Mutations found in multiple myopathies (Gueneau et al., 2009).
TNMD	X:133850979-133865577	Glycoprotein thought to be involved in tendon maintenance and healing and an inhibitor of angiogenesis (Shukunami et al., 2016).
TMEM164	X:142681405-142843494	Transmembrane protein, currently no identified role.
SCML2	X:161117192-161258214	Member of polycomb group of proteins. Involved in heterochromatin organisation during late spermatogenesis (Adams et al., 2018, Maezawa et al., 2018).
FIGF	X:164373271-164402659	Also known as VEGF-D, involved in angiogenesis and lymphangiogenesis (Lohela et al., 2009).
GPM6B	X:166238910-166389162	Glycoprotein thought to have a role in actin skeleton maintenance, neural development and osteoblast function (Drabek et al., 2011, Dere et al., 2015).
GM6472, RPL7A-PS12	X:152908777-152910296	Predicted pseudogenes of ribosomal protein S7 (GM6472) and ribosomal protein 7A (RPL7A-PS12) (Bult et al., 2019).
EGFL6	X:166522994-166585720	Signal peptide member of epidermal growth factor repeat superfamily. Possible roles in obesity and cancer metastasis (Oberauer et al., 2010, An et al., 2019).

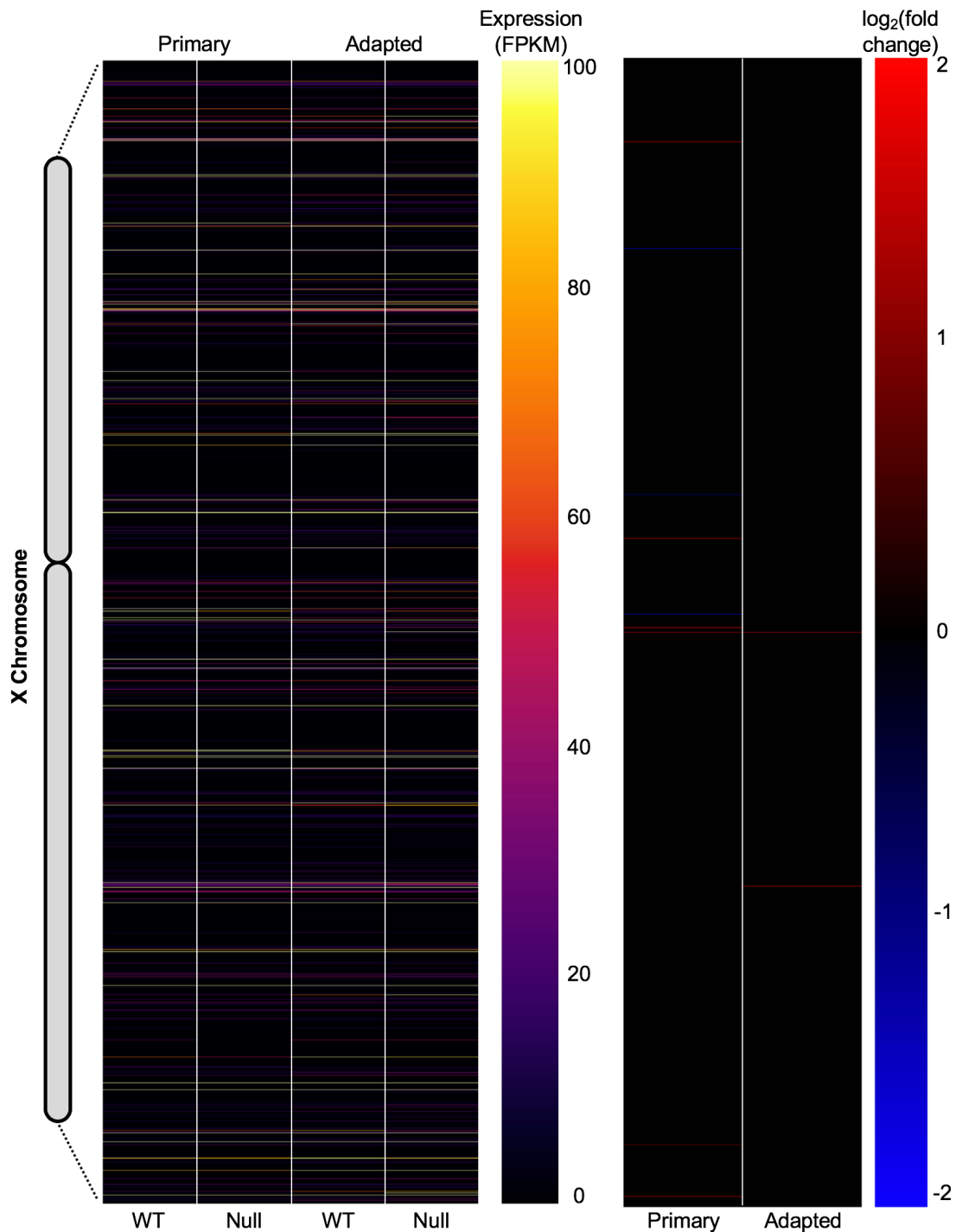


Figure 4.3 Effect of loss of *CIZ1* on X-linked gene expression

Left, heatmap showing expression (FPKM) of all transcription units located on the X chromosome ordered by location, in primary and culture adapted WT and *CIZ1* null cells. Right, heatmap showing the log₂(fold change) in expression of all transcription units from the X chromosome ordered by location in *CIZ1* null cells compared to WT for primary and culture adapted cells.

The effect of loss of CIZ1 on selected components of the XCI process was also investigated. Neither *Xist* or *Firre* lncRNA are significantly affected by CIZ1 loss in primary cells, nor are hnRNPU (SAF-A) or Spen (**Figure 4.4A**). Culture adaption appears to have a much greater effect on expression of these components however, due to high variability between culture adapted samples, only *Xist* is statistically significantly upregulated in CIZ1 null cells (**Figure 4.4A**).

To follow up protein level observations made by Robert Turner, a master's student in the Coverley group, on changes in polycomb repressive complexes in CIZ1 null cells (described in greater detail in **Chapter 5**), components of the polycomb repressive complexes were analysed at transcript level. All are unaffected by loss of CIZ1 with the exception of upregulation of RING1A, however EZH2 is significantly upregulated in WT cells following culture adaption (**Figure 4.4B**).

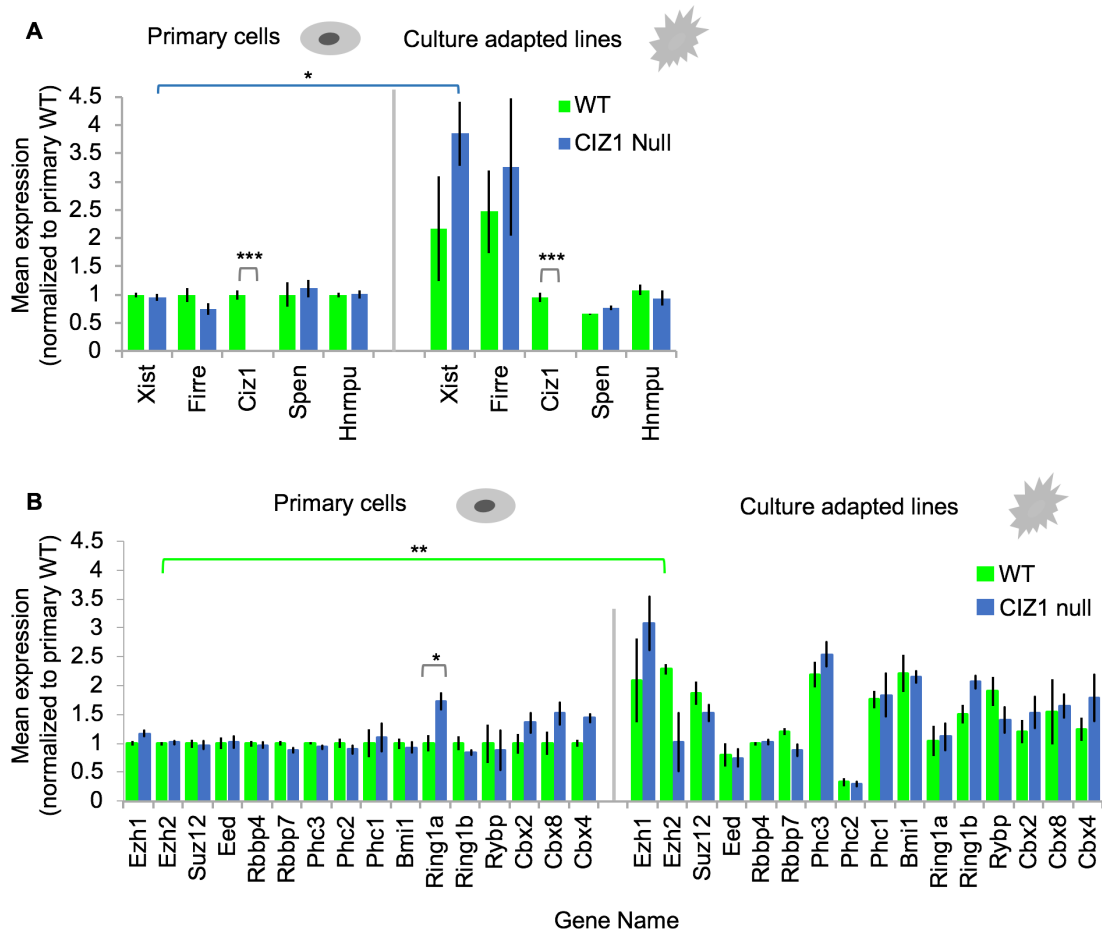


Figure 4.4 Expression of components of X chromosome inactivation

(A) Mean transcript levels of components in the process of X chromosome inactivation, normalized to primary WT levels, \pm SEM. Data is derived from FPKMs from three independent WT and CIZ1 null cell populations, showing little change in primary cells and inconsistent changes in culture-adapted derivatives bar CIZ1, which remains absent in CIZ1 null cells. Compared by *t*-test. B) As in A) but for PRC1 and PRC2 subunits.

Though there was no obvious change in PRC subunit gene expression, splice variant analysis was also performed to identify whether a change in isoform could be behind the loss of epigenetic mark enrichment. For example, if there was a shift to a catalytically inactive variant of RING1 or EZH2, the catalytic subunits of PRC1 and 2 respectively, this would result in inability to lay down epigenetic marks despite normal expression.

In general, there were few changes in PRC subunit splice variant expression in the absence of CIZ1 (**Figure 4.5**). Conversely, culture adaptation appeared to have some interesting effects, most notably the emergence of splice variant 'TCONS_00153598' of EZH2 in culture adapted cells. However, the sequence of this splice variant when translated (UniProt Q61188, **Appendix F**) matches the consensus sequence of EZH2 protein and would therefore not be expected to have an effect on EZH2 activity. Intriguingly 'TCONS_00153598' differs from 'TCONS_00153599', which is predominant in primary cells, by loss of a 5' expansion, some of which has previously been annotated as 5' UTR. As UTRs commonly contain elements controlling RNA stability and translation, it is possible that this loss could result in loss of a regulatory element which could inhibit or promote EZH2 protein expression (reviewed in (Mignone et al., 2002)).

Therefore it was concluded that loss of CIZ1 does not alter expression of XCI components in a way which would explain the absence of *Xist* localisation and epigenetic mark enrichment at the Xi. The lack of effect of CIZ1 loss on the X chromosome despite these absences implies that silencing is already established prior to CIZ1 becoming required for *Xist* localisation and epigenetic maintenance; and silencing can be maintained in an *Xist* and epigenetic mark independent mechanism, possibly for example by DNA methylation.

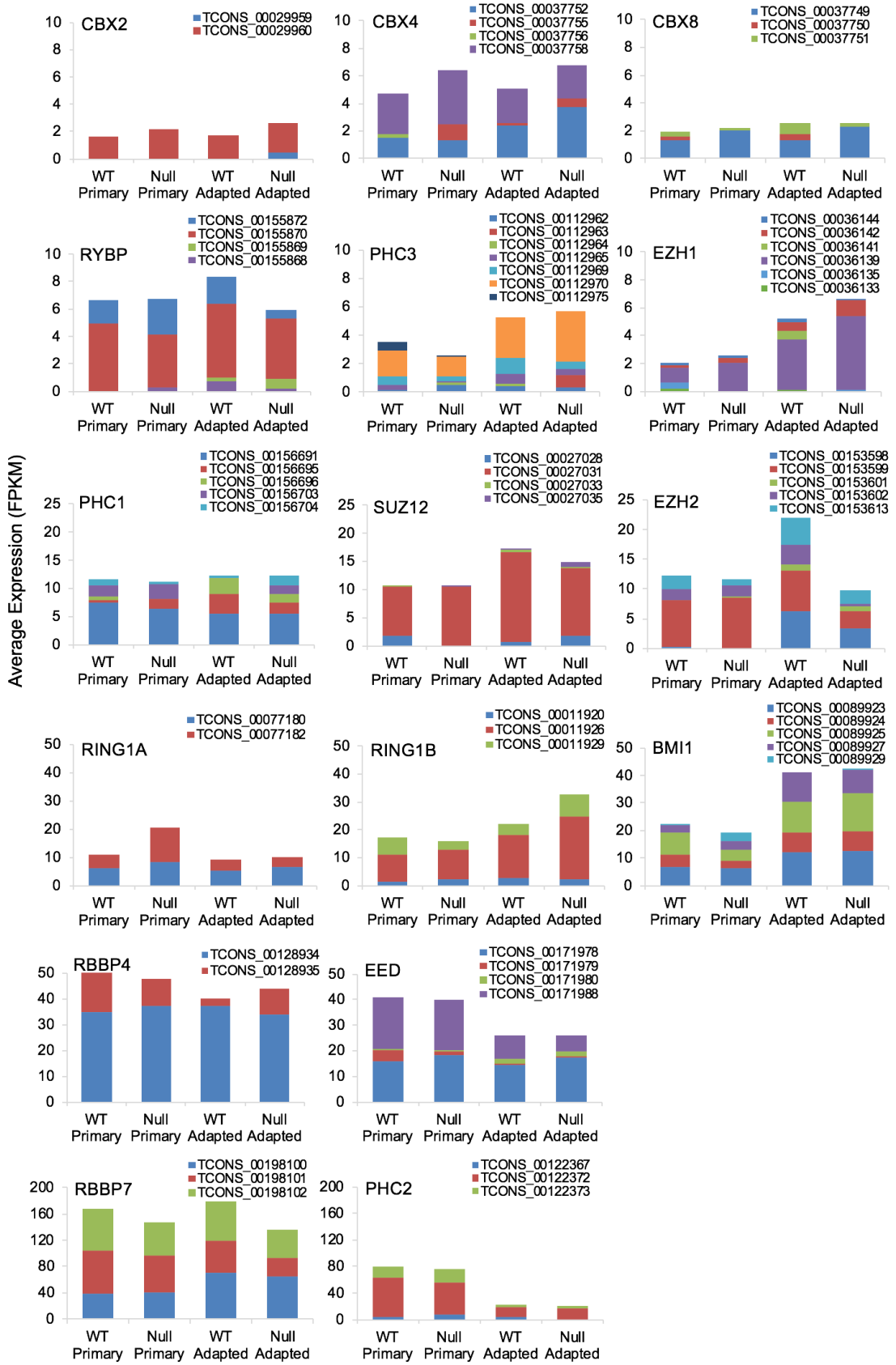


Figure 4.5 Isoforms of polycomb repressive complex subunits

Splice variant assemblies were derived from triplicate WT and CIZ1 null RNA-seq reads (primary cells and culture adapted derivatives). Histograms show the average total expression (FPMK) for each PRC subunit and the contribution of each splice variant (TCON).

4.4.3 Loss of CIZ1 results in PRC target deregulation genome wide

To identify candidate pathways affected by loss of CIZ1, Gene Set Enrichment Analysis (GSEA) was performed on the 266 CIZ1-dependent genes identified in primary cells (Subramanian et al., 2005). Probing the MSig Curated Gene Set database revealed targets of the polycomb repressive complexes were significantly affected following loss of CIZ1 (**Table 4.2**).

Table 4.2 PRC related gene sets from GSEA curated gene sets

PRC related gene sets from GSEA curated gene sets represented in 266 transcription units significantly changed between WT and CIZ1 null primary cells

Curated Gene Set	Percentage Overlap	p-value	q-value	Set Identifier
BENPORATH EED TARGETS	2.73	1.40E-12	7.10E-11	M7617
BENPORATH PRC2 TARGETS	3.07	7.28E-10	2.19E-08	M8448
BENPORATH SUZ12 TARGETS	3.08	3.53E-15	2.83E-13	M9898
DOUGLAS BMI1 TARGETS DN	2.23	1.73E-03	8.85E-03	M14279
DOUGLAS BMI1 TARGETS UP	2.65	6.25E-07	8.84E-06	M15103
KONDO EZH2 TARGETS	4.90	1.61E-08	3.36E-07	M5301
LU EZH2 TARGETS DN	2.17	4.71E-04	2.94E-03	M2140
LU EZH2 TARGETS UP	2.37	1.22E-03	6.48E-03	M2139
NUYTTEN EZH2 TARGETS UP	3.57	2.03E-19	3.99E-17	M4196
PASINI SUZ12 TARGETS DN	8.89	9.59E-25	4.53E-22	M2293
PASINI SUZ12 TARGETS UP	3.57	3.47E-03	1.59E-02	M2291
WIEDERSCHAIN TARGETS OF BMI1 AND PCGF2	10.53	1.10E-06	1.46E-05	M2316

Similar results were obtained when probing the MSig Oncogenic Signatures Gene Set database (**Table 4.3**). In general, the sets with the greatest overlaps and highest significance are reciprocal sets, i.e. genes which are *upregulated* upon loss of a PRC subunit or vice versa. As these are repressive complexes, this is the pattern which would be expected from direct targets of the complexes.

Table 4.3 PRC related gene sets from GSEA oncogenic signatures

PRC related gene sets from GSEA oncogenic signatures, represented in 266 transcription units significantly changed between WT and CIZ1 null primary cells

Oncogenic Signatures Gene Set	Percentage Overlap	p-value	q-value
BMI1 DN MEL18 DN.V1 DN	4.08	1.76E-04	6.28E-04
BMI1 DN MEL18 DN.V1 UP	10.34	1.08E-14	4.10E-13
BMI1 DN.V1 DN	3.47	1.24E-03	3.21E-03
BMI1 DN.V1 UP	10.88	6.63E-16	4.18E-14
MEL18 DN.V1 DN	4.05	1.83E-04	6.39E-04
MEL18 DN.V1 UP	11.35	3.63E-16	3.43E-14
PRC1 BMI UP.V1 DN	2.60	4.15E-03	8.71E-03
PRC1 BMI UP.V1 UP	1.05	0.27	0.32
PRC2 EED UP.V1 DN	0.52	0.64	0.68
PRC2 EED UP.V1 UP	1.55	0.09	0.11
PRC2 EZH2 UP.V1 DN	5.67	1.55E-08	1.04E-07
PRC2 EZH2 UP.V1 UP	2.56	4.42E-03	8.98E-03
PRC2 SUZ12 UP.V1 DN	3.14	6.77E-04	2.06E-03
PRC2 SUZ12 UP.V1 UP	1.03	0.28	0.32

This highly significant effect on PRC target genes is not detectable in culture adapted cells (**Figure 4.6**), consistent with the lack of coherency in CIZ1-dependent genes observed in (**Figure 4.2**). In fact, the process of culture adaption has a greater effect on PRC target gene expression than loss of CIZ1 (**Figure 4.6**). Therefore CIZ1 loss appears to result in deregulation of direct PRC target genes, which are also deregulated following adaption to long term culture.

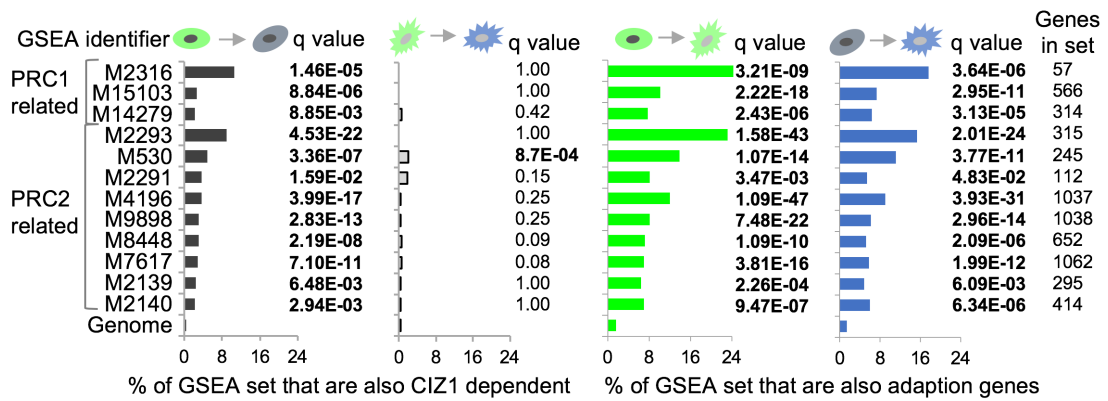


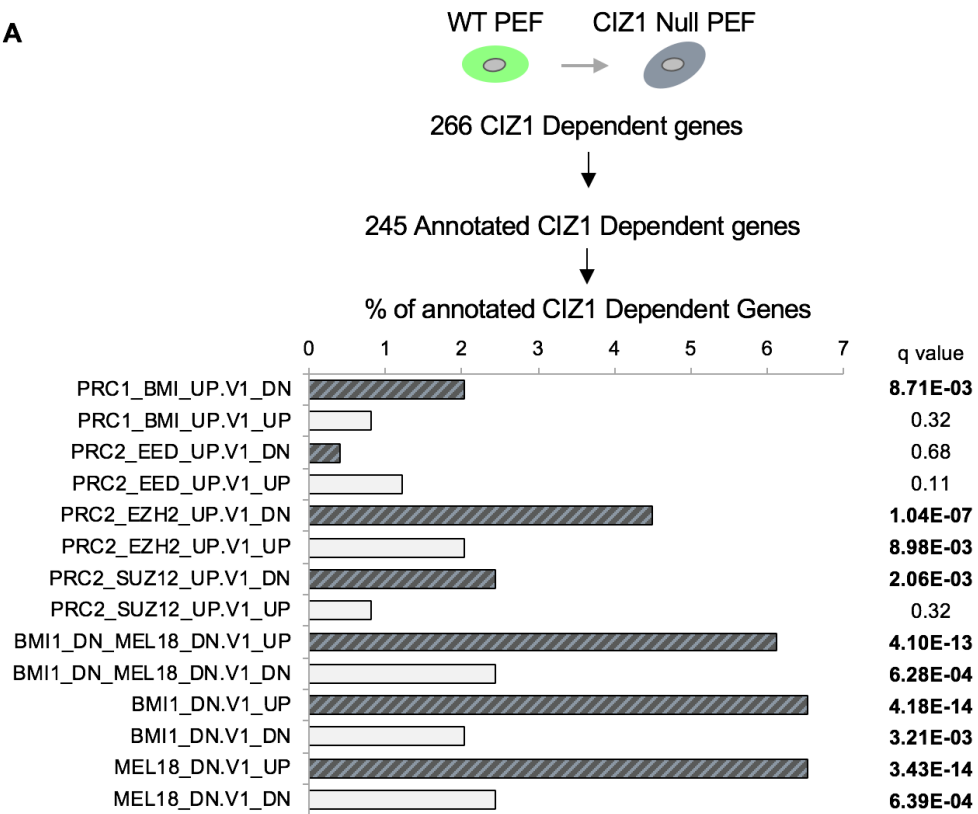
Figure 4.6 Loss of CIZ1 and culture adaption results in PRC target deregulation genome wide

Relationship between affected transcription units ($q < 0.05$, cuffdiff) in (Figure 4.2) and PRC1 or PRC2-related curated gene sets (GSEA MSig(Subramanian et al., 2005)), with significance expressed as q value. Set identifiers and number of genes in sets are indicated. Overlap with CIZ1-dependent genes in primary cells (dark grey) and culture-adapted cells (light grey). Overlap with genes affected by culture adaption of WT cells (green) and CIZ1-null cells (blue).

Though deregulation of PRC target genes may be linked to the absence of epigenetic marks at Xi in primary CIZ1 null cells, the direction of the change is opposite to what would be expected. Of the 266 CIZ1-dependent genes ($q < 0.05$) identified in primary cells, 245 were annotated transcripts. GSEA using the Oncogenic Signatures database revealed significant overlap with targets of the PRC as seen previously (Figure 4.7A).

However, when these CIZ1-dependent genes are separated based on direction of change, it is the genes which are downregulated upon loss of CIZ1 which have the strongest and most significant overlap with the reciprocal PRC sets (Figure 4.7B). As CIZ1 null PEFs lose PRC-mediated H2AK119Ub1 and H3K27me3 repressive epigenetic marks from the Xi, if we were to assume this observation translates genome-wide, this should result in overexpression of PRC target genes. However, there appears to be preferential downregulation of PRC targets genes.

A



B

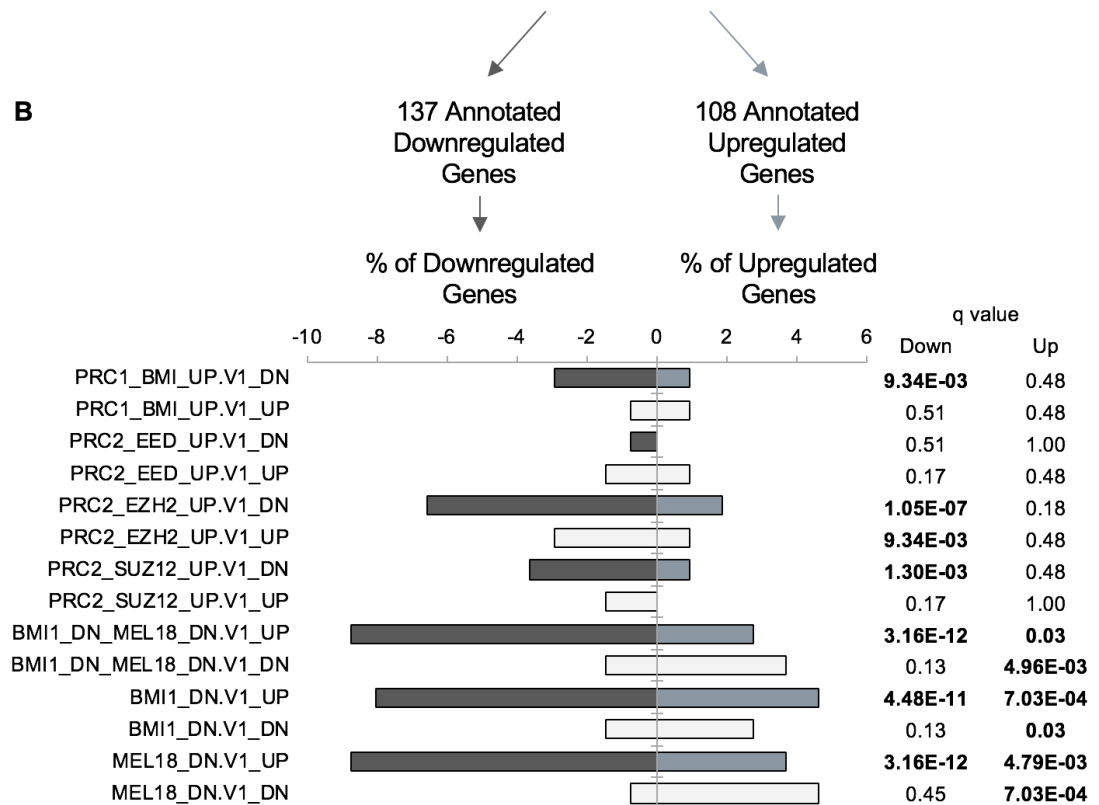


Figure 4.7 Downregulation of polycomb repressive complex targets in primary embryonic fibroblasts

(A) The percentage of the CIZ1-dependent annotated genes in primary embryonic fibroblasts which overlap with Oncogenic Signatures in the MSig GSEA database relating to the polycomb repressive complexes. The significance of these overlaps is shown to the right, significant q values are shown in bold. (B) Same analysis as (A), but when the CIZ1-dependent genes are separated into up- and down-regulation sets.

4.5 Conclusions

CIZ1 loss results in genome-wide deregulation of polycomb repressive complex targets though the mechanism remains unclear. Despite loss of enrichment of epigenetic marks at the Xi, loss of CIZ1 did not result in X chromosome reactivation, supporting the idea that silencing is maintained by a mechanism independent of epigenetic mark maintenance. Adaption to long term culture has a greater effect on PRC target gene expression than loss of CIZ1, demonstrating that investigations into the role of CIZ1 must be done in primary cells. I will follow up these findings *in vitro* to attempt to identify the mechanisms behind these changes.

5.0 Compromised epigenetic maintenance in CIZ1 null PEFs and cell lines

Text and figures contained within this chapter (except **Figure 5.1**) are my contributions to an item of published work which has been reproduced here (**Appendix D**) (Stewart et al., 2019).

5.1 Introduction

Genetic deletion of CIZ1 in primary embryonic fibroblasts results in loss of *Xist* localisation and loss of enrichment of the epigenetic marks H3K27me3 and H2AK119Ub1 from the inactive X chromosome (Xi) (Ridings-Figueroa et al., 2017, Sunwoo et al., 2017). This coincides with altered Polycomb Repressive Complex (PRC) target gene expression as demonstrated in **Chapter 4**. However the mechanism of how CIZ1 loss leads to these events was unclear.

It has been previously reported that the Xi shifts location between the nuclear periphery and nucleolus in S phase, in a mechanism dependent on the lncRNAs *Xist* and *Firre*, and that this behaviour is linked with maintenance of H3K27me3 (Zhang et al., 2007, Yang et al., 2015). Movement away from the nuclear periphery has also been reported to occur during G0/G1 transition (Lyu et al., 2018).

Robert Turner and Victoria Scott (Masters and Undergraduate students respectively, Coverley lab) confirmed Xi relocation, specifically during replication in WT primary embryonic fibroblasts (PEFs), and also demonstrated that Xi relocation does not take place in CIZ1 null PEFs (**Figure 5.1A,B**), allowing CIZ1 to be added to the model as the first protein component identified (Stewart et al., 2019). However, even in WT cells Xi relocation is absent in culture adapted cell lines derived from the same PEF lines (**Figure 5.1C**). Loss of relocation, both through loss of CIZ1 and adaption, coincides with a change both expression level and isoform of the PRC2 catalytic subunit EZH2 (**Figure 5.1D, E**).

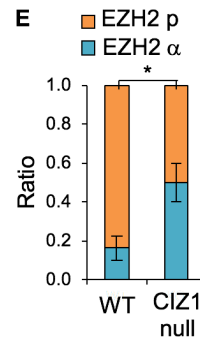
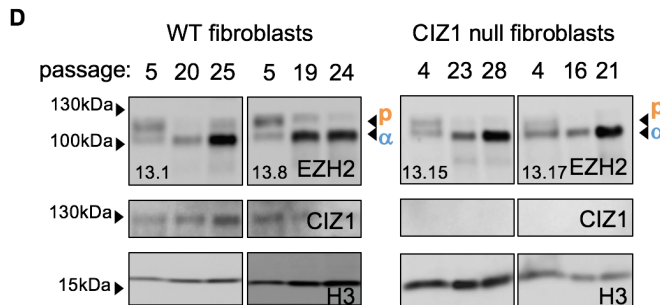
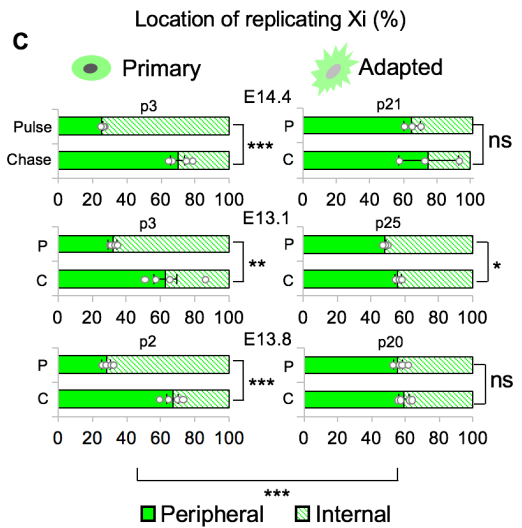
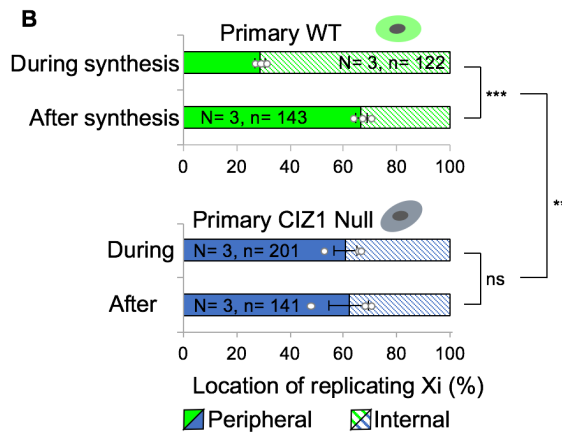
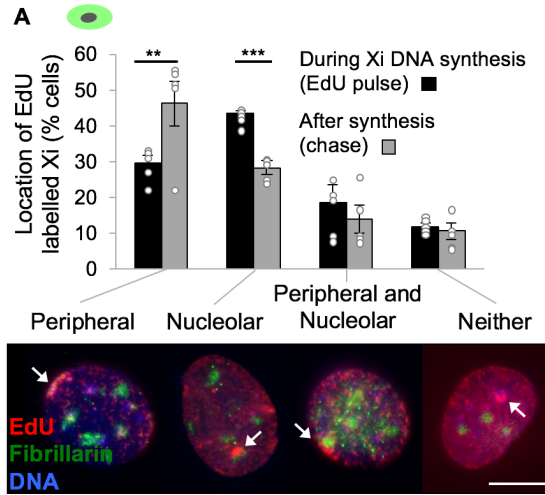


Figure 5.1 CIZ1-dependent Xi relocation during its replication

The contents of this figure are integral to the story but are primarily the work of Robert Turner and Victoria Scott, supported by myself and Professor Coverley. A) Location of Xi at the time of its replication (EdU pulse, black) or after a 30min chase in EdU-free media (grey). Nuclear position, illustrated below, of EdU-Xi (red, arrowed) is classified in relation to the nuclear periphery and in relation to fibrillarin (green), to generate four categories. Histogram shows mean data (13.1 and 14.4) \pm SEM, compared by t-test. Scatter plot of individual values contributing to the mean are overlaid. Bar is 10 microns. B) Binary classification of Xi location (peripheral or internal) at the time of incorporation of EdU (pulse) or after (chase), analysed by t-test. Graph shows mean data \pm SEM for three WT primary embryonic fibroblasts (green, 13.1, 13.8, 14.4 passage 2–3) and three CIZ1 null primary embryonic fibroblasts (blue, 13.15, 13.17, 14.2, passage 2–3). EdU-Xi in WT cells shifts preferred location during the chase period, but this remains unchanged in CIZ1 null cells, compared by ANOVA. C) Binary classification of Xi location (peripheral or internal) at the time of incorporation of EdU (pulse), or 30 min later (chase), analysed by t-test. Graphs show data \pm SEM for three WT primary populations (13.1, 13.8, 14.4 p2–3), and three derived MEF lines ($p > 20$). EdU-Xi in primary WT cells shifts preferred location during the chase period, but remains unchanged in most adapted lines, compared by ANOVA. D) Western blot showing PRC2 catalytic subunit EZH2 (antibody D2C9) in representative primary embryonic fibroblasts and derived culture-adapted populations, of WT and CIZ1 null genotypes, showing two prominent isoforms in early passage cells, designated p and α . E) Quantification of isoforms p and α in early passage WT and CIZ1 null cells, showing mean ratio for three independent populations of each type (WT; 13.1, 13.8, 14.4, CIZ1 null; 13.15, 13.17, 14.2), \pm SEM, significant changes are indicated (t-test). For all analyses * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$. N = independent cell lines, n = total nuclei scored. Individual replicates contributing to each mean are overlaid on bar charts.

5.2 Aims

The aims of my experiments were to investigate whether CIZ1 plays a role in maintenance of the epigenetic landscape by investigating the following questions:

- Is CIZ1 at the Xi differentially anchored within the nuclear matrix at certain points in cell cycle?
- Does the change in EZH2 seen in (**Figure 5.1D**) result in changes in EZH2 localisation and anchorage?
- Do CIZ1 and EZH2 co-localise in the nucleus?

5.3 Results and Discussion

5.3.1 CIZ1 at replicating Xi is resistant to extraction with RNase

CIZ1 at the inactive X chromosome is resistant to extraction with RNase in 3-10% of cells and is consistent between PEFs and MEFs (**Chapter 3.4.1, Figure 5.2**). As CIZ1 is known to have a role in initiation of DNA replication (Coverley et al., 2005) and the experiments were performed on asynchronous populations, one hypothesis for this transient RNase resistance was cell-cycle dependence. Is the cell replicating its Xi at the time of RNase resistance?

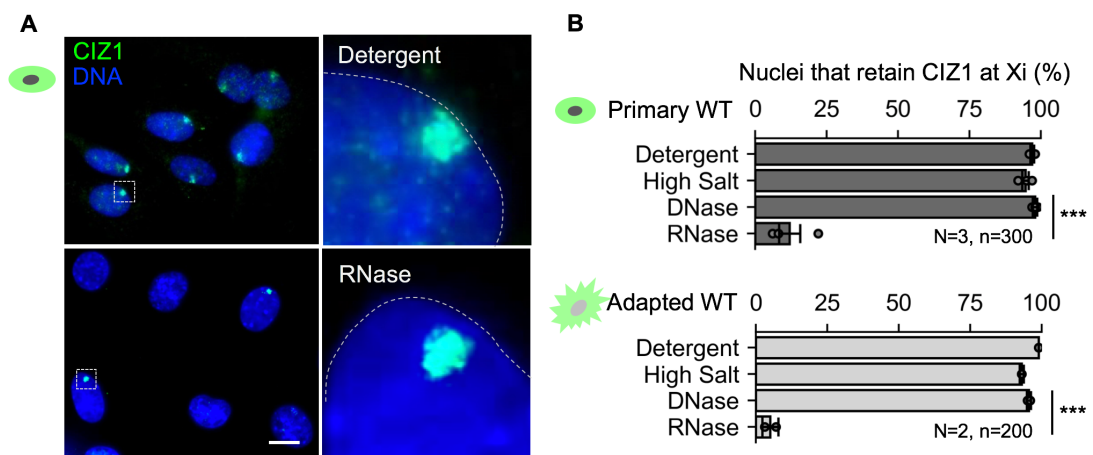


Figure 5.2 Xi-CIZ1 is RNase resistant in a subpopulation of cells

A) CIZ1 at Xi in cycling WT primary embryonic fibroblasts (PEFs), with and without extraction of RNA and associated proteins, detected with anti-CIZ1 antibody 1794 (green). DNA is blue, scale bar 10 microns. B) Proportion of cells that retain CIZ1 at Xi (Xi-CIZ1) at the indicated steps after extraction, for three independent populations of WT PEFs (means \pm SEM for 13.1, 13.8, 14.4 at passage 3–4, $n > 100$ for each step and each line. Significant differences are indicated (t-test). Similar results for culture-adapted derivative line 13.1 (passage > 20), and 3T3 cells is shown below, \pm SDEV. For all analyses $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. N = independent cell lines, n = total nuclei scored. Individual replicates contributing to each mean are overlaid on bar charts.

WT PEFs and MEFs were pulsed with the nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU) for 30 minutes prior to fixation, followed by codetection for EdU and CIZ1 (**Figure 5.3**). EdU labelling generated typical S phase patterns (O'Keefe et al., 1992), allowing identification of where in S phase cells were at the time of labelling. Early S phase cells display a low level of nucleus wide focal staining that is excluded from nucleoli. Mid-S phase cells display enrichment of EdU labelling around the nuclear periphery and peri-nucleolar regions. Late S phase cells have large spots highly enriched in EdU, consistent with late replication of dense heterochromatic regions.

Replicating Xi's were clearly visible and possible to reliably identify without requiring the CIZ1 co-stain, as expected due to the synchronous replication of the Xi chromatin (Casas-Delucchi et al., 2011). Most Xi were identified as replicating during mid-S phase, with the remaining replicating during early S phase, conflicting with the late S phase replication of other heterochromatic regions but consistent with some literature on Xi (Casas-Delucchi et al., 2011, Koren and McCarroll, 2014).

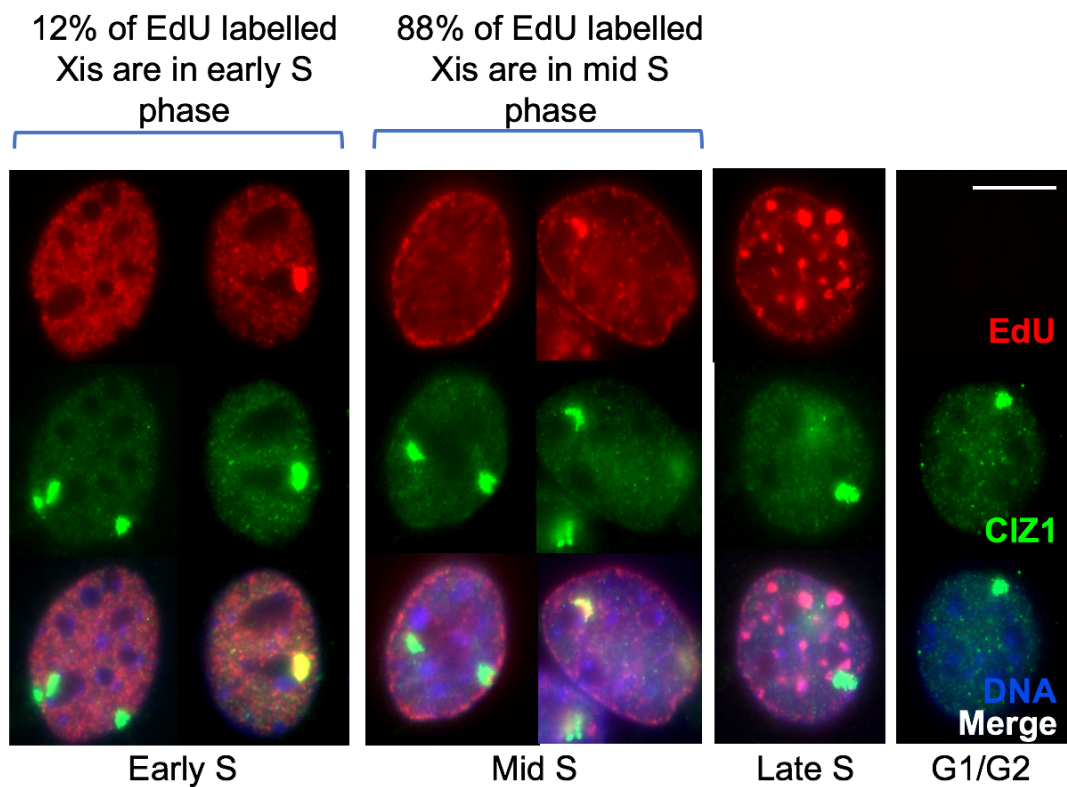


Figure 5.3 *The inactive X chromosome replicated during mid-S phase*
 Example images of 3T3 cell nuclei labelled with EdU (red) for 30 minutes immediately prior to isolation, classified as G1/G2 (no label), or early, mid or late S phase based on pattern of EdU incorporation (O'Keefe et al., 1992). Scale bar is 10 microns. Coimmuno-staining for CIZ1 (green) allows positive identification of the Xi and ability to distinguish between large spots of general heterochromatic replication in late S phase and possible replication of the Xi. This allowed estimation of the stage in S phase in which Xi incorporates EdU (given as %, n=100 replicating Xis). The data are consistent with Xi replication in mid-S phase.

To identify if Xi-CIZ1 is more likely to be resistant to RNase extraction around the time of its replication, the same experiment was performed in WT MEFs with the additional step of prior extraction with RNase before fixation. The percentage of cells with Xi-CIZ1 were analysed for presence of EdU, and EdU positive cells were assigned as 'Early', 'Mid' or 'Late' S phase according to the above criteria (**Figure 5.4A**). Following RNase extraction, significantly fewer cells with Xi-CIZ1 were in G1/G2, and significantly more cells were in Mid S phase, implying a transition to RNA-independent anchorage during its replication, rather than pre- or post-replication.

To investigate further whether anchorage occurs during Xi replication, the percentage of cells with RNase resistant Xi-CIZ1 for each subpopulation was investigated in WT PEFs and MEFs (**Figure 5.4B**). As seen previously, approximately 7% of all cells retained CIZ1 at the Xi following RNase extraction. The percentage of cells retaining Xi-CIZ1 increases when only looking at cells in S phase (EdU positive) and increases again when looking at cells in mid-S phase as expected from (**Figure 5.4A**). Importantly, the biggest enrichment is seen when analysing cells which are replicating their Xi, with almost 60% of these cells retaining Xi-CIZ1 following RNase extraction (**Figure 5.4B, C**). This adds further weight to the hypothesis that anchorage of CIZ1 at the Xi occurs during its replication, though the function of this interaction is unknown.

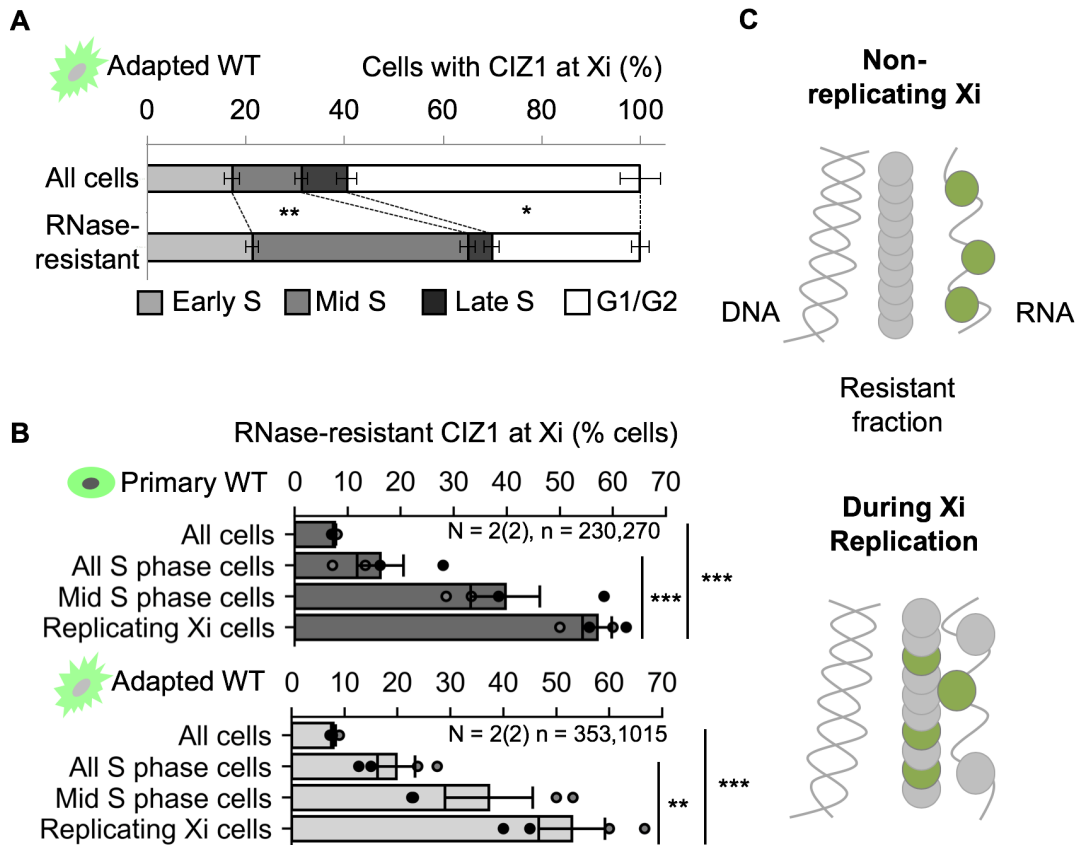


Figure 5.4 CIZ1 at replicating Xi is resistant to extraction with RNase

A) Frequency of interphase (unlabelled) and S phase stages in nuclei with CIZ1 marked Xis, among a complete cycling population of 3T3 cells, compared to the frequency of each stage with RNase-resistant Xi-CIZ1. After RNase treatment preferential removal of Xi-CIZ1 occurs in G1/G2 phase, whereas preferential retention occurs in mid-S phase. Data, \pm SEM, is mean of 4 replicate analyses, $n=1015$, compared by *t*-test. B) Proportion of cells with RNase-resistant Xi-CIZ1 after pulse-labelling with EdU to classify cells by S-phase stage. Upper graph shows mean data for WT primary cells 13.1 and 13.8 (passage 3–4), each in duplicate, \pm SEM (*t*-test). Lower graph shows the same analysis for WT MEFs (13.1, passage > 20), and 3T3 cells, each in duplicate. C) Schematic showing transient change in CIZ1 anchorage during Xi replication. For all analyses $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. N = independent cell lines, n = total nuclei scored. Individual replicates contributing to each mean are overlaid on bar charts, shaded according to cell line where necessary.

The coincident timing of this anchorage with the CIZ1-dependent chromosomal relocation seen in (**Figure 5.1B**) suggests CIZ1-mediated anchorage could be an essential part of the chromosomal relocation process. However this anchorage is also observed in MEFs, where chromosomal relocation is lost. Therefore, for this hypothesis to be correct, loss of chromatin relocation in culture adapted cells must reflect perturbation of another essential component of the pathway.

5.3.2 Nuclear Myosin 1 as a possible component of chromatin movement machinery

Previously identified components of the chromosomal relocation machinery, CIZ1, *Xist* and *Firre*, are not significantly changed at the transcript level following culture adaption (**Figure 5.5A**). However expression of *Myo1c*, the most abundantly expressed myosin isoform in PEFs (**Figure 5.5B**), is significantly downregulated in culture adapted cells (**Figure 5.5A**). Nuclear Myosin 1 (NM1), a splice variant of *Myo1c* (**Figure 5.5C,D**), has been previously reported to have a role in chromosomal movement during DNA repair and entry into quiescence and is therefore an interesting candidate (Chuang et al., 2006, Mehta et al., 2010, Kulashreshtha et al., 2016). Though it must be noted that all actin isoforms present in culture adapted cells are also downregulated (**Figure 5.5E**).

Further investigation would be required to prove whether NM1 is involved in Xi relocation. For example, it must be verified whether actin and *Myo1C* are also downregulated at the protein level. Additionally, investigating whether inhibitors of actin/myosin dynamics prevents Xi relocation in PEFs would also add weight to this hypothesis.

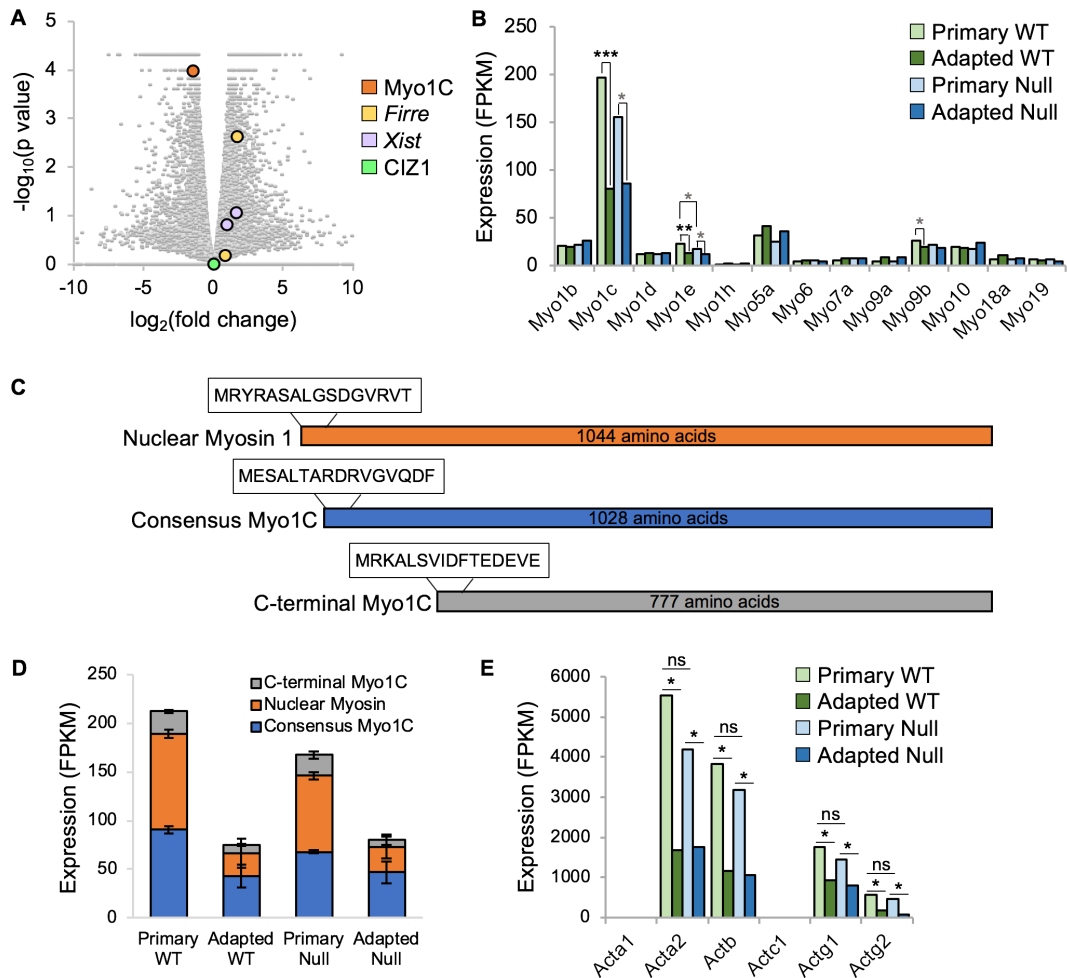


Figure 5.5 Downregulation of Nuclear Myosin 1 and actins following adaption to long term culture

A) Volcano plot showing mean fold change in transcript level (FPKM) against significance (P -value, calculated by cuffdiff) for all 65530 annotated transcription units assigned to mouse genome assembly GRCm38 in culture adapted WT cells compared to the primary WT embryonic fibroblast lines they were derived from (13.1 p4, 13.8 p4, 14.4 p4). Previously identified components of chromosomal relocation machinery are indicated. B) Mean myosin expression levels, \pm SEM. Data is derived from three independent WT and CIZ1 null cell populations. C) Schematic of the differences between splice variants of Myo1C. D) Mean expression levels of Myo1C splice variants, \pm SEM. E) As B) but for actin expression. For all analyses $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$.

5.3.3 Isoform preference of different EZH2 antibodies

Xi relocation during replication has been linked to maintenance of H3K27me₃ (Yang et al., 2015), which is laid down by EZH2, the catalytic subunit of Polycomb Repressive Complex 2 (PRC2). H3K27me₃ is no longer enriched at the Xi in CIZ1 null cells and Robert Turner showed a change in EZH2 isoform at the protein level both upon loss of CIZ1 in primary cells, and culture adaptation of either genotype (**Figure 5.1**). My work in (**Chapter 4**) shows that PRC target gene deregulation occurs under both of these conditions (**Figure 4.6**), suggesting a link between EZH2 and CIZ1. It also poses the question of whether changes in EZH2 isoform are a driver, compensation mechanism or consequence of chromosomal relocation loss.

The smaller (higher mobility) isoform of EZH2 seen in CIZ1 null and culture adapted cells is consistent in apparent MW to the known EZH2 isoform EZH2 α (Grzenda et al., 2013). The higher MW isoform seen in primary cells was designated EZH2p ('primary') and was detectable with two of three different EZH2 antibodies from different suppliers, but with variable preference (**Figure 5.6A**). These molecular weights correlate with the two most abundantly expressed splice variants detected using RNA-seq (**Figure 5.6B**), with the longer isoform including an additional N-terminal sequence, part of which has been previously annotated as 5'UTR (**Figure 5.6C**). However no upstream canonical start codon was detected, and further work would be needed to determine whether the higher MW isoform is derived from this additional EZH2 sequence, post translational modifications or binding to other molecules such as RNA.

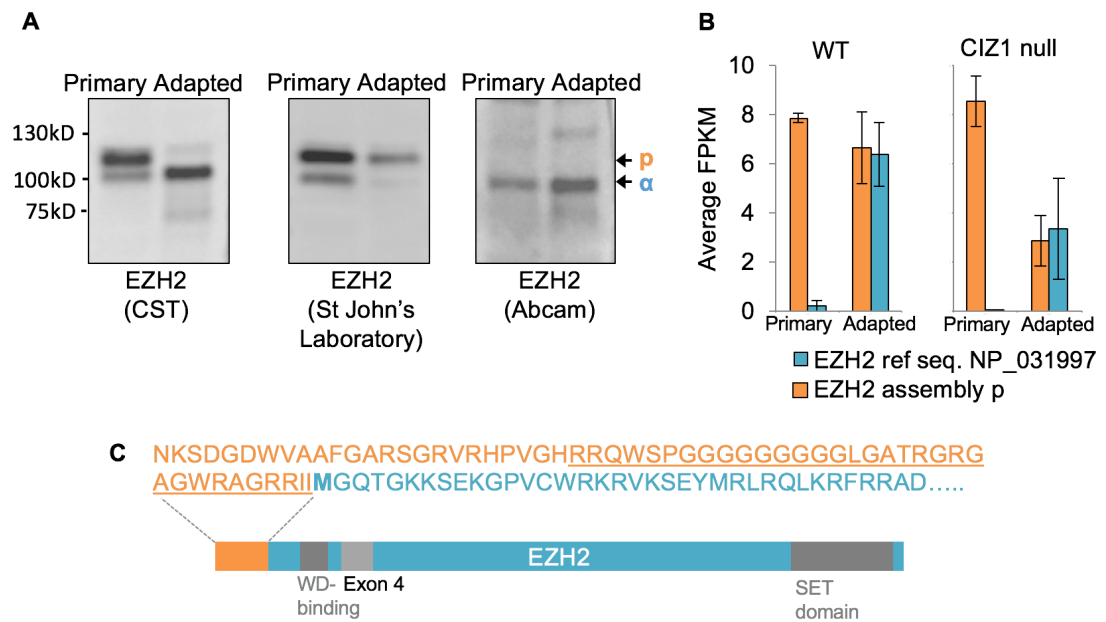


Figure 5.6 EZH2 isoforms, and their relative antigenicity.

A) Whole cell lysates from murine WT primary (p5) and culture-adapted cells (p21) showing dominant expression of EZH2_p protein (upper form) in primary cells and EZH2_α in later passage cells (lower form), detected with rabbit EZH2 antibody CST D2C9 (used in all other figures unless stated), which is centred on Arg354 of Q15910 (EZH2_α). EZH2_p is also detected with rabbit EZH2 antibody STJ112944 (St John's Laboratory, unspecified epitope) but not rabbit EZH2 antibody ab191080 (Abcam), raised against a peptide epitope within amino-acids 1-100 of Q15910. B) Splice variant assemblies derived from triplicate WT and CIZ1 null RNA-seq reads (primary cells and culture adapted derivatives) with mean frequency, ±SEM, for the two most abundant assemblies. These are canonical EZH2 (blue, which matches the reference sequence NP_031997), and a longer variant (orange, denoted p for primary) that includes additional 5' sequence currently annotated as UTR but is otherwise identical. Notably, canonical EZH2 is virtually undetectable in primary cells, while culture adapted cells express both variants at similar levels. C) Translation of the upstream sequence in EZH2 assembly p gives rise to the putative sequence shown in orange. Canonical EZH2 begins with the M highlight in bold in blue. Approximately 50% of mammalian transcripts contain an upstream ORF (uORF), and many non-AUG triplets can act as alternative start codons for uORF translation (Lee et al., 2012).

To investigate the location of these two isoforms, primary WT and culture adapted WT cells were co-stained for CIZ1 and the EZH2 antibodies CST D2C9 (slight preference for EZH2 α) and STJ112944 (preferentially detects EZH2p). Though there was no obvious co-localisation with CIZ1, the two EZH2 antibodies gave different staining patterns (**Figure 5.7A**). CST D2C9 gave a nucleus-wide staining pattern, excluded from nucleoli but not enriched at any other structure. STJ112944 had a similar pattern to CST D2C9 in the adapted WT cells where EZH2p is absent but appeared to be enriched in the perinucleolar region in the primary WT cells (**Figure 5.7A**). Nucleolar location was confirmed using co-staining for nucleolar marker Fibrillarin (**Figure 5.7B**). Perinucleolar enrichment of EZH2 has previously been described, along with SUZ12 (another PRC2 subunit) and Snf2h (component of the ACF1-ISWI chromatin remodelling complex known to be involved in heterochromatin replication) (Zhang et al., 2007, Chen et al., 2008, Zhao et al., 2008).

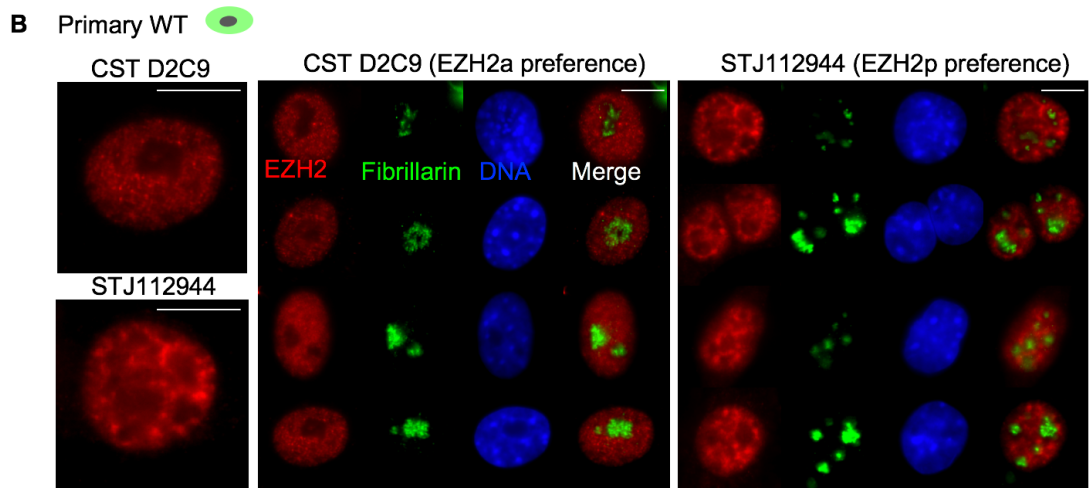
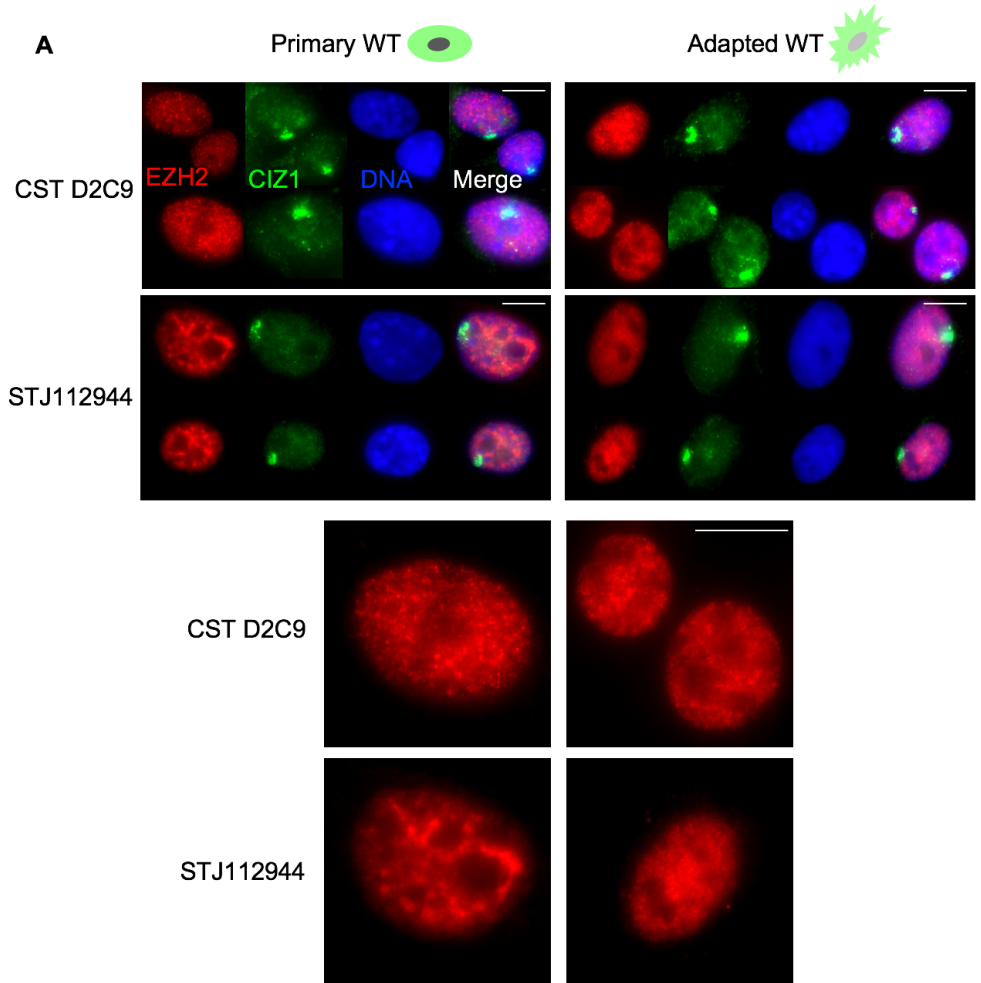


Figure 5.7 Differential isoform detection with EZH2 antibodies

A) Immuno-detection of CIZ1 (hC221a) and EZH2 (CSTD2C9 and STJ112944) in primary (14.4, p3) and culture adapted (13.8, p23) WT murine embryonic fibroblasts. Below, high-magnification images. B) Immuno-detection of EZH2 with two different EZH2 antibodies (CSTD2C9 and STJ112944), and of the nucleolar protein fibrillarin in primary mouse embryonic fibroblasts (14.4 p3) showing distinct EZH2 staining patterns, and peri-nucleolar enrichment with STJ112944. Scale bar is 10 microns.

5.3.4 EZH2p is resistant to extraction

As the EZH2 isoforms localise differently, nuclear matrix extraction was performed with detection by both immunofluorescence and western blot to determine whether the change in isoform affects anchorage (**Figure 5.8A**). Detection was performed with CST D2C9 as this antibody is capable of detecting both isoforms when present (**Figure 5.6A**).

In the following description, numerals in brackets refer to biochemical fractions described in (**Chapter 3**). Both EZH2p and EZH2 α are resistant to detergent (ii) and 0.5M salt extraction (iv) in WT primary cells, however approximately half of the EZH2 α expressed in culture adapted cells of both genotypes is solubilised by 0.5M salt (**Figure 5.8B**). This shows that although EZH2 α is capable of binding to nuclear components in a salt-resistant manner, approximately half of the EZH2 α is soluble, possibly reflecting a saturation of binding sites due to the increased levels (**Figure 5.1**). Quantification of immunofluorescence signal intensity confirms these findings, with signal intensity dropping by approximately 50% following 0.5M salt extraction in culture adapted WT cells (**Figure 5.8C**).

Further extraction in primary cells shows that EZH2p is in fact resistant to all extraction methods (vi, viii), while the remaining EZH2 α is largely solubilised by digestion of nuclear RNA (v, vii, **Figure 5.8D**), confirmed by quantification of immunofluorescence intensity (**Figure 5.8C, E**). Previous analyses of EZH2 in embryonic stem cells found EZH2 to be bound to chromatin and RNA (Beltran et al., 2016), more in line with findings for EZH2 α . However the exact EZH2 antibody used in these analyses is not stated and, as shown previously (**Figure 5.6A**), different EZH2 antibodies have different preferences for the two isoforms.

Overall these analyses show the shift in isoform to EZH2 α produces a less spatially constrained enzyme (**Figure 5.8F**), possibly resulting in the loss of enrichment from the peri-nucleolar region where EZH2p is anchored in cells that are in the primary state.

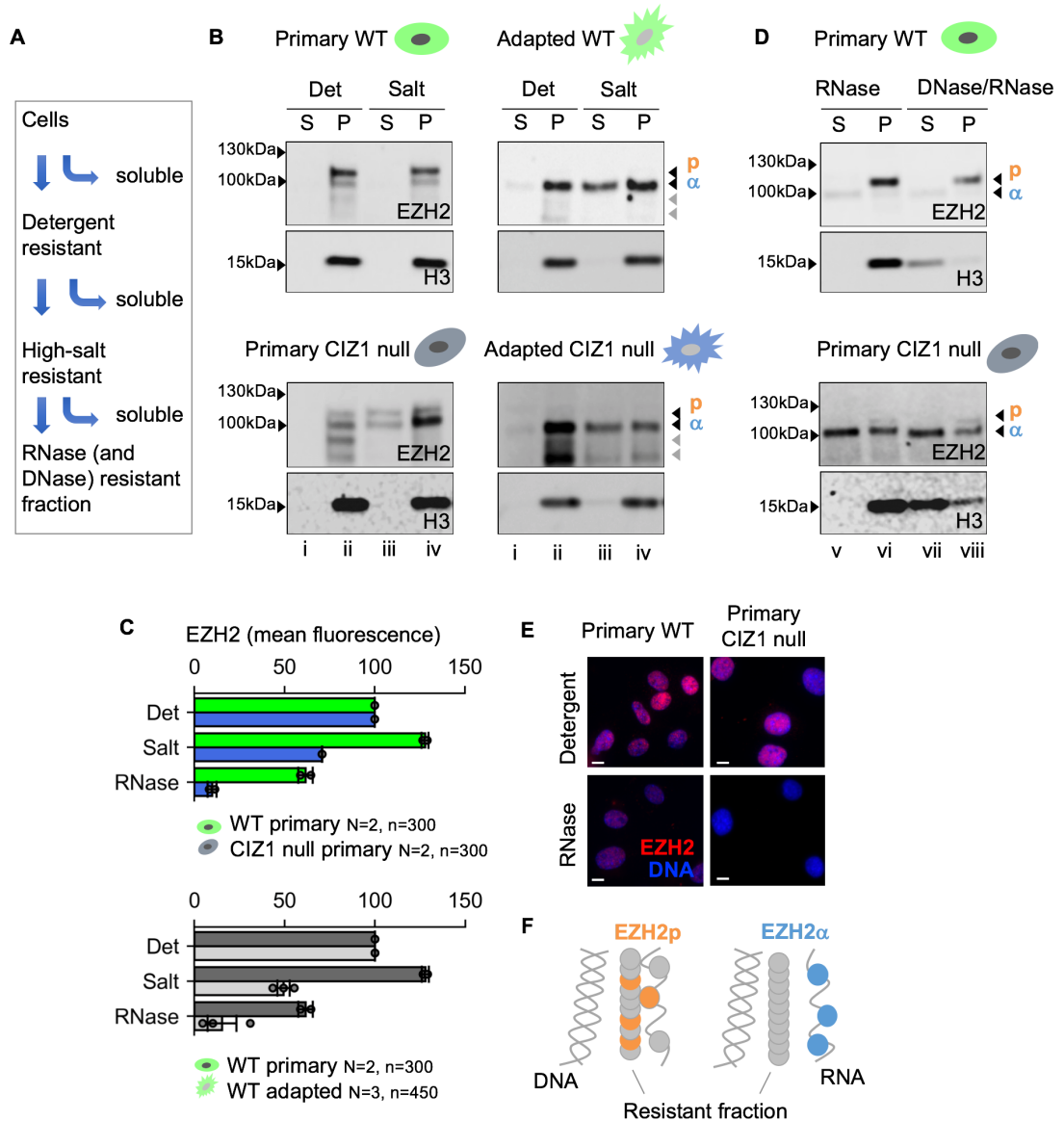


Figure 5.8 Loss of CIZ1 drives changes in EZH2 binding

A) Schematic showing sequential extraction with detergent-containing buffer (0.1% Triton X100), high-salt buffer (0.5 M NaCl), and nucleases (RNase and/or DNase1). B) Representative western blots showing partitioning of EZH2p and EZH2 α (antibody D2C9) into soluble (S) and pellet (P) fractions, after incubation with detergent (Det) or with detergent plus 0.5 M NaCl (Salt). All fractions within an extraction series are cell-equivalents. Results are shown for primary WT (13.8 p3) and CIZ1 null cells (13.17 p5) and culture-adapted derived populations (p28–29) of each genotype. C) Quantification of EZH2 immuno-stain in two primary cell populations of WT (13.1, 13.8) and CIZ1 null (13.17, 14.2) genotypes after the indicated extraction steps. Histogram shows mean fluorescence intensity normalized to detergent-treated samples, \pm SEM, where $n > 50$ for each genotype and condition, and $N = 2$. Below, the same analysis comparing two WT primary cells and three WT culture-adapted cells (13.1, 14.4, 3T3 cells). Individual data points contributing to the mean are overlaid. D) The salt-resistant pellet fraction (iv) from primary cells, was further partitioned by digestion to solubilize RNA, or both RNA and DNA as indicated, revealing complete nuclease-resistance of EZH2p in WT cells, despite efficient release of histone 3 (H3) into the soluble fraction. In contrast, in CIZ1 null cells (expressing EZH2 α) more than half is sensitive to solubilisation with RNase. E) Images show EZH2 (red, D2C9) after detergent-extraction and after RNase-extraction. DNA is blue, scale bar is 10 microns. F) Illustration of the differing modes of attachment of EZH2 α (released along with RNA), and EZH2p (immobilized by attachment to RNase and DNase-resistant assemblies).

5.3.5 EZH2 and CIZ1 are in close proximity in primary WT cells

Though there was no obvious colocalization between EZH2 and CIZ1 by standard immunofluorescence (**Figure 5.7A**), the nuclear matrix anchorage of EZH2p in the peri-nucleolar region, and the transient occupation of a peri-nucleolar zone by Xi-CIZ1 during Xi translocation in S phase, poses the hypothesis that only a small subpopulation of CIZ1 and EZH2 meet when functionally required. To identify these events, proximity a ligation assay was performed using antibodies against CIZ1 and EZH2 in WT primary and culture adapted cells, with primary CIZ1 null cells used as a negative control (**Figure 5.9**).

In correlation with the previous data, the average number of 'spots', where CIZ1 and EZH2 were found to be within 40nm of each other, is reduced in adapted cells and these spots are more likely to occur on the nuclear periphery, consistent with the absence of chromosomal relocation in these cells.

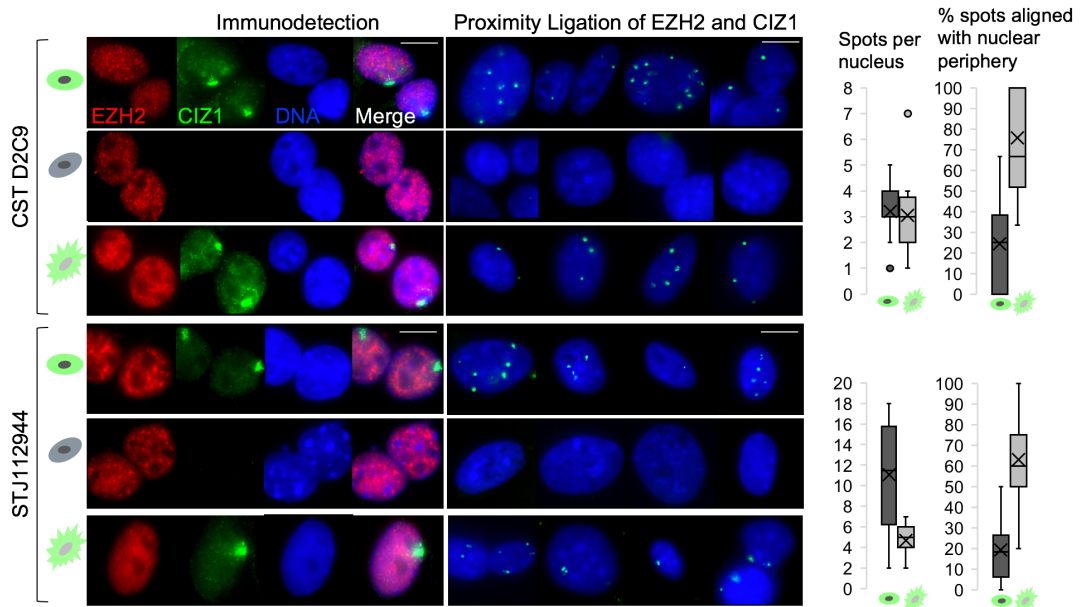


Figure 5.9 Proximity Ligation Assay between CIZ1 and EZH2

Immuno-detection (left) and proximity ligation (middle) of two different EZH2 antibodies (CSTD2C9 and STJ112944) with CIZ1 (hC221) in primary WT (14.4 p3), primary CIZ1 null (13.15 p2) and culture adapted WT (13.8 p23) embryonic fibroblasts. For EZH2 image intensity is differentially adjusted to reveal pattern rather than relative signal intensity. Scale bar is 10 microns. Right, quantification of proximity ligation assay, conducted by Dawn Coverley, showing the number of spots per nucleus ($n = 20$) and the percentage of spots aligned with the nuclear periphery for the primary and culture adapted WT cells, for both EZH2 antibodies. X represents the mean, the mid-line shows the median, the top and bottom line of the box the 3rd and 1st quartiles, and the whiskers the minimum and maximum values, excluding outliers (points, greater than 1.5 times the interquartile range).

5.3.6 Re-establishment of epigenetic marks in culture adapted cells

We have determined that the change in isoform of EZH2 produces a less spatially constrained enzyme (**Figure 5.8**), and the emergence of this isoform coincides with loss of chromosomal relocation (**Figure 5.1**) and deregulation of target genes (**Figure 4.6**).

Strikingly, following culture adaption of CIZ1 null cells, in which H3K27me3 and H2AK119Ub1 at the Xi are normally absent, both marks began to re-immerge (**Figure 5.10**), sometimes with a more dispersed pattern than in WT cells. This suggests that the dependence on chromosomal relocation for maintenance of these epigenetic marks is overridden during prolonged culture by elevation and solubilization of the enzymes, resulting in the changes seen in PRC target gene expression (in both genotypes) following culture adaption.

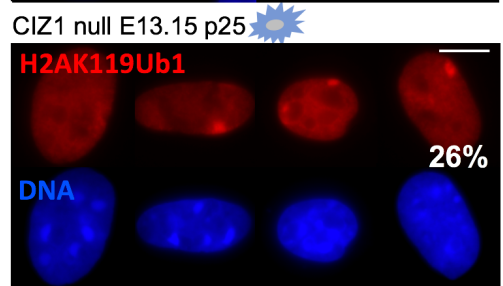
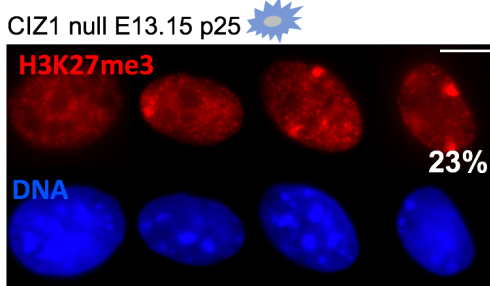
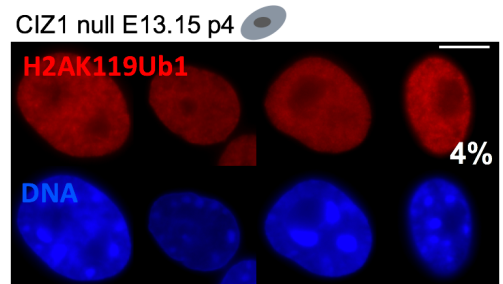
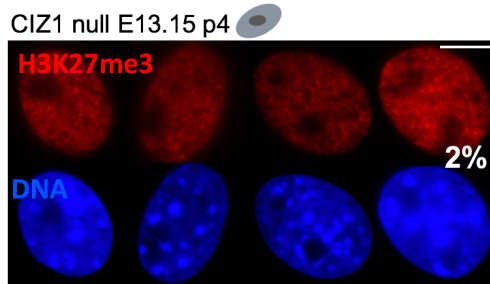
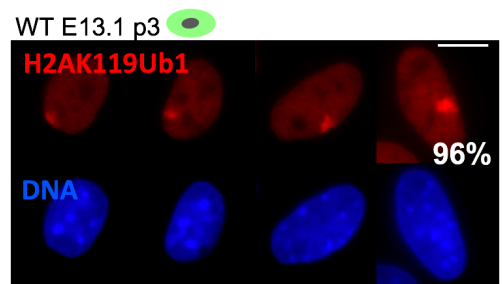
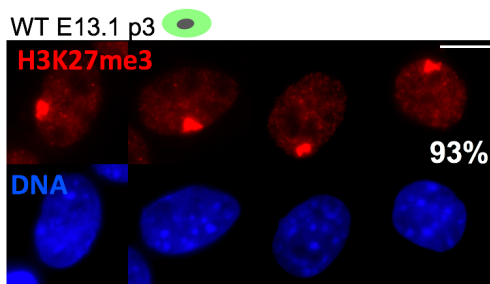
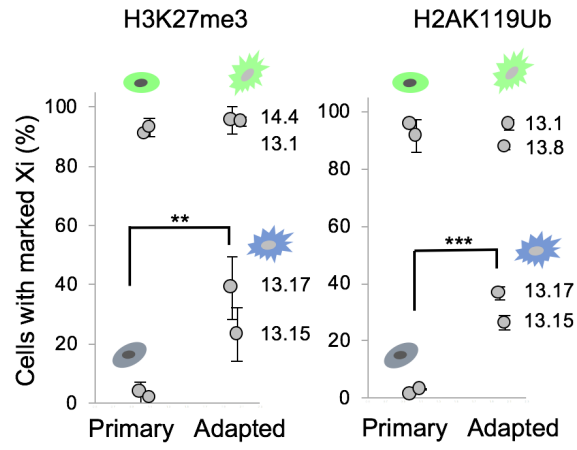


Figure 5.10 Reinstatement of epigenetic marks in culture adapted cells
Representative immuno-stain of PRC1-mediated H2AK119Ub1 and PRC2-mediated H3K27me3 in primary WT cells (13.1, p3), absence of these marks in primary CIZ1 null cells (13.15, p4), and re-emergence in culture-adapted CIZ1 null MEFs (13.15, p25). Percentage of cells with marked Xis is shown. Scale bar is 10 microns. Above, quantification by Robert Turner showing the proportion of primary and culture-adapted cells with H3K27me3-marked Xis (left) and H2AK119Ub-marked Xis (right), in CIZ1 null (blue) and WT (green) genotypes, \pm SEM from triplicate analyses for each line. Significant changes upon culture adaptation are indicated and compared by t-test. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

5.3.7 Proposed model

Based on these findings we propose a model (**Figure 5.11**) whereby, over the process of culture adaption or loss of CIZ1, chromosomal movement is lost, resulting in reduced capability to maintain H3K27me3 and H2AK119Ub1. To compensate for the loss, a more soluble isoform of EZH2 is expressed. There could be selective pressure towards cells which are capable of maintaining methylation as H3K27me3 is known to be involved in silencing repetitive sequences such as transposable elements (Ishak et al., 2016), which could lead to genomic instability if de-repressed. This more soluble EZH2 enzyme re-instates H3K27me3 across the genome in a less targeted manner, leading to deregulation of targets and a step down in epigenetic control.

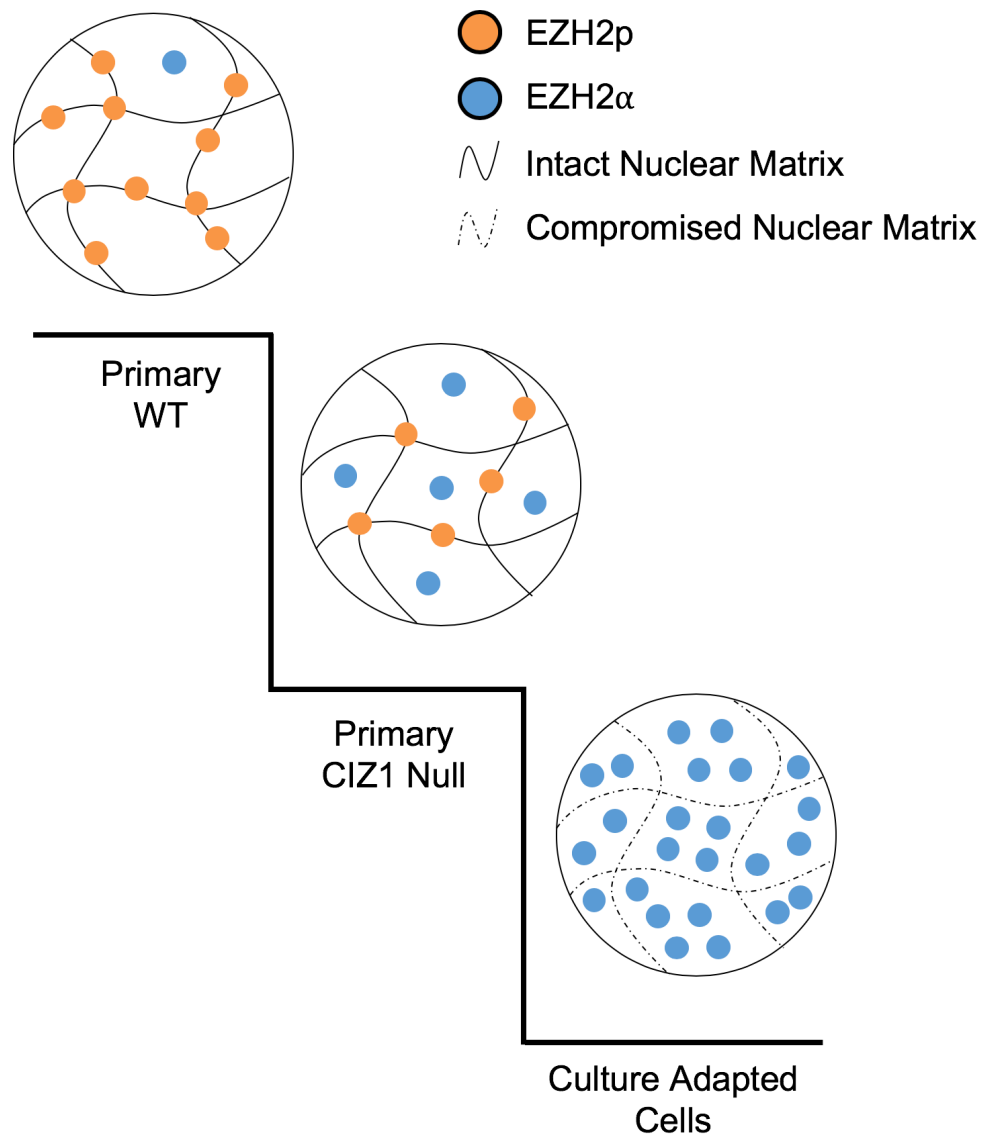


Figure 5.11 Model of step down of epigenetic control through adaption
 Model illustrating two drivers of the shift away from immobilized EZH2p (primary WT cells) towards the expression of high levels of non-immobilized EZH2α (culture-adapted cells). CIZ1 null primary cells represent an intermediate state.

5.4 Conclusions

CIZ1 anchors the inactive X chromosome during its replication, coinciding with Xi relocation to the nucleolus and back. Following culture adaption chromosomal relocation is lost, EZH2 is released from spatial constraint and PRC target expression is deregulated. Primary CIZ1 null cells are already displaying many of the same phenotypes, showing that loss of CIZ1 mimics the effects of culture adaption. Crucially, this shows that investigations into the role of CIZ1 must be performed in primary cells.

6.0 Loss of CIZ1 in young adult splenic lymphocytes

6.1 Introduction

Loss of CIZ1 in mice yielded no obvious embryonic phenotype. However, 100% of CIZ1 null adult females displayed a fully penetrant lymphoproliferative disorder, pathologically consistent with B-cell lymphoma with T-cell infiltration (Ridings-Figueroa et al., 2017). This is the first time CIZ1 has been linked to lymphoma but is consistent with a previous mouse model linking CIZ1 loss with leukaemia susceptibility (Nishibe et al., 2013), and well-established links between CIZ1 and other solid tumours (Swarts et al., 2018), including cancers of the lung (Higgins et al., 2012, Coverley et al., 2016), prostate (Liu et al., 2015), liver (Lei et al., 2016, Wu et al., 2016), gallbladder (Zhang et al., 2015) and Ewing tumours (Rahman et al., 2007).

Previous studies have indicated lymphocytes are particularly susceptible to perturbations in XCI. Studies have shown that females have immunological advantage over males and an increased incidence of autoimmune diseases (Eaton et al., 2007, Libert et al., 2010). In addition, there is a higher proportion of immunity-related genes on the human X chromosome than autosomes (Ross et al., 2005). *Xist* itself has been identified as a haematological tumour suppressor, with conditional loss of *Xist* in murine haematopoietic stem cells resulting in female specific lethality between 1.5 months and 2 years of age and hyperproliferation of all haematopoietic lineages (Yildirim et al., 2013, Agrelo et al., 2009).

While the exact mechanism of this sensitivity is unknown, it is thought to be linked to the unique method of XCI maintenance in lymphocytes (Wang et al., 2016, Ridings-Figueroa et al., 2017, Syrett et al., 2019). In naïve splenocytes derived from 6-week old female C57BL/6 mice, *Xist* is dispersed and there is no enrichment of CIZ1 at the inactive X chromosome. However, both return to the Xi upon receipt of activation signals (**Figure 6.1**) and return of *Xist* is dependent on CIZ1 (Ridings-Figueroa et al., 2017).

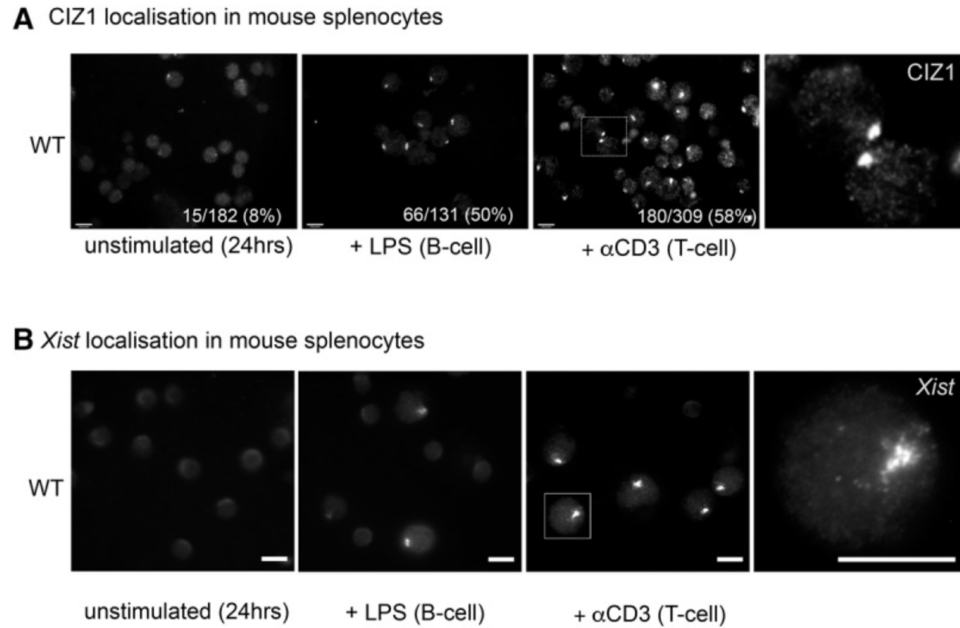


Figure 6.1 Recruitment of CIZ1 and Xist to the Xi upon activation

Figure from (Ridings-Figueroa et al., 2017) showing recruitment of CIZ1 (A) and Xist (B) to the inactive X chromosome following lymphocyte activation.

This suggests a second round of XCI occurs in these cells, which could (i) predispose females to partial reactivation of the Xi in these lineages and/or (ii) be compromised in the absence of CIZ1. This could result in aberrant over-expression of a subset of genes on the Xi, including some immunity-related genes, which could explain the increased incidence of autoimmune diseases in females.

6.2 Aims

The aims of these experiments were to:

- Identify candidate pathways affected by loss of CIZ1 in naïve and activated lymphocytes and spleens
- Identify possible avenues of investigation to determine the cause of lympho-proliferative disorder
- Identify possible reasons for the female specificity of the CIZ1 null phenotype

6.3 Experimental Design

To identify the effect of loss of CIZ1 on splenic lymphocyte activation, and how these changes could lead to the lymphoproliferative disorder phenotype, RNA-sequencing was performed on naïve and activated lymphocytes and snap-frozen spleen samples.

Lymphocytes were isolated from spleens of two 6-week old WT and two CIZ1 null female mice by Justin Ainscough, as described in (Ridings-Figueroa et al., 2017). The CIZ1 null mice showed no obvious signs of lympho-proliferative disorder. The lymphocytes were cultured for 24 hours either without activation (naïve) or in the presence of lipo poly-saccharides (LPS) to activate B cells, or α -CD3 to activate T cells, before I isolated the RNA (**Figure 6.2**).

Spleens were previously isolated by Justin Ainscough from adult (>9-month old) WT and CIZ1 null female mice and snap frozen. All CIZ1 null mice were displaying lympho-proliferative disorder. RNA was extracted from small sections (~20mg) of two WT and two CIZ1 null spleens by myself, supervised by Justin Ainscough (**Figure 6.2**).

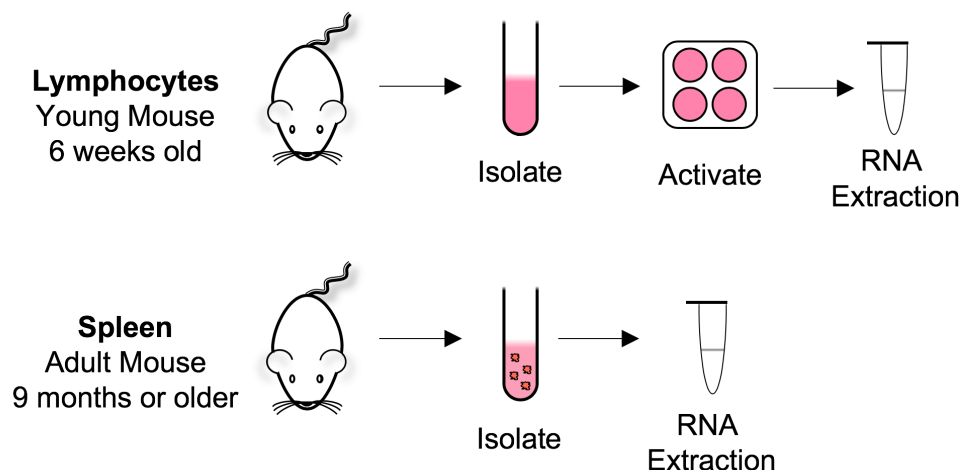


Figure 6.2 Schematic of RNA isolation from lymphocytes and spleens
Lymphocytes and whole spleens were isolated from WT and CIZ1 null mice. Lymphocytes were grown in culture either without stimulus or in the presence of B or T cell activating stimuli for 24 hours before RNA was extracted. Spleen RNA was extracted directly from small sections of adult mouse spleen tissue.

RNA quality was analysed at the University of York Genomics Facility and libraries generated as described (**Chapter 2.6.2**) for sequencing at the Leeds Institute for Molecular Medicine (LIMM) using the Illumina HiSeq 3000 system.

Sequence reads were processed by Katherine Newling as stated in (**Chapter 2.6.3**) to assemble a transcriptome from which gene expression differences between genotypes and activation status could be investigated. Isolation of RNA of a single WT and a single CIZ1 null spleen was performed previously by Rebeca Ridings-Figueroa, and sequenced separately. All data were combined and analysed identically to assemble a single transcriptome from 3 replicates for spleen. Principle component analysis performed by Katherine Newling showed no preferential grouping by experiment implying no batch effect.

6.4 Results and Discussion

6.4.1 Generation of RNA

Quality of RNA was assessed by Bioanalyzer (**Figure 6.3**). A rRNA ratio [28S/18S] of above 1.8 is consistent with an undegraded RNA sample.

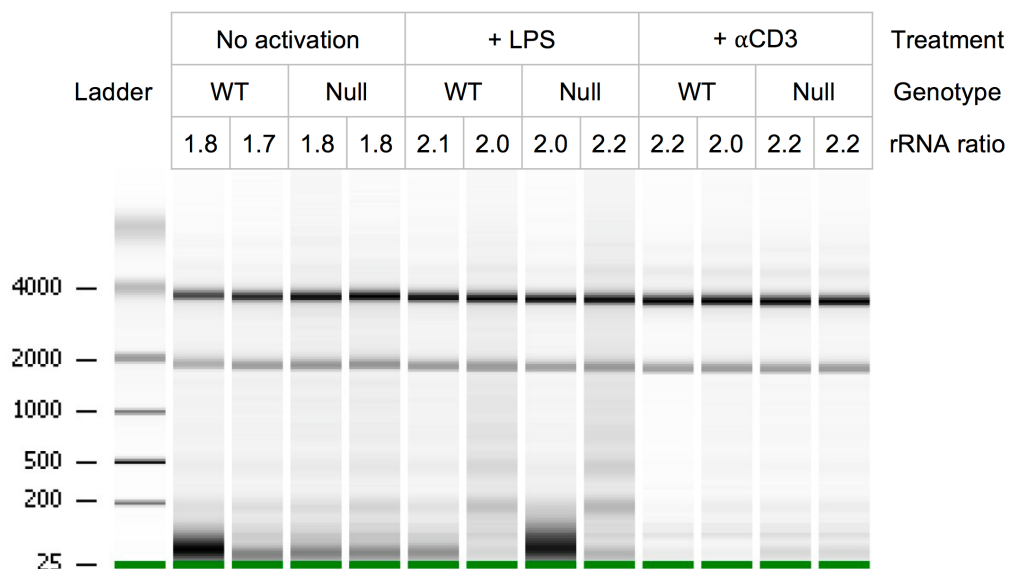


Figure 6.3 Analysis of lymphocyte RNA quality

Bioanalyzer results of all 12 lymphocyte RNA samples including the rRNA ratio [28S/18S] showing undegraded RNA for sequencing.

6.4.2 Effects of CIZ1 loss on gene expression from the X Chromosome

Due to the female specific nature of the lympho-proliferative disorder, the effect of both activation, and CIZ1 loss, on expression from the X chromosome was analysed. Activation of B cells had very little effect on gene expression from the X chromosome (**Figure 6.4**). Activation of T cells had a much greater effect, but was very similar between WT and CIZ1 lymphocytes. WT and CIZ1 null cells of the same activation state were then compared. As expected, there was little difference between WT and null lymphocytes in the naïve state (N), and once the lymphocytes had been activated (B) and (T). In fact, no CIZ1-dependent genes were identified in the naïve state ($q < 0.05$).

There was a much greater difference between WT and CIZ1 null spleens (Sp), showing some deregulation of expression from the X chromosome during disease. However, the 50 significantly changed genes ($q < 0.05$, **Appendix G**) represent less than 2% of genes located on the X chromosome, demonstrating no greater effect of loss of CIZ1 on the X chromosome than the genome as a whole. These could be further investigated on an individual basis to attempt to identify a rationale for the female specificity of the phenotype, however my further analysis has focussed on looking at genome-wide expression changes.

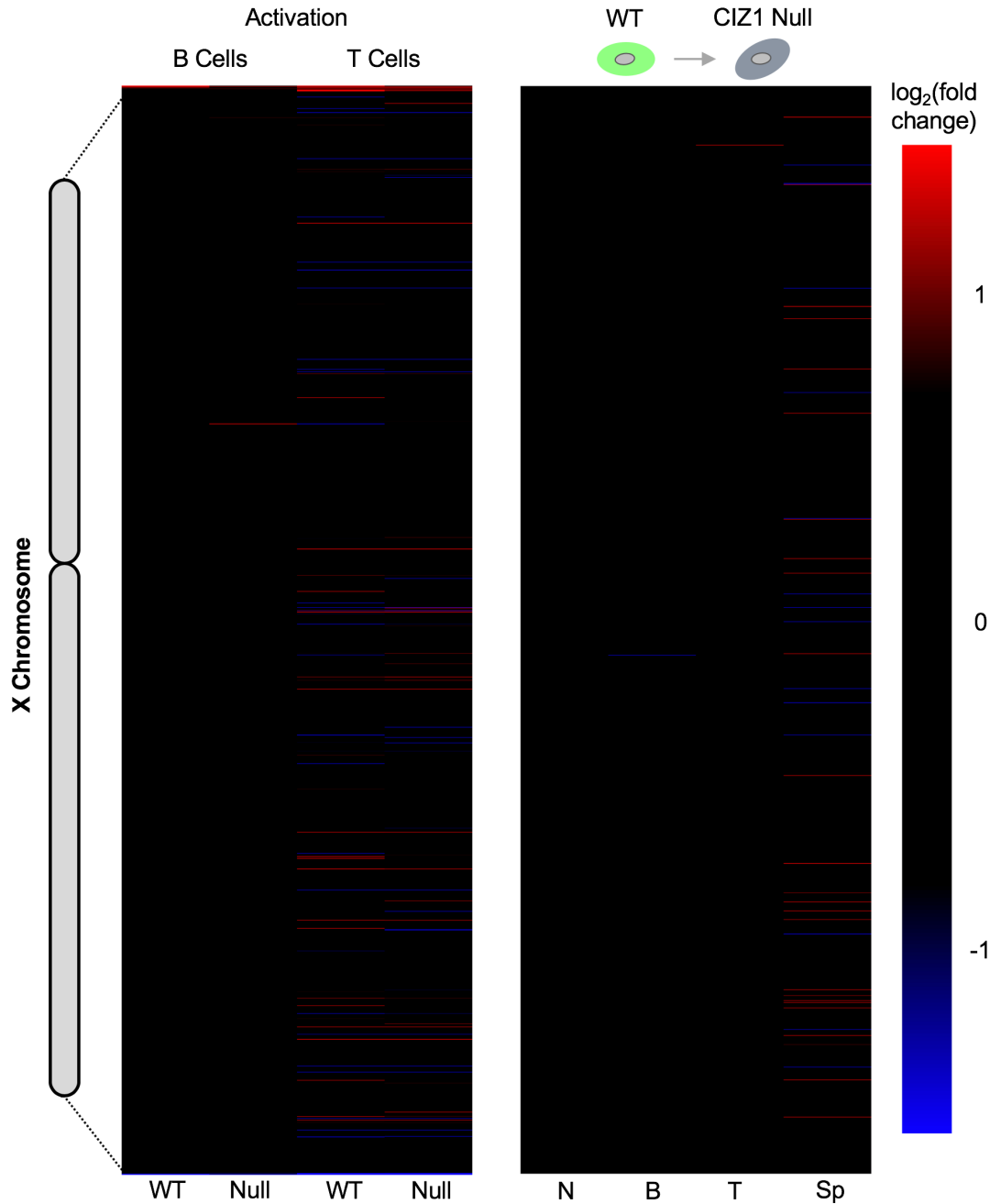


Figure 6.4 Effect of CIZ1 loss on expression from the X chromosome
 Heatmap showing the $\log_2(\text{fold change})$ in expression of genes from the X chromosome ordered by location on the chromosome. (Left) Effect of lymphocyte activation on WT and CIZ1 Null lymphocytes. (Right) Effect of CIZ1 loss on naïve (N), B cell activated (B) and T cell activated (T) lymphocytes and spleens (Sp).

6.4.3 Genome wide effects of CIZ1 loss

Volcano plots showing genome-wide expression in naïve CIZ1 null lymphocytes, compared to WT, revealed a small set of genes that are compromised even prior to activation (**Figure 6.5A**). More than 75% of these were upregulated in the absence of CIZ1. These expression changes were generally conserved upon activation, with 60% of CIZ1-dependent genes in activated B cells and 50% in activated T cells being genes which were already significantly changed in CIZ1 null lymphocytes before activation (**Figure 6.5B**).

61 of the 79 significantly different genes ($q < 0.05$) in CIZ1 null compared to WT lymphocytes were annotated. A Gene Set Enrichment Analysis (GSEA) using the MSig Database Gene Ontology terms showed strong correlation with immune and inflammatory response pathways (**Figure 6.5C**), suggesting that CIZ1 null lymphocytes could be in a primed state of activation in the absence of stimuli.

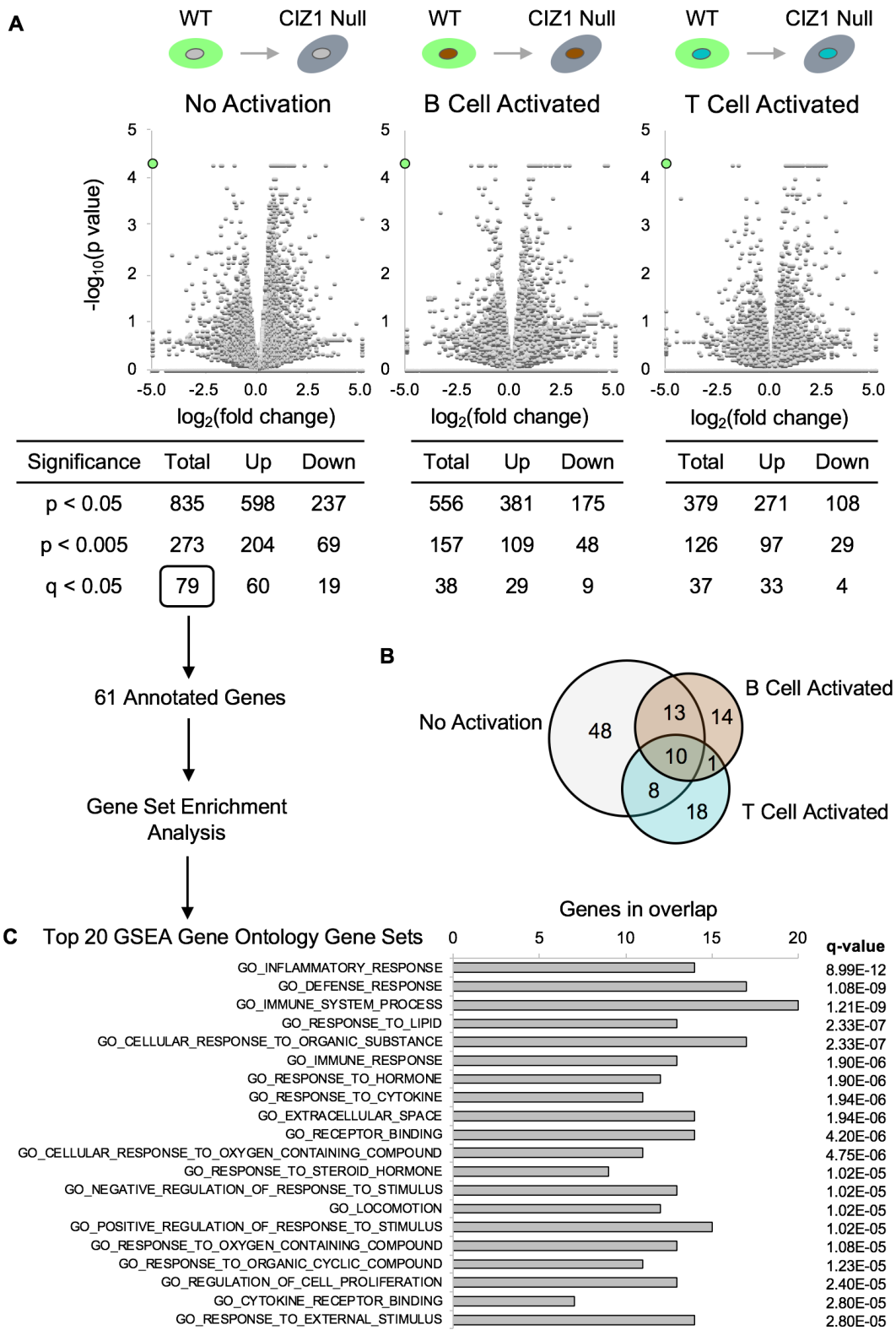


Figure 6.5 CIZ1 loss results in up-regulation of immunity related genes in lymphocytes in the absence of activation signals

(A) Volcano plots comparing genome-wide expression between CIZ1 null and WT lymphocytes, in the absence of activation signals (left), upon activation of B cells (centre) and upon activation of T cells (right). Number of genes changed to three significance cut offs are shown beneath each comparison. CIZ1 expression is shown in green. (B) Venn diagram showing overlap between significantly changed genes ($q < 0.05$) upon loss of CIZ1 in the absence of activation signals (79 genes, grey), upon B cell activation with LPS (38 genes, brown) and upon T cell activation with α -CD3 (37 genes, blue). (C) Of the 79 genes significantly affected by loss of CIZ1, Gene Set Enrichment Analysis was performed on the 61 annotated genes using the MSig Database Gene Ontology terms. The top 20 overlaps are shown, alongside the significance (q-value) of these overlaps.

Specific activation of WT B and T cells (**Figure 6.6A**) had a much greater effect on gene expression than loss of CIZ1, providing a set of gene expression changes in response to controlled activation against which loss of CIZ1 can be compared (**Figure 6.6B**). The dramatic difference in number of affected genes (1052 B; 4899 T; compared to 79 CIZ1-dependent) is not unexpected as directed activation results in large-scale cellular reprogramming and remodelling events (Waters et al., 2018, Wang et al., 2011). Interestingly, T cell activation affected >4 fold the number of genes affected by B cell activation.

These WT activation gene sets were then compared with the 79 CIZ1-dependent genes in naïve lymphocytes to investigate whether CIZ1 null lymphocytes are partially activated even in the absence of stimuli (**Figure 6.6C**). Of the 79 CIZ1-dependent genes, 61 (77%) overlapped with either of the WT activation sets, with the strongest overlap with the genes significantly changed upon T cell activation. Interestingly, 81% of the genes significantly changed upon B cell activation are also changed upon T cell activation, suggesting that there is a large core set of genes involved. However, this analysis did not discriminate between direction of change, so this overlap could contain reciprocal changes (e.g. a gene upregulated upon CIZ1 loss but downregulated upon lymphocyte activation).

Indeed, only 20% of the CIZ1-dependent genes which overlap with B or T cell activation were changed in the same direction (**Figure 6.6D**). In general, genes upregulated in CIZ1 null lymphocytes are down-regulated upon B or T cell activation. This explains conservation of CIZ1-dependent genes across activation status seen in (**Figure 6.5B**) as over 80% CIZ1 dependent genes which remain changed in CIZ1 null lymphocytes following activation of B or T cells respectively are consistently upregulated.

Therefore CIZ1 loss seems to result in persistent overexpression of genes which would typically be down-regulated upon lymphocyte activation.

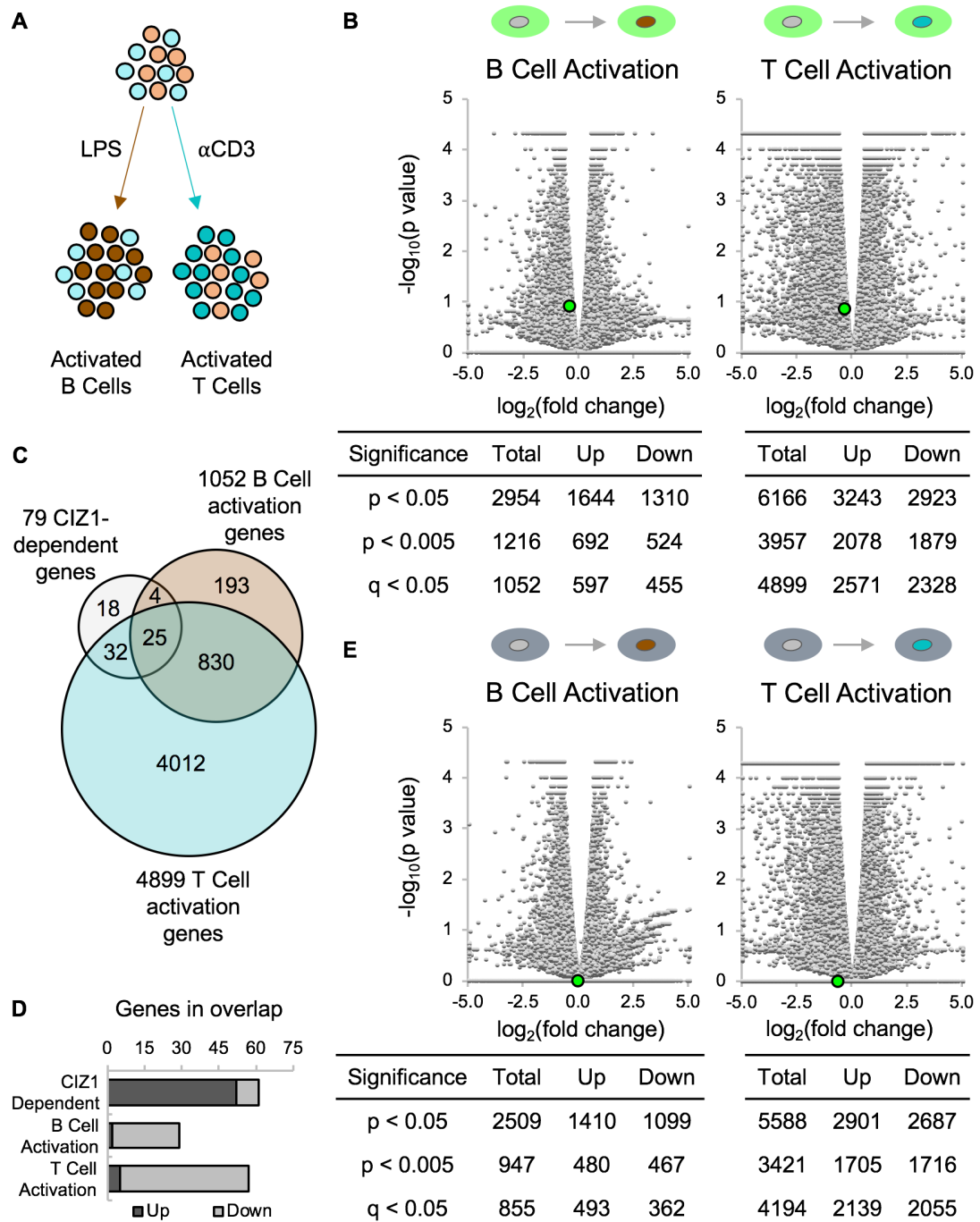


Figure 6.6 Effect of activation on WT and CIZ1 null lymphocytes

(A) Specific activation of either B or T cells using LPS or α -CD3. (B) Volcano plots comparing genome-wide gene expression in WT lymphocytes where either the B (left) or T (right) cells have been activated. Genes changed to certain significance cut offs are shown beneath each comparison. CIZ1 expression is shown in green. (C) Overlap between the 79 CIZ1-dependent genes and genes changed upon activation of WT B or T cells. (D) Direction of overlaps between CIZ1-dependent genes and activation genes shown in (C). (E) Volcano plots as (B) but for CIZ1 null lymphocytes.

A potential cause for the lympho-proliferative disorder in CIZ1 null females could arise from differences in response to activation. Indeed, CIZ1 null lymphocyte activation resulted in slightly fewer significantly changed genes than WT cells (**Figure 6.6E**). However, there was no overwhelming bias towards up- or down- regulation; similar to activation of WT lymphocytes. Although the number of significantly changed genes was similar between WT and CIZ1 null activating lymphocytes, this analysis does not reveal whether the same genes are changed, or if they are changed in the same direction.

6.4.4 Identification of activation gene sets for further analysis

Significantly changed genes ($q < 0.05$) were split into up- and down- regulated lists for WT and CIZ1 null lymphocytes for both B and T cell activation. WT and CIZ1 null gene lists for each condition were compared to identify genes which behaved differently upon activation (**Figure 6.7A**). This allowed determination of 4 categories:

- 1) Inappropriately upregulated. Genes upregulated in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes which are not upregulated in activated WT lymphocytes compared to naïve WT lymphocytes.
- 2) Failure to upregulate. Genes upregulated in activated WT lymphocytes compared to naïve WT lymphocytes which are not upregulated in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes.
- 3) Inappropriately downregulated. Genes downregulated in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes which are not downregulated in activated WT lymphocytes compared to naïve WT lymphocytes.
- 4) Failure to downregulate. Genes downregulated in activated WT lymphocytes compared to naïve WT lymphocytes which are not upregulated in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes.

Extent of overlap between B and T cell activation in each of these 4 categories was compared (**Figure 6.7B**). In general overlap was low, suggesting that CIZ1-dependent genes are cell type specific.

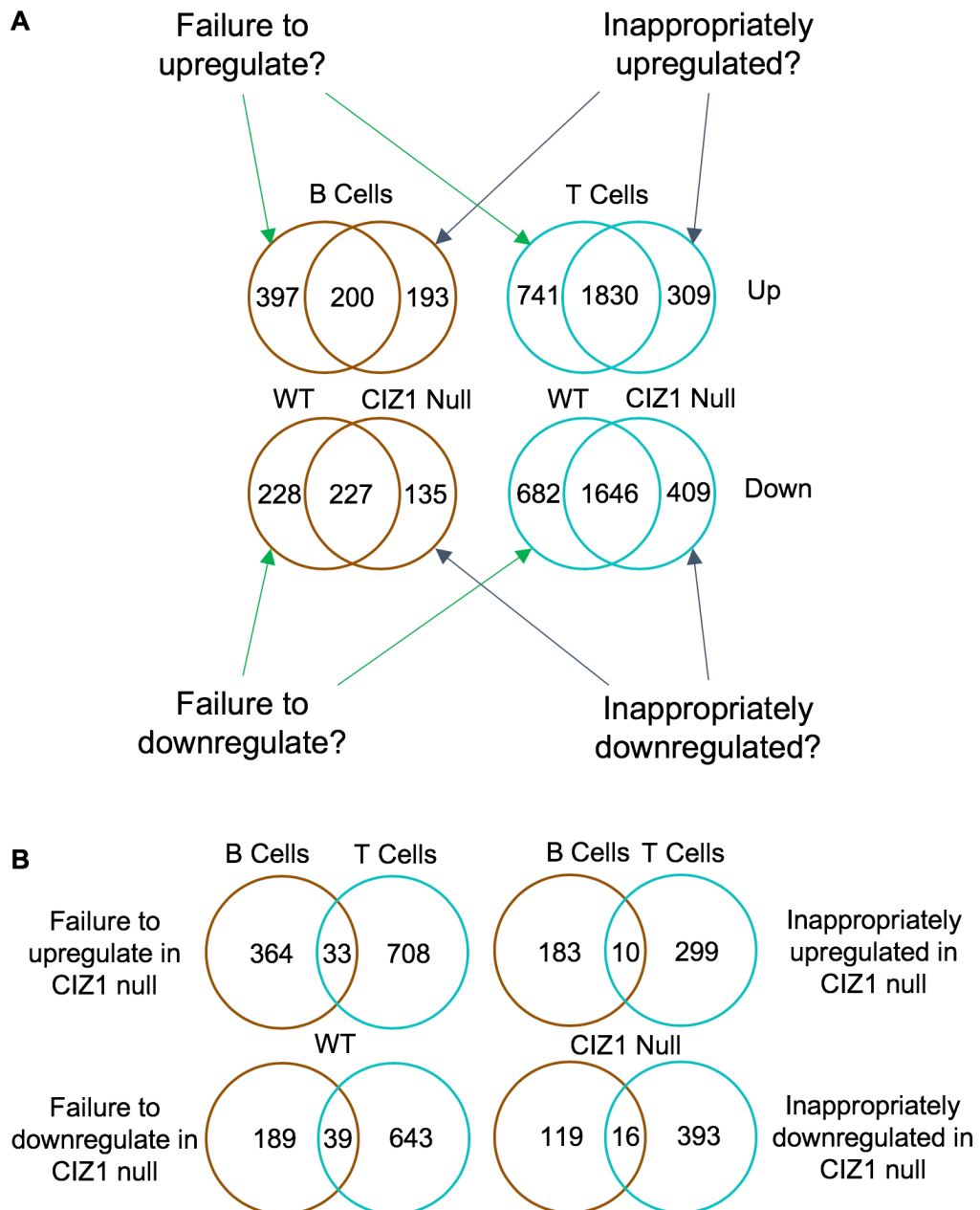


Figure 6.7 Identification of activation gene sets for further analysis

(A) Genes significantly changed ($q < 0.05$) upon lymphocyte activation for both genotypes were separated into up- and down-regulated gene lists and compared, giving four categories of genes for further analysis. (B) The genes in each of the four categories were compared between B and T cell activation to investigate the lineage specificity of these sets.

One way of identifying possible pathways upon which CIZ1 might impinge is to perform a gene set enrichment analysis (GSEA). However GSEA can only be performed on annotated (previously named) genes. Therefore any transcripts arising from an unannotated locus would be overlooked. The proportion of unannotated genes in the four identified categories was investigated for each cell type (**Figure 6.8**).

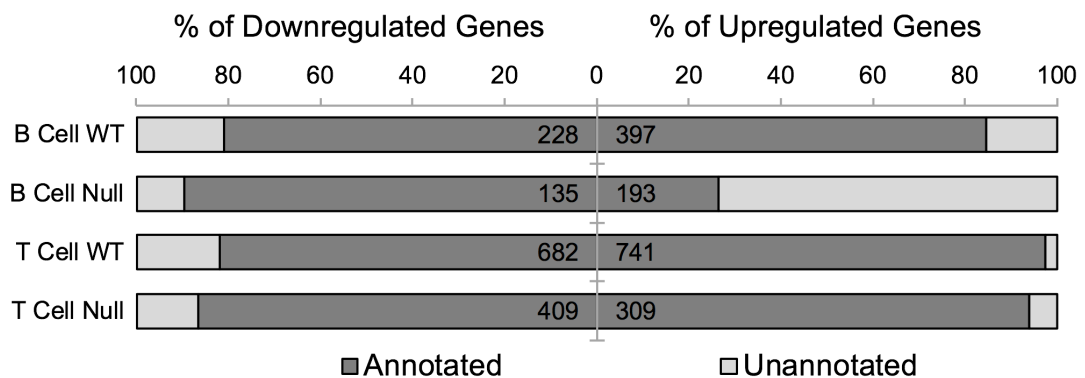


Figure 6.8 Unannotated transcripts in gene activation sets

Comparison of the proportion of genes in each category which are annotated and unannotated for both B and T cells. The number of genes in the category is shown at the base of each bar.

Generally, unannotated transcripts made up less than 20% of each gene set. However, they accounted for almost 75% of the inappropriately upregulated genes following B cell activation of CIZ1 null lymphocytes. Strikingly 100% of these upregulated unannotated transcripts had an infinite \log_2 (fold change) value, meaning there is no expression of any of these transcripts in the naïve CIZ1 null lymphocytes.

Therefore loss of CIZ1 appears to result in aberrant expression of unannotated transcripts specifically in the B cell lineage.

GSEA analysis was performed on the annotated genes significantly changed ($q < 0.05$) in each category using both the Curated Gene Sets and Gene Ontology collections of the MSig Database. This analysis was performed independently for B and T cell activation sets and the Top 20 Outputs from each comparison extracted (**Appendix H**). Gene sets common to both cell types were highlighted.

Notably, though there was little overlap between the gene lists derived from activation of the two cell types (**Figure 6.7B**), over half of the gene sets to which they belong are the same in the inappropriate upregulation category (**Table 6.1**). This implies that although the genes in this category may be different between cell lineages, they are part of the same pathways. Inappropriate upregulation is also a category of particular interest as loss of CIZ1 results preferentially in gene upregulation (**Figure 6.5A**).

Table 6.1 *The number of gene sets which appear in both B and T cell top 20 outputs of GSEA analysis for all four categories.*

Gene Category	Curated Genes Sets	Gene Ontology Sets
Inappropriate Upregulation	12/20	13/20
Failure to Upregulate	7/20	10/20
Inappropriate Downregulation	4/20	6/20
Failure to Downregulate	5/20	1/20

This analysis yielded interesting outputs, highlighting multiple possible avenues for further investigation. This included deregulation of pathways involved in **cell cycle progression and DNA damage response**, inappropriate downregulation of **proapoptotic genes**, and failure to downregulate **targets of transcriptional regulators**.

6.4.5 Deregulation of cell cycle genes in activated CIZ1 null lymphocytes compared to naïve CIZ1 null lymphocytes

Gene ontology analysis of the genes inappropriately upregulated upon activation of CIZ1 null lymphocytes showed strong overlap with sets involved in cell cycle regulation and DNA damage response (**Appendix H.1-2, Figure 6.9**). Interrogating the Curated Gene Set database returned overlaps with targets of the DREAM (dimerization partner, RB-like, E2F and multi-vulval class B) complex (Fischer et al., 2016), genes correlated with DNA damage repair protein BRCA1 and DNA damage checkpoint protein CHEK2 (Pujana et al., 2007), and genes upregulated upon loss of function of the cell cycle regulator protein RB1 (Markey et al., 2007) (**Appendix H.1-2, Figure 6.10A**). This implies inappropriate upregulation of genes involved in the cell cycle in the absence of CIZ1, particularly those involved in overcoming cell cycle

checkpoints. Interestingly, CIZ1 is already known to have a role in initiation of DNA replication thought to be mediated by direct interaction with cyclin A and CDK2 (Coverley et al., 2005, Copeland et al., 2010).

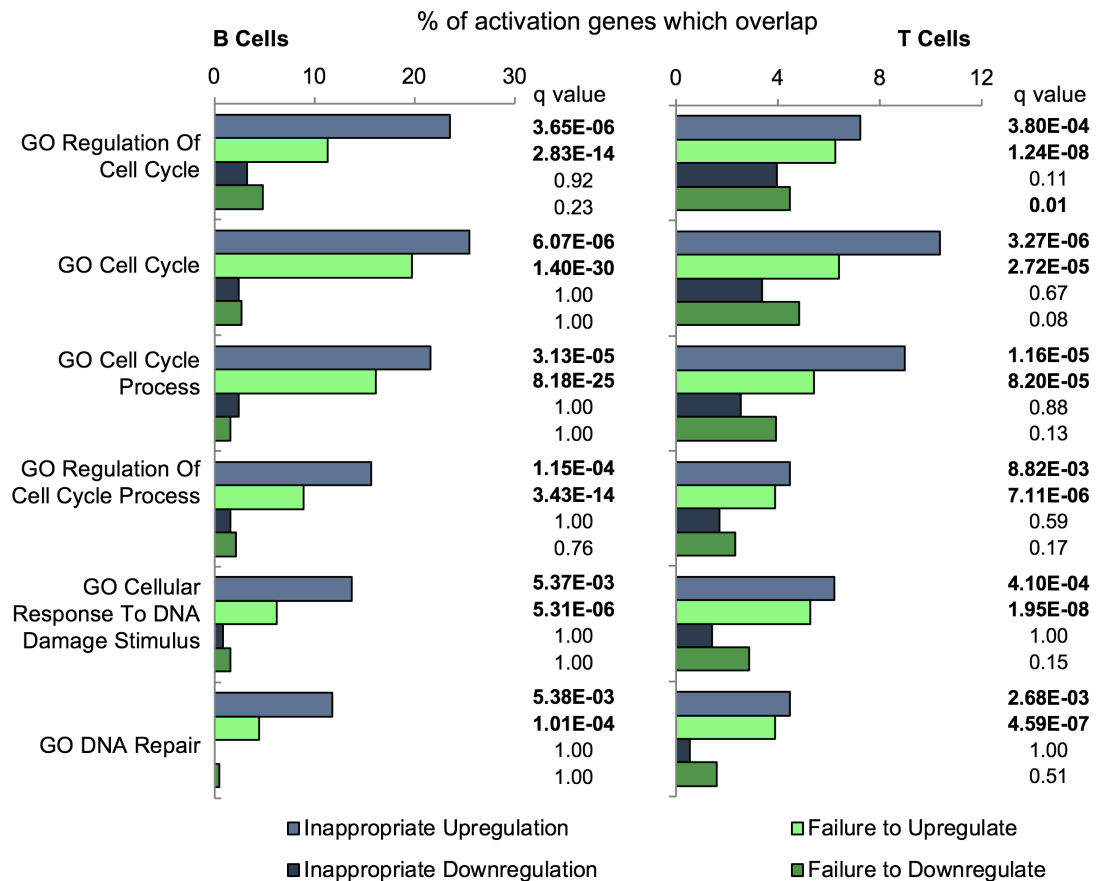


Figure 6.9 Inappropriate upregulation and failure to upregulate cell cycle and DNA damage response genes upon lymphocyte activation

The percentage of genes in each of the four categories which overlap with Gene Ontology gene sets in the MSig GSEA database relating to cell cycle and DNA damage response for B cell (left) and T cell (right) activation. The significance of these overlaps are shown to the right of each graph, significant q values are shown in bold.

However, the same gene sets also overlap significantly with genes in the failure to upregulate category (**Figure 6.9 and 6.10A**).

Therefore, more detailed analysis was carried out on 'Fischer DREAM Targets', 'Pujana BRCA1 PCC Network' and 'Markey RB1 Acute Loss of Function Up' to determine what is happening to these sets as a whole upon activation (**Figure 6.10B**). These sets were chosen as they were the highest overlapping and most significantly changed sets in each of the three processes of interest. I chose not to focus on the E2F4 set as it is a subunit of the DREAM Complex (Sadasivam and DeCaprio, 2013), or the CHEK2 network set as it is also involved in DNA damage response (Matsuoka et al., 2000) and therefore highly similar to the BRCA1 network set.

In all cases genes in these sets are preferentially upregulated in response to lymphocyte activation and in 5 of the 6 sets over 50% of the genes upregulated upon activation are upregulated in both genotypes. Upregulation of these sets is expected in these cells as naïve lymphocytes are quiescent (Yusuf and Fruman, 2003) and need to re-enter the cell cycle upon activation. Exit from and re-entry into cell cycle also involves nucleus-wide chromosomal rearrangement, which is likely to result in DNA damage which must be repaired (Mehta et al., 2010, Laporte et al., 2013).

Therefore CIZ1 loss appears to result in deregulation of processes which are already occurring in the cells upon transition, resulting in inappropriate upregulation of a subset of genes within a certain process undergoing general upregulation, while failing to upregulate another subset of genes within the same cellular process.

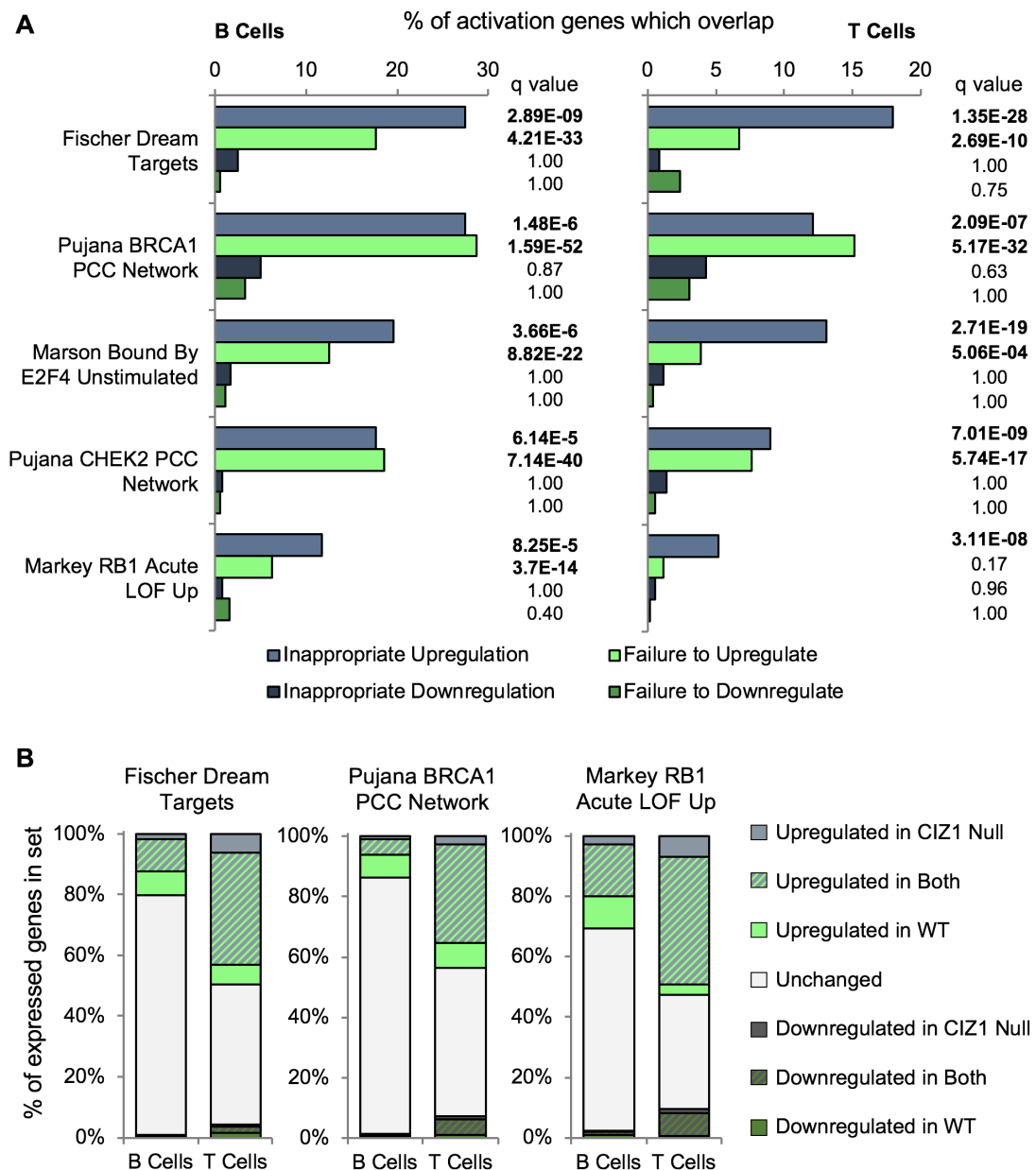


Figure 6.10 Upregulation of cell cycle and DNA damage response genes upon lymphocyte activation

(A) The percentage of genes in each of the four activation categories which overlap with Curated Gene Sets in the MSig GSEA database relating to cell cycle progression and DNA damage response for B cell (left) and T cell (right) activation. The significance of these overlaps are shown to the right of each graph, significant *q* values are shown in bold. (B) Detailed analysis of the effect of lymphocyte activation on all genes within three of the Curated Gene Sets seen in (A) which are expressed in WT naïve lymphocytes.

These analyses highlight the importance of investigating the direction of changes in gene expression. Previously, analysis of primary embryonic fibroblasts showed significant overlap between CIZ1-dependent genes and genes which are targets of the polycomb repressive complexes (PRC1/2) (**Chapter 4, Figure 4.6**, (Stewart et al., 2019)). However, directional analysis revealed that CIZ1 loss results in downregulation of previously identified direct PRC targets under conditions where the ability of these repressive complexes to stably modify chromatin is apparently impaired (**Chapter 4, Figure 4.7**). Thus, our original interpretation of these data was over-simplified, and assumed that loss of repressive marks on the X chromosome would translate to up-regulation of specific genes.

Taken together, the data are consistent with a hypothesis that during a biologically relevant transition, in the absence of CIZ1, a subset of genes which should escape regulation by a prevailing pathway are inappropriately targeted, and some genes which should be targeted are missed. These transitions could include differentiation in the embryo to generate the fibroblasts used in (**Chapter 4**), and entry into cell cycle as modelled by lymphocyte activation. In these transitions target selection would be expected to vary due to lineage specificity.

There are multiple possible mechanisms of how this could occur. For example, CIZ1 might be directly involved in controlling gene selection and targeting of regulatory complexes, by anchoring complexes at the site of their targets. Alternatively, CIZ1 could directly bind and block access to genes which should not be targeted. Indirect possibilities include recruitment of chromatin modifiers in such a way that complexes cannot access inappropriate target genes. Further focussed investigation is required into the effect of loss of CIZ1 on biological transitions. Insights might be gleaned from comparison of sequencing data from populations of WT and CIZ1 null ES cells prior to and after differentiation, or studying lymphocytes which have been sorted by FACS to select purer populations. Avenues of investigation could be assessed in cells to elucidate the mechanism, for example, looking for interaction between CIZ1

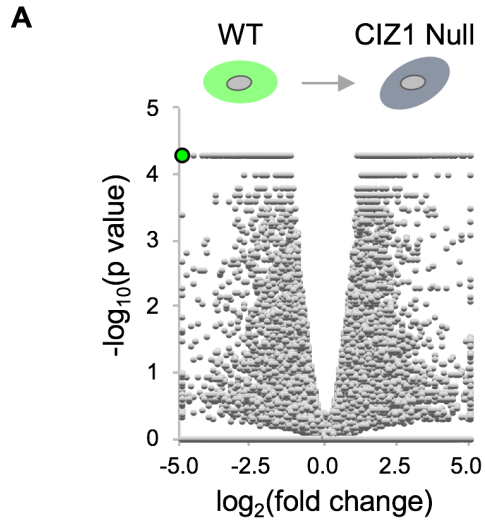
and putative affected complexes such as DREAM subunits, or by looking for occupancy of these complexes on affected genes promoters.

6.4.6 Deregulation of mitotic checkpoint genes in CIZ1 null spleens

To focus more closely on potential mechanisms most likely to underlie the lymphoproliferative disorder, expression differences between three WT and three CIZ1 null spleens from adult mice aged >9-months were investigated. All three of the CIZ1 null spleens displayed the diseased phenotype. If deregulation of a specific pathway was behind this, downstream consequences of that deregulation might be evident in the diseased spleens.

Volcano plots showing genome-wide expression changes in CIZ1 null compared to WT spleens revealed a much larger population of changed genes, compared to the equivalent populations in naïve lymphocytes (**Figure 6.11A**). Notably, however, preferential upregulation of genes in the absence of CIZ1 was retained, with over 60% of the CIZ1-dependent genes falling into this category. This further emphasises inappropriate upregulation as a key consequence of CIZ1 loss.

Gene set enrichment analysis on the significantly ($q < 0.05$) changed genes revealed highly significant overlaps with many of the same sets identified in lymphocytes (**Figure 6.11B**). Over 25% of the significantly upregulated genes in CIZ1 null spleens are targets of the DREAM complex, and over 20% are associated with BRCA1. As these are pro-cycling genes, they would be expected to be upregulated in cancerous tissue compared to normal. RB1 has been extensively profiled as a tumour suppressor, meaning loss of function could also be expected (Sellers and Kaelin, 1997). Additionally, these sets are highly related, with many of the same genes featuring in the three sets (**Figure 6.11C**).



Significance	Total	Up	Down
$p < 0.05$	3586	2091	1495
$p < 0.005$	1769	1084	685
$q < 0.05$	1714	1047	667

911 annotated genes 551 annotated genes

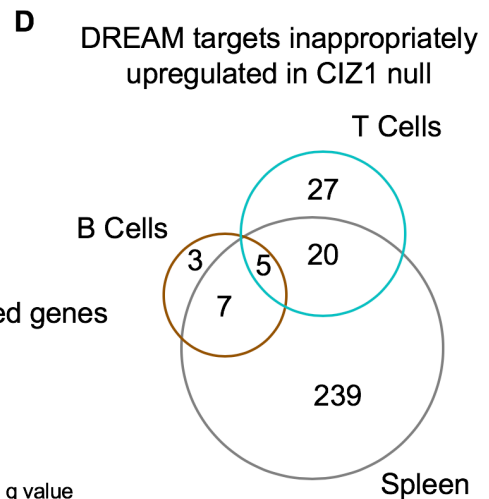
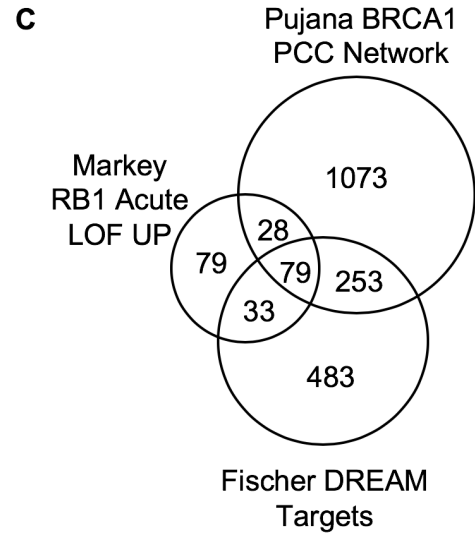
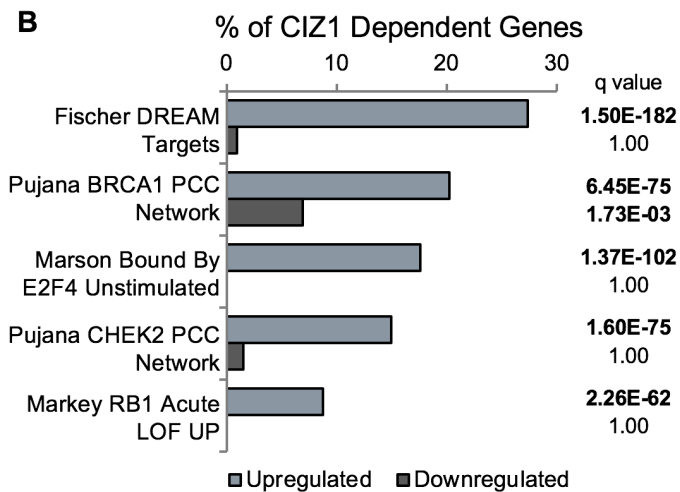


Figure 6.11 CIZ1 dependent deregulation of DREAM targets

(A) Volcano plot comparing genome-wide gene expression in spleens from 3 independent CIZ1 null mice compared to spleens from 3 independent WT mice. Genes changed to certain significance cut offs are shown beneath each comparison. CIZ1 expression is shown in green. (B) Of the 1714 genes significantly affected by loss of CIZ1, Gene Set Enrichment Analysis was performed on the 911 annotated upregulated and 551 annotated downregulated genes using the MSig Curated Gene Set database. Overlaps with sets previously identified are shown, alongside the significance (q-value) of these overlaps. (C) Venn diagram showing overlap between the genes contained within sets related to cell cycle progression and DNA repair. (D) Venn diagram showing overlap between the genes inappropriately upregulated ($q < 0.05$) upon activation of CIZ1 null lymphocytes and genes which are significantly upregulated in CIZ1 null spleens compared to WT which overlap with Fischer DREAM Targets.

To isolate CIZ1-dependent changes from possible non-specific cancer changes, the 271 Fischer DREAM Target set genes, which were significantly upregulated in CIZ1 null spleens, were compared to the inappropriately upregulated-upon-lymphocyte-activation Fischer DREAM Target set genes (**Figure 6.11D**). This identified 5 genes which are CIZ1-dependent in all three analyses – **ASPM** (Abnormal Spindle Microtubule Assembly), **BUB1B** (BUB1 Mitotic Checkpoint Serine/Threonine Kinase B), **KIF22** (Kinesin Family Member 22), **NCAPG** (Non-SMC Condensin I Complex Subunit G) and **ZWILCH** (ZWILCH Kinetochores Protein).

Despite DREAM targets spanning a large range of pathways (**Figure 6.12A**), all five of these genes are contained within the cell cycle pathway (**Figure 6.12B**). In fact these genes are all involved in mitotic checkpoint activation, with roles in chromatin condensation and regulating kinetochores function (Kouprina et al., 2005, Guo et al., 2012, Yu et al., 2014, Williams et al., 2003, Kimura et al., 2001). This is particularly interesting as the publication from which the Fischer DREAM Targets set originates found that p21 (a known

binding partner of CIZ1) is required for downregulation of TP53 targets, and that genes downregulated by TP53 are cell cycle genes bound by the DREAM complex (Fischer et al., 2016). As these are the subset of genes identified as CIZ1 dependent in this study, could interaction between p21 and CIZ1 be behind the fine tuning of target selection of these complexes? Additionally BRCA1 and CHEK2 are both established DREAM target genes (Engeland, 2018), also shown by strong overlap between DREAM targets and DNA repair pathways (**Figure 6.11C**, **Figure 6.12A**), so could deregulation of these networks actually be a downstream effect of deregulation of the DREAM pathway?

Furthermore, chronic checkpoint activation is known to be a driver of tumourigenesis, as it is only the cells which overcome the checkpoint barrier that will then be able to proliferate (Dalton and Yang, 2009). As these cells have non-functioning cell cycle checkpoints they may be particularly vulnerable to transformation.

Therefore, overall, targets of the DREAM complex, in particular those involved in mitotic checkpoint and chromatin condensation are coherently affected and of particular interest for further focussed investigation.

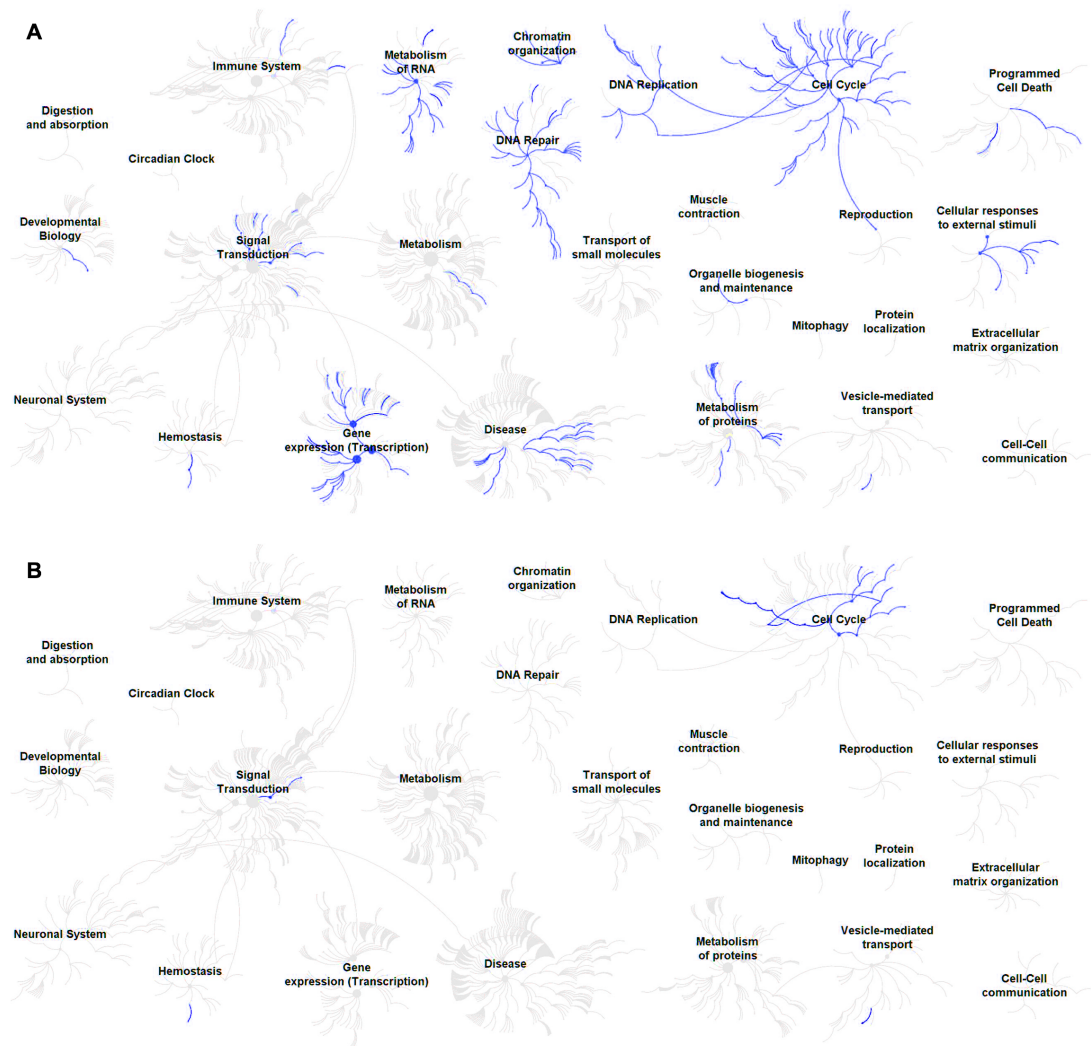


Figure 6.12 CIZ1 dependent deregulation of DREAM targets

(A) All of the genes contained within the Fischer DREAM Targets set which are expressed in WT naïve lymphocytes were analysed by the Reactome Pathway Database (Fabregat et al., 2018). Significant pathway overlap is indicated by blue highlighting. (B) The Reactome Pathway Database was instead probed with the five CIZ1 dependent genes significantly inappropriately upregulated upon activation of B and T CIZ1 null lymphocytes and in CIZ1 null spleens compared to WT which are within the Fischer DREAM Target set. Significant pathway overlap is indicated by blue highlighting.

To further understand this, the 50 genes significantly changed on the X chromosome (**Figure 6.4**) were analysed by probing GSEA and the top 15 hits investigated (**Figure 6.13**). Strikingly, the BRCA1 and CHEK2 network sets are high on this list, alongside multiple cancer related sets. Reactome Mitotic Prometaphase also significantly overlaps, which is the same set to which highly CIZ1 dependent DREAM targets *BUB1B*, *NCAPG* and *ZWILCH* also belong (Fabregat et al., 2017).

Overall this demonstrates the small number of gene expression changes seen on the X chromosome represent the same sets returned by genome-wide analysis.

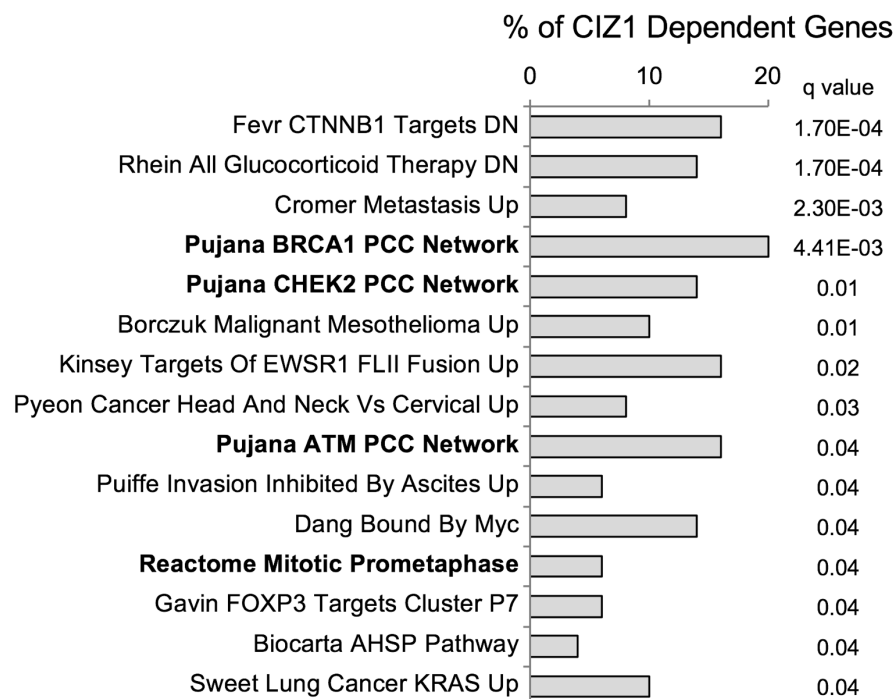


Figure 6.13 Gene set enrichment analysis of X-linked CIZ1-dependent genes in diseased spleens

Gene Set Enrichment Analysis was performed on the 50 genes significantly changed upon CIZ1 loss on the X chromosome in spleens using the MSig Curated Gene Set database. The top 15 sets are shown, alongside the significance (q-value) of these overlaps.

6.4.7 Additional avenues investigated

6.4.7.1 TCF21

GSEA curated gene set analysis of the genes which fail to be downregulated upon activation of CIZ1 null lymphocytes showed strong overlap with targets of the tumour suppressor Transcription Factor 21 (TCF21, also known as Pod-1, capsulin, or epicardin). This leads to the question of whether oncogenes normally suppressed by this transcription factor are overexpressed in CIZ1 null lymphocytes upon activation (**Appendix H.7**). TCF21 is of particular interest as it is known to play a role in spleen organogenesis (Lu et al., 2000).

However, both failure-to-downregulate and inappropriate-downregulation sets overlap significantly with genes that are downregulated by TCF21. Additionally, there is significant overlap with genes upregulated by TCF21, with both those that are inappropriately upregulated and those that fail to be upregulated upon lymphocyte activation (**Figure 6.14A**). Further analysis of revealed a similar pattern to the cell cycle and DNA damage response sets, whereby loss of CIZ1 results in altered gene expression in the same direction as what is happening to other target genes in that set (**Figure 6.14B**). **Therefore this analysis aligns well with the hypothesis that CIZ1 plays a role in target selection during biological transitions.**

6.4.7.2 Apoptosis

Another pathway of interest was apoptosis, as genes upregulated following induction of apoptosis (pro-apoptotic genes) overlap significantly with genes that are inappropriately downregulated upon activation of CIZ1 null lymphocytes (**Appendix H**). Further analysis, however, did not reveal a consistent trend (**Figure 6.14C**). In general there was a much greater overlap with apoptosis related genes when apoptosis was induced by the chemotherapy drug doxorubicin compared to more natural methods such as deprivation of serum. Furthermore, in T cells there was no preferential directional overlap, as both downregulated and upregulated upon apoptosis genes overlapped similarly with all four activation categories. Overall there is no clear avenue of further investigation for this observation.

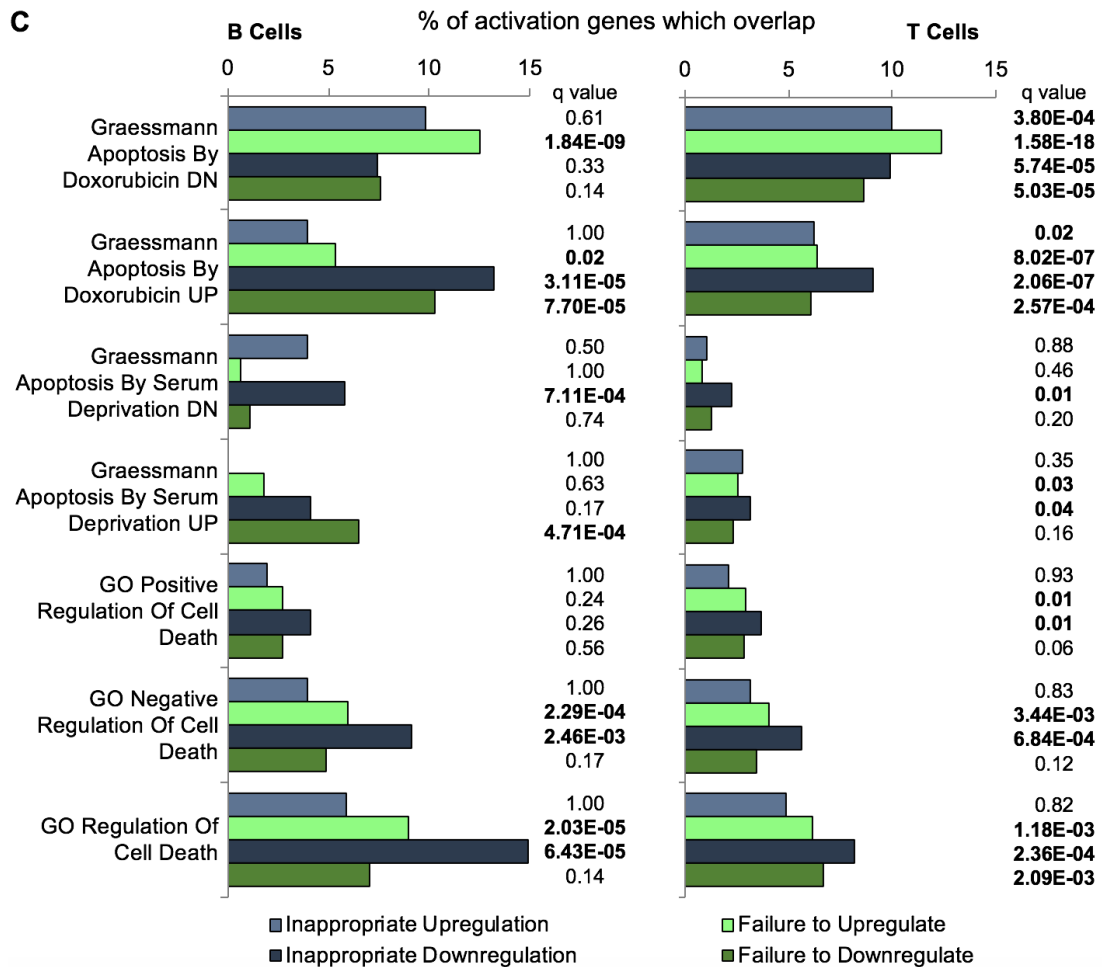
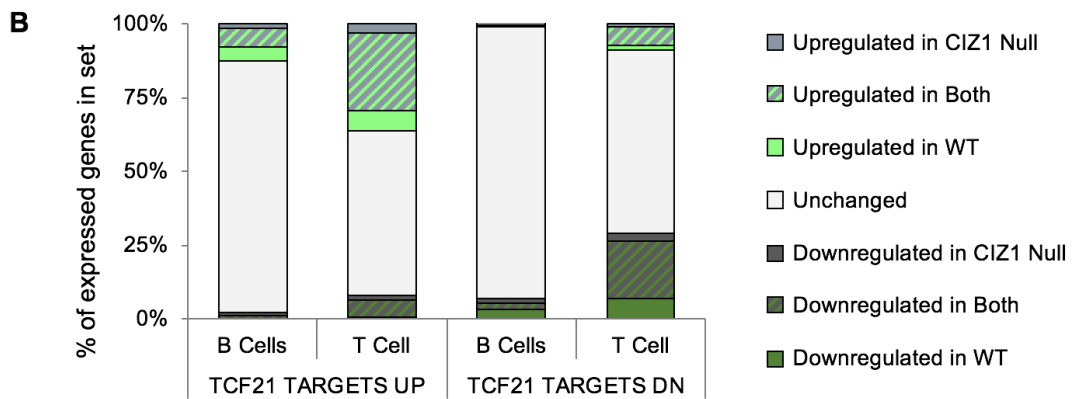
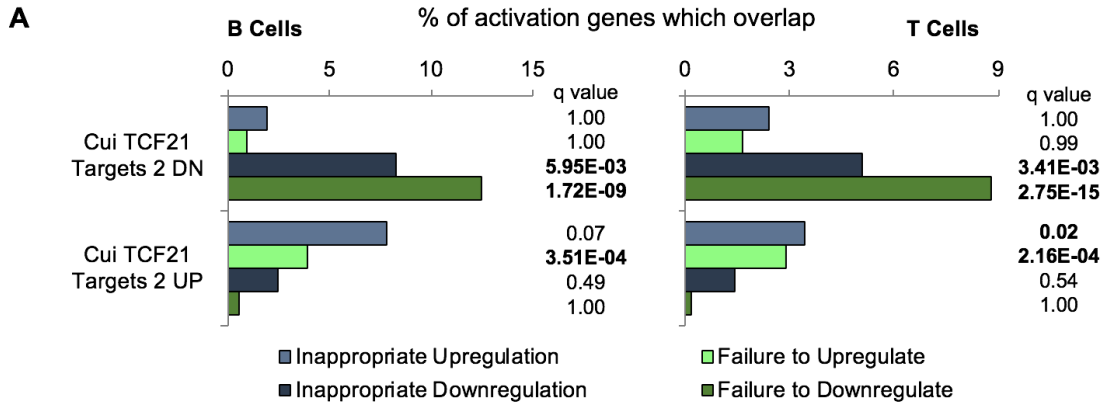


Figure 6.14 Targets of TCF21 and genes induced during apoptosis

(A) The percentage of genes in each of the four activation categories which overlap with Curated Gene Sets in the MSig GSEA database relating to TCF21 for B cell (left) and T cell (right) activation. The significance of these overlaps are shown to the right of each graph, significant q values are shown in bold. (B) Detailed analysis of the effect of lymphocyte activation on all genes within the TCF21 related Curated Gene Sets seen in (A) which are expressed in WT naïve lymphocytes. (C) Same analysis as (A), but for Curated and Gene Ontology gene sets relating to apoptosis.

6.4.7.3 FOXP3

Also highly associated with genes in the failure to downregulate category are genes that are bound by the master transcriptional regulator Forkhead box P3 (FOXP3), which is critical in the development of regulatory T cells that generally suppress the immune response (Hori et al., 2003). Intriguingly FOXP3 defects are known to cause X-linked syndrome (IPEX) (Bennett et al., 2001), also known as X-linked autoimmunity-immunodeficiency syndrome, and is therefore of particular interest here due to the female specific nature of the CIZ1 null phenotype. Further investigation revealed that although there is strong overlap with the failure to downregulate category, there is also significant overlap with the other three categories in some of the FOXP3 sets (**Figure 6.15**).

Strikingly, although there is significant overlap with genes bound by FOXP3, there is little to no overlap with targets of FOXP3, identified in the same study (**Figure 6.15**). Yet the 'target' sets are much smaller than the 'bound by' sets, for example 1023 genes were found to be bound by FOXP3 in the Marson stimulated set, but only 30 genes were found to be upregulated targets and 12 genes downregulated targets. This suggests FOXP3 binding is only one part of the mechanism regulating target genes.

It should be noted that these targets of FOXP3 were identified in a hybridoma cell line, which is very different from the primary lymphocyte cultures used in

my analysis. Similarly, the Zheng study was conducted on specific subsets of regulatory T cells, compared to the mixed population in my study. Thus, any lineage specific changes in targets in my study could be undetectable against the background of the unchanged expression in the other cell lineages. Therefore, the CIZ1-dependent genes in this study which overlap with genes 'bound by FOXP3' could still be FOXP3 targets. Alternatively FOXP3 could have another role that CIZ1 could be affecting.

Despite the complexity of these observations, genes bound by FOXP3 appear to be deregulated in lymphocytes in some manner. This could imply a lack of or inappropriate occupancy of FOXP3 on these genes in the genome. Overall, this analysis appears to provide further evidence in support of the idea that CIZ1 is involved in targeting factors to their appropriate sites of action.

To test this hypothesis ChIP for FOXP3 could be performed in WT and CIZ1 null lymphocytes of a more uniform background (for example separately into B and T cells by FACS). This could help determine if FOXP3 occupancy is affected by loss of CIZ1, in addition to the downstream effects on gene expression that have been quantified here.

6.4.7.4 Deregulation of TCF21 targets and genes bound by FOXP3 in CIZ1 null spleens compared to WT spleens

Gene set enrichment analysis on the CIZ1 dependent genes ($q < 0.05$) identified in diseased CIZ1 null spleens compared to WT (**Figure 6.16A**) also revealed significant overlap with 'bound by FOXP3' and TCF21 target sets (**Figure 6.16B**). However this overlap was much smaller than for the cell cycle related sets (**Figure 6.11B**). Additionally, although genes bound by FOXP3 remained a significantly changed set, the trend was not consistent with that previously identified.

Zheng 'bound by FOXP3' had significant overlap with the 'failure-to-downregulate' activation set (**Figure 6.15**), so these genes would be expected to be upregulated in CIZ1 null spleens in comparison to WT spleens if status was maintained from young lymphocyte to adult spleen. However, Zheng 'bound by FOXP3' was more strongly associated with the downregulated genes in spleens (**Figure 6.16B**), despite the overall bias towards CIZ1-dependent upregulation generally. Furthermore, detailed analysis revealed greater overlap between the genes which failed to be downregulated upon activation of CIZ1 null lymphocytes and genes downregulated in the spleens, with no overlap with upregulated CIZ1 dependent genes in spleen (**Figure 6.16C**). Therefore FOXP3 related sets are less coherently CIZ1-dependent, and further analysis was not performed.

To isolate the most CIZ1-dependent changes within the TCF21 target sets the 67 'Cui TCF21 Targets 2 Up' genes, which were significantly upregulated in CIZ1 null spleens, were compared to the genes within the same set which were inappropriately upregulated-upon-lymphocyte-activation (**Figure 6.16D**). There was strong overlap between these sets, with 21 strongly CIZ1-dependent genes identified (**Table 6.2**). However, further analysis revealed that 19/21 of these genes are also found in the 'Fischer DREAM Targets' set, including previously identified *ASPM* and *KIF22* in addition to another non-SMC condensin *NCAPH*, and 17/21 are within the term 'GO Mitotic Cell Cycle'. Therefore this further correlates with the findings of the earlier analysis

showing deregulation of genes involved in mitosis and mitotic checkpoint activation but does not provide any new avenues of investigation.

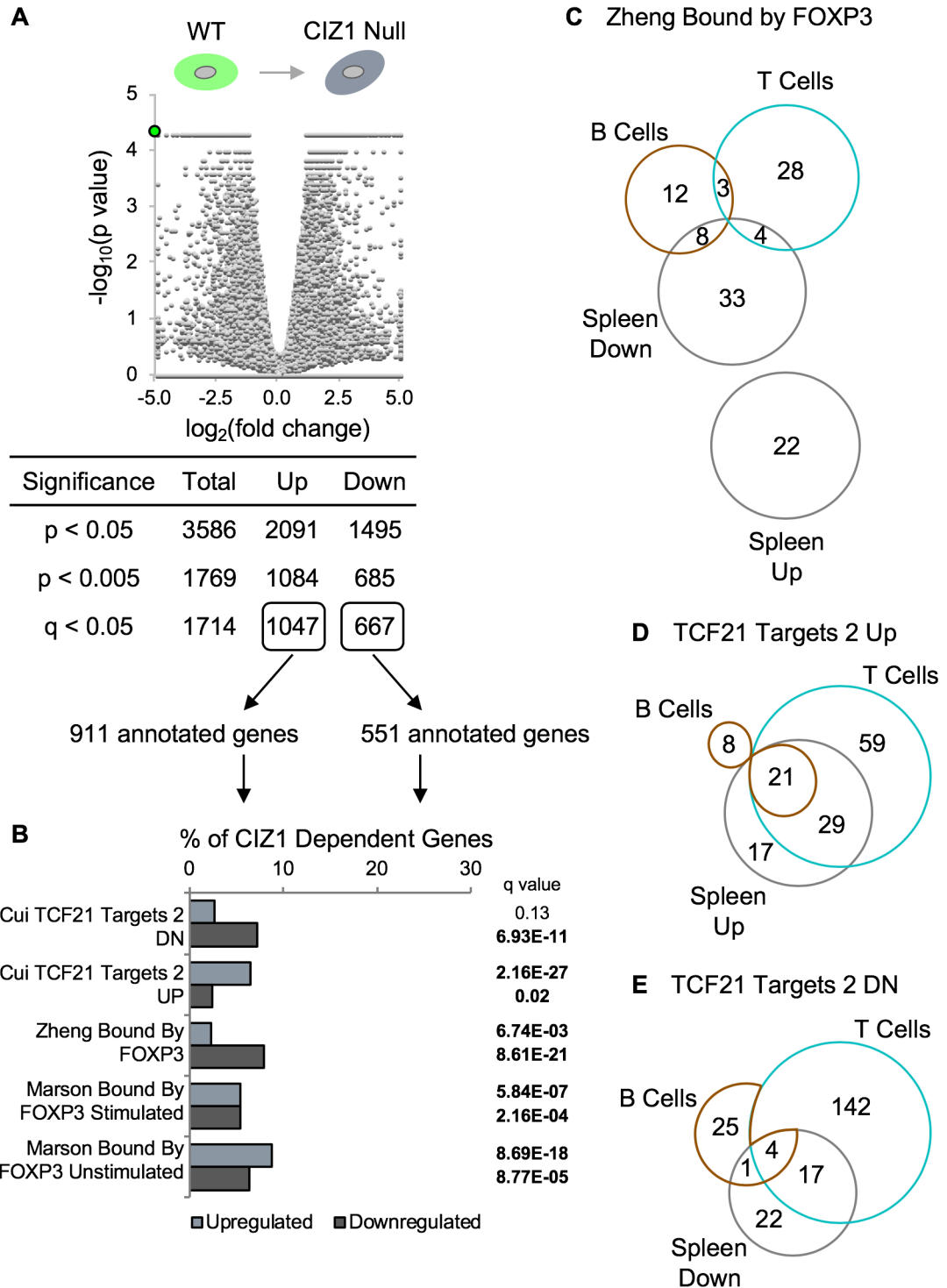


Figure 6.16 CIZ1 dependent deregulation TCF21 targets and genes bound by FOXP3

(A) Volcano plot comparing genome-wide gene expression in spleens from 3 independent CIZ1 null mice compared to spleens from 3 independent WT mice. Genes changed to the indicated significance cut offs are shown beneath each comparison. CIZ1 expression is shown in green. (B) Of the 1714 genes significantly affected by loss of CIZ1, Gene Set Enrichment Analysis was performed on the 911 annotated upregulated and 551 annotated downregulated genes using the MSig Curated Gene Set database. Overlaps with sets previously identified are shown, alongside the significance (q-value) of these overlaps. (C) Venn diagram showing overlap between the genes which fail to be downregulated ($q < 0.05$) upon activation of CIZ1 null lymphocytes and genes which are significantly changed in CIZ1 null spleens compared to WT which overlap with Zheng Bound by FOXP3. (D) Venn diagram showing overlap between the genes which are inappropriately upregulated ($q < 0.05$) upon activation of CIZ1 null lymphocytes and genes which are significantly upregulated in CIZ1 null spleens compared to WT which overlap with Cui TCF21 Targets 2 Up. (E) Venn diagram showing overlap between the genes which are inappropriately downregulated ($q < 0.05$) upon activation of CIZ1 null lymphocytes and genes which are significantly downregulated in CIZ1 null spleens compared to WT which overlap with Cui TCF21 Targets 2 Down.

The same analysis on the 44 'Cui TCF21 Targets 2 Down' genes which were significantly downregulated regulated in CIZ1 null spleens revealed 4 strongly CIZ1-dependent genes (**Figure 6.16E**) – **ARHGEF18** (*Rho/Rac Guanine Exchange Factor 18*), **CD55** (*Complement*), **MYADM/PRKCG** (*Myeloid Associated Differentiation Marker/Protein Kinase C Gamma*) and **NOSTRIN** (*Nitrogen Oxide Synthase Trafficking*). However these genes do not have an obvious connection to each other, so no further analysis was conducted.

Table 6.2 The 21 CIZ1-dependent genes identified in (Figure 6.16D).

Gene	Aliases	Role of Protein Product
<i>ASPM</i>	Abnormal Spindle Microtubule Assembly	Normal mitotic spindle function (Kouprina et al., 2005)
<i>CCNA2</i>	Cyclin A2	Promote transition through cell cycle (Copeland et al., 2010)
<i>CDCA3</i>	Cell Division Cycle Associated 3, TOME-1	Trigger for mitotic entry (Ayad et al., 2003)
<i>CDCA7</i>	Cell Division Cycle Associated 7, JPO1	Myc-responsive gene (Prescott et al., 2001)
<i>CDCA8</i>	Cell Division Cycle Associated 8, Borealin	Chromosomal passenger complex (CPC) component, involved in control of attachment-error-correction and chromosomal alignment (Gassmann et al., 2004)
<i>CHAF1B</i>	Chromatin Assembly Factor 1 Subunit B, M-Phase Phosphoprotein 7	Assembly of histones onto newly replicated DNA, possible role in DNA repair (Nabatiyan and Krude, 2004)
<i>ESPL1</i>	Separase, Separin, Extra Spindle Poles-Like 1 Protein	Final separation of sister chromatids prior to anaphase (Sun et al., 2009)
<i>GALK1</i>	Galactokinase 1	Galactose metabolism (Park et al., 2009)
<i>STMN1</i>	Stathmin1, Oncoprotein 18, Leukemia-Associated Phosphoprotein P18	Destabilises microtubules (Marklund et al., 1996)
<i>HMMR</i>	Hyaluronan Mediated Mobility Receptor, RHAMM	Forms a complex with BRCA1 and BRCA2, involved in cell motility (Maxwell et al., 2011)
<i>INCENP</i>	Inner Centromere Protein	Chromosomal passenger complex (CPC) component, key regulator of mitosis (Adams et al., 2001)
<i>IQGAP3</i>	IQ Motif Containing GTPase Activating Protein 3	Regulation of cellular proliferation through the Ras/ERK signalling cascade (Nojima et al., 2008)
<i>KIF22</i>	Kinesin Family Member 22, KID	Move chromosomes during cell division, metaphase chromosome alignment and maintenance (Yu et al., 2014)
<i>LMNB1</i>	Lamin B1	Component of the nuclear lamina (Luderus et al., 1992)
<i>MCM5</i>	Minichromosome Maintenance Complex Component 5, CDC46	Component of MCM2-7 complex involved in initiation of DNA replication (Chen et al., 1992)
<i>NCAPH</i>	Non-SMC Condensin I Complex Subunit H	Chromosome condensation (Kimura et al., 2001)
<i>NUF2</i>	NDC80 Kinetochores Complex Component NUF2, Cell Division Cycle-Associated Protein 1	Component of NDC80 complex, required for kinetochores integrity and organising stable microtubule binding sites (Nabetani et al., 2001)
<i>PRIM1</i>	DNA Primase Subunit 1	Synthesises small RNA primers for Okazaki fragments (Shiratori et al., 1995)
<i>SPC24</i>	SPC24 NDC80 Kinetochores Complex Component	Component of NDC80 complex, required for kinetochores integrity and organising stable microtubule binding sites (Bharadwaj et al., 2004)
<i>TPX2</i>	Targeting protein for Xklp2	Required for chromatin and/or kinetochores dependent microtubule nucleation (Wittmann et al., 2000)
<i>UBE2C</i>	Ubiquitin Conjugating Enzyme E2 C, Mitotic-Specific Ubiquitin-Conjugating Enzyme	Ubiquitin conjugating enzyme (E2) involved in promotion of mitotic exit (Townesley et al., 1997)

6.4.8 Summary of avenues for investigation

6.4.8.1 Unannotated transcripts

The transcription of unannotated transcripts in CIZ1 null B cells upon activation and deregulation of mitosis and mitotic checkpoints are most clearly worthy of further investigation.

These unannotated transcripts are typically between 400bp and 1kb in length, meeting the 200bp requirement to be classified as a lncRNA (Hung and Chang, 2010), and generally located in between genes. The mechanism of how these transcripts are being produced could be equally, if not more relevant than what the transcripts themselves do. For example are they repressive transcripts being expressed in an attempted compensatory mechanism to shut down targets in the absence of CIZ1? Are the promoters of these transcripts usually bound by a regulator in the presence of CIZ1, normally preventing their transcription?

6.4.8.2 Mitosis

Genes encoding proteins involved in mitosis, particularly prometaphase and mitotic checkpoint activation were also found to be coherently significantly upregulated in the absence of CIZ1. Whether this translates to change in protein level could be confirmed, and the effect this has on CIZ1 null cells investigated.

6.4.8.3 Limitations

The biggest limitation of this analysis was the small number of replicates for the lymphocyte samples and the mixed cell populations. By sorting lymphocytes into B and T cells using FACS, analysis could be performed on data produced from a more uniform cell background, allowing easier identification of lineage specific changes. Additional replicates would allow more confidence in the determination of CIZ1-dependent genes.

Further direction could be found from additional analysis of the adult female mouse pathology. If a more definitive diagnosis of the type of lympho-

proliferative disorder could be determined, this could indicate individual lineages which would be of particular interest in further studies.

6.5 Conclusions

Overall, the effect of CIZ1 loss on young adult female splenic lymphocytes was investigated, yielding promising avenues for further research that might determine the cause of the lympho-proliferative disorder phenotype. However, none of these analyses provided an explanation for the female specificity of the phenotype. Additional analysis on individual lineages with more replicates, paired with information on disease pathology could give further insight.

7.0 Discussion and Future Directions

7.1 How is the nuclear matrix organized at the Xi?

Nuclear matrix extraction followed by quantification of immunofluorescence intensity showed that many epitopes at the Xi are not accessible to antibodies prior to removal of chromatin (**Chapter 3**). This is in concordance with previous literature on SAF-A (hnRNPU) (Nakagawa and Prasanth, 2011). Interestingly, the buried epitopes investigated in this study are sensitive to extraction with RNase. This suggests that the compartment in which these proteins are located is dependent on interaction with RNA, most likely the lncRNA *Xist* to which many of these proteins have been shown to bind (Helbig and Fackelmayer, 2003, Hasegawa et al., 2010, Jeon and Lee, 2011, Chu et al., 2015, Ridings-Figueroa et al., 2017, Sunwoo et al., 2017).

A recent publication has hypothesised liquid-liquid phase separation, the process of forming membrane-less organelles, which are involved in various biological functions (Fay and Anderson, 2018), as a driver of X chromosome inactivation. They hypothesise that *Xist* seeds this phase-separated compartment and progressively sequesters its binding partners, including CIZ1, creating a high-density compartment (Cerese et al., 2019). The data presented in this thesis correlate well with this hypothesis, with the high-density of the compartment explaining why antibody inaccessibility is so prevalent in Xi territory and *Xist* dependency explaining RNase sensitivity. However, my nucleus-wide quantitative data on CIZ1 (**Figure 3.4**) suggests that these RNase-sensitive buried epitopes are not exclusive to the Xi territory. Though it is possible that liquid-liquid phase separation is a common feature of the nuclear matrix, there is no data presented here to support this.

7.2 CIZ1-dependent chromosomal relocation

CIZ1 is required for chromosomal relocation during replication, moving the inactive X chromosome from its anchorage site at the nuclear periphery (Chen et al., 2016) to the nucleolus and back (**Chapter 5**) (Stewart et al., 2019). The catalytic subunit of PRC2, EZH2, is anchored to the nuclear matrix and is enriched at this nucleolar location. This is consistent with literature suggesting that relocation is required for maintenance of H3K27me3 (Zhang et al., 2007, Yang et al., 2015) and suggests that CIZ1 is required for accurate engagement of enzyme and template, specifically the template to be remodelled. However, the active motor behind this relocation remains to be identified.

Previous literature (Kulashreshtha et al., 2016, Chuang et al., 2006, Mehta et al., 2010) and analysis of RNA-sequencing data (**Chapter 5**) has already proposed Nuclear Myosin 1 (NM1) and actins as possible candidates. This could be further investigated by studying the effects of inhibitors or mutants of actins and NM1 on chromosomal relocation (Chuang et al., 2006) and identifying whether NM1 and actins are bound to, or in close proximity to, components of the movement machinery such as CIZ1, using techniques such as proximity ligation assay.

Additional work that could identify possible components of the movement machinery include investigation of CIZ1 binding partners, both protein and RNA. Although CIZ1 is anchored to the nuclear matrix at the Xi, relocation of Xi occurs in only a small proportion of cells in an asynchronous population (Stewart et al., 2019). Thus, if crosslinking methods are used (Chu et al., 2015), enrichment for cells undergoing chromosomal relocation prior to crosslinking would be beneficial. Possible strategies for enrichment of this population include cell synchronisation with thymidine arrest and release (Chen and Deng, 2018), and nuclear matrix extraction (Stewart and Coverley, 2018). To negate the need for crosslinking, alternative approaches such as Bio-ID (Roux et al., 2013) for identification of protein binding partners is also worthy of consideration.

Further work using CIZ1 mutants could also help identify which regions of CIZ1 are required for chromosomal relocation, and begin to unravel the mechanism by which it functions. For example, CIZ1 can be separated into its two functional domains of DNA replication and nuclear matrix anchorage (Ainscough et al., 2007, Coverley et al., 2005). Can cells expressing only one of these functional domains maintain chromosomal relocation or does this mechanism require both functions? This could be followed by more directed mutagenesis of cyclin binding motifs, phosphorylation sites or individual zinc-fingers for further mechanistic insight (Ainscough et al., 2007, Copeland et al., 2010, Copeland et al., 2015).

7.3 Deregulation of chromatin remodelling complex targets

I also performed bioinformatic analysis on RNA-sequencing data obtained from WT and CIZ1 null primary embryonic fibroblasts and young adult splenocytes. The aim was to correlate the loss of *Xist* localisation and epigenetic mark enrichment at the Xi observed in CIZ1 null primary embryonic fibroblasts with changes in gene expression, and identify pathways affected by loss of CIZ1 in lymphocytes which could underpin the lymphoproliferative disorder phenotype.

The investigations presented in this thesis found little overlap between CIZ1-dependent genes in primary embryonic fibroblasts (PEFs, **Chapter 4**) and lymphocyte activation (**Chapter 6**), and GSEA analysis returned few of the same pathways. Though this might be expected as these cells are derived from very different lineages, if CIZ1 was a key component of a specific signaling cascade, that pathway would likely be deregulated in both contexts. Therefore, CIZ1 appears to have a role independent of a specific pathway.

However, loss of CIZ1 did result in deregulation of targets of complexes in both cases. PEF analysis showed **down-regulation** of targets of the polycomb repressive complexes (**Chapter 4**) while lymphocytes showed **up-regulation** of DREAM complex targets following lymphocyte activation (**Chapter 6**). Crucially, the prevailing directions of change in target gene expression were

opposite to what might be expected. For example, in the absence of repressive enrichment of epigenetic marks at the Xi in PEFs (Ridings-Figueroa et al., 2017), it would be expected that target genes would no longer be repressed, resulting in upregulation.

Primary embryonic fibroblasts from the embryo have undergone differentiation, during which polycomb repressive complexes aid repression of one of the X chromosomes in female cells (Pintacuda et al., 2017, da Rocha et al., 2014). During lymphocyte activation, naïve lymphocytes re-enter cell cycle from quiescence, activating DREAM target genes (Sadasivam and DeCaprio, 2013). Therefore, even though the targets of different complexes were affected in the different cell types, the complexes involved and direction of change in both cases is consistent with what is already going on in that cell type. Both of these processes involve extensive chromatin remodelling (Chen et al., 2016, Mehta et al., 2010).

7.4 Expression of unannotated transcripts

Another observation made in lymphocytes was that, following activation of CIZ1 null B cells with LPS, there was a significant upregulation in expression of unannotated transcripts (**Chapter 6**). These RNAs do not contain an obvious open reading frame to suggest they may be unannotated protein-coding genes and, at between 400bp and 1kb length, meet the requirements to be classified as lncRNAs (Hung and Chang, 2010). What these transcripts do and the mechanism behind how they are generated remains unclear. Located between genes, it is possible that they are repressive transcripts produced as part of a compensatory mechanism to shut down nearby targets in the absence of CIZ1 (Garitano-Trojaola et al., 2013, Kopp and Mendell, 2018). It is also possible that the transcripts have no function, and in fact their presence is a consequence of altered genome architecture in the absence of CIZ1, leading to the promoters of these transcripts being aberrantly accessible to RNA polymerases. This could be investigated using chromatin conformation capture techniques on WT and CIZ1 null cells prior to and following activation to identify changes in genome architecture (Mumbach et al., 2019).

7.5 A role for CIZ1 in chromatin remodelling?

The data discussed here are consistent with the hypothesis that CIZ1 plays a role in chromatin remodelling during biologically relevant transitions, for example differentiation (**Chapter 4, 5**) and entry into or exit from quiescence (**Chapter 6**). CIZ1 is required to accurately select targets to be remodelled. In the absence of CIZ1, accurate target selection is no longer maintained as CIZ1-directed convergence of enzyme and template cannot occur. Support for this hypothesis includes:

- 1) Compensation for loss of enzyme-template convergence by releasing spatial constraint on the enzyme, resulting in unbiased target selection. This leads to changes in gene expression of a subset of targets that should escape regulation of the prevailing pathway, and a subset of genes which should be targeted are potentially missed. Overall, this leads to **deregulation of targets of complexes active in the cell**.
- 2) Absence of enzyme-template convergence or possible error-prone compensation mechanisms leading to inaccurate remodelling events could also lead to previously inaccessible promoters becoming accessible to RNA polymerase, leading to random transcription events which could include **unannotated transcripts**.
- 3) Incorrect chromatin remodelling in mitosis would result in mitotic checkpoint activation (Yao and Dai, 2012). **The DREAM complex targets, which are coherently upregulated in CIZ1 null activated splenocytes and spleen tissue, are mitotic checkpoint proteins**.

7.6 Limitations of using cell lines

7.6.1 Differences between primary cells and ‘normal’ cell lines

Normal differentiated primary cells have limited replicative capacity, governed by the shortening of telomeres (Bodnar et al., 1998). Once cells reach the end of this capacity, known as the Hayflick limit, cells exit cell cycle and enter replicative senescence or crisis (Hayflick and Moorhead, 1961). In order to generate culture adapted cell lines with limitless replicative potential, cells must be cultured through this ‘crisis’ and re-enter cell cycle, usually due to acquiring telomerase activity via mutation or transfection (Lee et al., 2004).

Therefore, though cell lines generated from ‘normal’ cells are still generally considered to be ‘normal’, they display some hallmarks of cancer (Hanahan and Weinberg, 2011). For example, all culture adapted cell lines must have ‘enabled replicative immortality’ and the cell lines used in these analyses have ‘genome instability and mutation’, as demonstrated by the presence of multiple inactive X chromosomes in subpopulations of culture adapted cells seen throughout (**Chapter 3**) and (**Chapter 5**). Abnormal karyotypes are a common feature in cell lines, with cells of the ‘normal’ human mammary epithelial lines MCF-12A and MCF-12F cell line having between 65 and 71 chromosomes instead of 46 (Pauley et al., 1993).

7.6.2 Suitability of cell lines when investigating the role of CIZ1

In this analysis, loss of CIZ1 and adaption to long term culture resulted in many of the same phenotypes; including loss of replicating Xi relocation, change of EZH2 isoform, loss of EZH2 anchoring and deregulation of the same pathways (Stewart et al., 2019). However, RNA-independent nuclear matrix anchorage of CIZ1 during Xi replication occurred in both primary and culture adapted cells.

Previous literature comparing nuclear matrix composition (the structure to which CIZ1 is anchored) between primary breast tissue, ‘normal’ breast cell line MCF-10 and primary cancer breast tumour tissue demonstrated that the MCF-10 cell line had a phenotype intermediate between the normal and

cancer tissue (Khanuja et al., 1993). As CIZ1 has been linked to several cancers including lymphoma and cancers of the colon, gallbladder, liver, lung and prostate (Higgins et al., 2012, Yin et al., 2013, Wang et al., 2014, Zhang et al., 2015, Liu et al., 2015, Lei et al., 2016, Wu et al., 2016, Ridings-Figueroa et al., 2017), it is not unexpected that mechanisms involving CIZ1 will be compromised in the culture adaption process.

The data and literature imply that both the processes in which CIZ1 is involved, and the structure to which it is anchored, are corrupted in even 'normal' cell lines. Crucially, these data reinforces that 'normal' cell lines do not fully represent the functional mechanisms that occur in primary cells and, by implication, *in vivo*.

7.6.3 Implications

Analysis of EZH2 in primary cells and cell lines revealed EZH2 α (commonly reported isoform) is upregulated upon culture adaption and is more soluble than EZH2 β (primary cell specific isoform). A partial shift from β - α is already evident in primary CIZ1 null PEFs. One hypothesis behind this release is that the lower MW EZH2 α isoform is missing the additional sequence or modification that mediates nuclear matrix binding. However, as the nuclear matrix itself is thought to be disrupted in cell lines, it is possible that the change in EZH2 isoform may not be the direct cause of solubility. EZH2 α may be capable of anchorage to the nuclear matrix, however destabilisation of the nuclear matrix structure may result in a lack of binding sites. I favour the former possibility since other nuclear matrix proteins retain ability to bind to a nuclear matrix structure in the same culture adapted cells. For example, transient RNA-independent anchorage of CIZ1 still occurs during Xi replication (**Chapter 5**) and buried RNA-protein nuclear matrix bound epitopes of SAF-A and YY1 are detected (**Chapter 3**).

There are a large number of publications that investigated EZH2 in cell lines. These include studies of interaction partners (reviewed in (Yamaguchi and Hung, 2014)), enzyme kinetic studies (Diaz et al., 2012) and even targeting EZH2 as a possible cancer therapeutic (Yamagishi and Uchimaru, 2017, Han

Li and Chen, 2015). Based on the data in this thesis, the isoform of EZH2 studied in these publications may not be the isoform expressed *in vivo*. As the differences between these isoforms are not fully delineated, with the exception of nuclear matrix binding capability, the data from some of these studies may not be applicable to the normal state in the body.

These findings also have implications for further study. As mentioned previously, gene editing tools such as CRISPR/Cas9 (Zhang et al., 2014) could be used to ask questions such as whether chromosomal relocation can still occur if key functional motifs in CIZ1 are removed. The selection processes required for these techniques means they can only be done in cell lines *in vitro*. As compromised chromosomal relocation necessitates the use of primary cells, these manipulations would instead need to be carried out *in vivo*, which is much more expensive and therefore unsustainable to use for screening multiple small manipulations. To do these kinds of analyses a better model would be required.

7.6.4 Developing a better model for mechanism of movement

To avoid studying the role of CIZ1 in a system that is already compromised, without extensive animal use, a different model needs to be developed. One possibility is the use of embryonic stem (ES) cell lines. ES cells already have unlimited replicative potential and do not inactivate an X chromosome until stimulated to differentiate (Okamoto et al., 2004). Therefore, it is possible that changes seen in cell lines produced from differentiated cells will not have occurred, and inducing differentiation in ES cell lines could produce cells more similar to the primary cells isolated from animals, which would maintain the ability to efficiently relocate chromatin. If investigations proved that differentiated cells from ES cell lines exhibit chromosomal relocation, genetic manipulations could be performed in the ES cell lines prior to stimulating differentiation, bypassing the need for animal use.

7.7 Limitations of splenocyte analysis

Studying the effect of loss of CIZ1 in splenocyte activation and spleen tissue yielded interesting results including upregulation of DREAM targets, in particular those involved in mitotic checkpoint activation, and transcription of unannotated transcripts in activated B cells (**Chapter 6**).

However, analysis was performed on unsorted splenocytes. As approximately 50% of isolated splenocytes recruited *Xist* to an inactive X chromosome following B or T cell activation (**Figure 6.1**)(Ridings-Figueroa et al., 2017), it was estimated that the B and T numbers in all samples were approximately equal. Additionally, there are many cell subsets within B and T cells (LeBien and Tedder, 2008, Golubovskaya and Wu, 2016). Therefore, though only the B or T cells within the population were activated, changes in gene expression in a certain cell type might be lost within the background expression of genes in other cell types within the sample.

Approximately equal numbers of B and T cells in samples may be unexpected as activation should cause re-entry of splenocytes into cell cycle. However, cells enter an extended G1 phase following re-entry from quiescence (Brooks, 1976, Zetterberg and Larsson, 1985, Coverley et al., 2002). Previous analysis modelling cell cycle in murine proliferating lymphocytes found a mean cell cycle rate of between 12 and 18 hours, and that G1 made up only a small proportion of total division time (Dowling et al., 2014). Therefore, taking into account an extended G1 phase, it is possible that lymphocytes have not gone through a full cell cycle following 24 hours of activation.

Further analysis could be performed on lymphocytes that have been sorted, for example using flow cytometry to separate B and T cells. To determine the most relevant lineage for further studies, further investigations into the cell types involved in the lymphoproliferative disorder phenotype could be carried out.

Additionally, this thesis has implicated a possible role for CIZ1 in chromatin remodelling. If the cause of the lymphoproliferative disorder is due to aberrant

events in lymphocyte activation such as chromatin remodelling events as lymphocytes re-enter cell cycle from quiescence, kinetics could play an important role. Previous investigations using primary human fibroblasts showed that chromosomal relocation can occur within 15 minutes of receiving stimulus to enter quiescence (Mehta et al., 2010). The only time point investigated in this study was after 24 hours of activation, which may be long after the events of interest have occurred.

Therefore, further investigations could be carried out in specific lineages of splenocyte based on findings from further investigation of disease pathology, and over a larger range of times post activation.

7.8 Conclusions

In summary, I have optimised a protocol capable of profiling how proteins are anchored within the nucleus (**Chapter 3**). I have generated and combined bioinformatic and *in vitro* data to demonstrate a role for CIZ1 in maintaining the epigenetic landscape in primary embryonic fibroblasts and identified that this is also corrupted in cells that have been adapted to long term culture (**Chapters 4 and 5**). I have also generated interesting avenues for further research into the cause behind the female specific lymphoproliferative disorder phenotype displayed by CIZ1 null mice (**Chapter 6**), however these findings are yet to be validated. Based on the data obtained from the primary embryonic fibroblasts and splenocytes, I have also hypothesised a role for CIZ1 in chromatin remodelling.

The data has met the aims of this thesis and provides greater insight into mechanisms of how enzymes and template meet, maintenance of transcriptional silencing and new insight into changes which occur during cell transformation which could be relevant to development of cancer biomarkers and therapeutics.

Appendix A

PDF of Ridings-Figueroa R, **Stewart ER**, Nesterova TB, Coker H, Pintacuda G, Godwin J, et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* 2017.

The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory

Rebeca Ridings-Figueroa,^{1,7} Emma R. Stewart,¹ Tatyana B. Nesterova,² Heather Coker,² Greta Pintacuda,² Jonathan Godwin,² Rose Wilson,^{1,8} Aidan Haslam,¹ Fred Lilley,¹ Renate Ruigrok,³ Sumia A. Bageghni,³ Ghadeer Albadrani,^{3,4} William Mansfield,⁵ Jo-An Roulson,⁶ Neil Brockdorff,² Justin F.X. Ainscough,^{1,3} and Dawn Coverley¹

¹Department of Biology, University of York, York YO10 5DD, United Kingdom; ²Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; ³Leeds Institute of Cardiovascular and Metabolic Medicine (LICAMM), University of Leeds, Leeds LS2 9JT, United Kingdom; ⁴Princess Nourah Bint Abdulrahman University (PNU), Riyadh, Kingdom of Saudi Arabia; ⁵Stem Cell Institute, University of Cambridge, Cambridge CB2 1QR, United Kingdom; ⁶Leeds Institute of Molecular Medicine (LIMM), University of Leeds, Leeds LS9 7TF, United Kingdom

The nuclear matrix protein Cip1-interacting zinc finger protein 1 (CIZ1) promotes DNA replication in association with cyclins and has been linked to adult and pediatric cancers. Here we show that CIZ1 is highly enriched on the inactive X chromosome (Xi) in mouse and human female cells and is retained by interaction with the RNA-dependent nuclear matrix. CIZ1 is recruited to Xi in response to expression of X inactive-specific transcript (Xist) RNA during the earliest stages of X inactivation in embryonic stem cells and is dependent on the C-terminal nuclear matrix anchor domain of CIZ1 and the E repeats of *Xist*. CIZ1-null mice, although viable, display fully penetrant female-specific lymphoproliferative disorder. Interestingly, in mouse embryonic fibroblast cells derived from CIZ1-null embryos, Xist RNA localization is disrupted, being highly dispersed through the nucleoplasm rather than focal. Focal localization is reinstated following re-expression of CIZ1. Focal localization of Xist RNA is also disrupted in activated B and T cells isolated from CIZ1-null animals, suggesting a possible explanation for female-specific lymphoproliferative disorder. Together, these findings suggest that CIZ1 has an essential role in anchoring Xist to the nuclear matrix in specific somatic lineages.

[*Keywords:* CIZ1; *Xist*; X-chromosome inactivation; nuclear matrix; lymphoproliferative disorder]

Supplemental material is available for this article.

Received January 5, 2017; revised version accepted April 20, 2017.

In mammals, dosage compensation for X-linked transcripts is achieved by the developmentally regulated inactivation of one of the two X chromosomes in female cells. X-chromosome inactivation (XCI) is initiated by X inactive-specific transcript (*Xist*), a noncoding RNA ~17 kb in length (Brockdorff et al. 1992; Brown et al. 1992; Lee et al. 1996; Penny et al. 1996). *Xist* RNA is expressed from the inactive X chromosome (Xi) and accumulates *cis* to form a domain over the entire chromosome, serving as a trigger for a cascade of chromatin modifications that results in the progressive transition toward a stable, heritable, repressed state (for review, see Heard and Disteche 2006).

Analysis of *Xist* transgenes has revealed that *Xist*-mediated chromosome silencing and Xist RNA localization are

conferred by distinct elements (Wutz et al. 2002). Silencing activity maps in large part to the A repeat, a short tandemly repeated region located at the 5' end of *Xist*. In contrast, localization maps to several redundantly acting elements, including the tandem repeat regions C, E, and F (Wutz et al. 2002; Jeon and Lee 2011; Makhlof et al. 2014; Yamada et al. 2015).

Xist spreading occurs through a sequence of events dictated by the architecture of the X chromosome. Xist RNA searches for binding sites in three dimensions, leading to modification of chromosome structure, before spreading to newly accessible locations (Engreitz et al. 2013; Simon et al. 2013). While the nature of Xist RNA-binding sites is poorly defined, it is known that Xist RNA localizes to the perichromatin compartment that corresponds to the nuclear matrix (NM) (Clemson et al. 1996; Smeets et al. 2014). Accordingly, the NM proteins scaffold attachment

Present addresses: ⁷Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK; ⁸Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.

Corresponding author: dawn.coverley@york.ac.uk

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.295907.117>. Freely available online through the *Genes & Development* Open Access option.

© 2017 Ridings-Figueroa et al. This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution-Non-Commercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

factor A (SAF-A; hnRNPU) and hnRNP UL1 are involved in anchoring Xist RNA within Xi (Helbig and Fackelmayr 2003; Hasegawa et al. 2010; Sakaguchi et al. 2016). Additionally, NM protein 1 (NMP1; YY1) (Guo et al. 1995) has been implicated in tethering Xist RNA to chromatin (Jeon and Lee 2011; Makhlof et al. 2014). In naïve B and T cells, Xist RNA is dispersed in the nucleoplasm but is recruited to Xi upon lymphocyte activation in a process that involves YY1 and SAF-A (Wang et al. 2016).

Recent proteomic and genetic screens have identified several novel *Xist*-interacting factors, some of which now have confirmed roles in *Xist*-mediated silencing (Chu et al. 2015; McHugh et al. 2015; Moindrot et al. 2015; Monfort et al. 2015). With the exception of SAF-A identified previously (Hasegawa et al. 2010), none of the factors investigated in detail was found to affect *Xist* localization (Chu et al. 2015; McHugh et al. 2015).

Random XCI of the paternal or maternal chromosome occurs around embryonic day 5.5 (E5.5) and is propagated through subsequent cell divisions. This can be recapitulated in XX mouse embryonic stem cells (mESCs), allowing fine-tuned analysis of the timing and the earliest events in the initiation process. Examples include deposition of histone 3 Lys27 trimethylation (H3K27me3) and hypoacetylation of H4, which occur rapidly at the onset of Xist RNA expression, and DNA methylation of promoters of X-linked genes and association of the variant histone macroH2A, which occur at later stages of the XCI process (Pollex and Heard 2012). *Xist*-dependent silencing occurs within a developmental window of opportunity corresponding to early stages of differentiation in the XX ESC model (Wutz and Jaenisch 2000). Beyond this time, XCI enters a maintenance phase in which *Xist* is largely dispensable (Brown and Willard 1994; Csankovszki et al. 1999) despite continued expression.

Here we define a novel *Xist* localization factor, the NM protein Cip1-interacting zinc finger protein 1 (CIZ1). CIZ1 was characterized previously as a factor with roles in initiation of DNA replication (Coverley et al. 2005), interacting directly with CDK2, p21/CIP1 (Mitsui et al. 1999), and cyclins (Copeland et al. 2010) and supporting cyclin recruitment to the NM (Copeland et al. 2015). The NM anchor domain in the C-terminal third of CIZ1 (Ainscough et al. 2007) mediates immobilization of CIZ1 and its interaction partners but retains the ability to become incorporated into the NM in the absence of cyclin cargo. CIZ1 has also been linked with post-replicative functions in male germ cell differentiation (Greaves et al. 2012). *CIZ1* transcripts are alternatively spliced to yield at least 22 variants (Rahman et al. 2010), most of which are not characterized. Aberrant alternative splicing underlies its links with a range of pathologies, including pediatric tumors and common adult-onset cancers of the breast (den Hollander et al. 2006) and lung (Higgins et al. 2012) as well as neurological abnormalities (Xiao et al. 2016). Here we describe *Xist*-dependent recruitment of CIZ1 to Xi and a requirement for CIZ1 to maintain Xist RNA localization at Xi in fibroblasts and splenocytes.

Results

CIZ1 accumulates at Xi in female cells

Immunolocalization of endogenous CIZ1 via epitopes in its N-terminal DNA replication domain (Coverley et al. 2005; Copeland et al. 2010) or C-terminal NM anchor domain (Fig. 1A; Ainscough et al. 2007) reveal one or two high-intensity domains within the nucleus of female human or mouse cultured fibroblasts plus smaller nucleus-wide foci in both male and female cells (Fig. 1B; Supplemental Fig. S1A). In immortalized or primary embryonic fibroblasts (PEFs), CIZ1 domains are discrete, while, in cancer-derived cell lines, they are more irregular (shown for MCF7 breast cancer cells in Supplemental Fig. S1A). We hypothesized that the high-intensity domains present only in female cells correspond to Xi. Consistent with this, immunostaining of H3K27me3, a marker for Xi, colocalizes with CIZ1 in PEFs (Fig. 1B) and a range of other female cell types (Supplemental Fig. S1A). CIZ1 did not colocalize with the active chromatin mark H3K4me3 or constitutive heterochromatin mark H3K9me3 (Supplemental Fig. S1B). Identification of the CIZ1 domains observed in female PEFs as the silenced X chromosome was confirmed by immuno-FISH for CIZ1 and Xist RNA (Fig. 1C), which revealed localization within the same chromosome territory.

To determine at which stage of the XCI process CIZ1 is recruited, we analyzed CIZ1 localization in PGK12.1 XX mESCs at time points following the initiation of differentiation. CIZ1 localization to Xi was observed from day 1 and persisted throughout the time course (Fig. 1D, Supplemental Fig. S1C). This closely correlates with the dynamics of Xist RNA expression reported previously (Sheardown et al. 1997). As in PEFs, CIZ1 domains colocalized with H3K27me3, identifying their location as Xi. CIZ1 is lost from Xi in late metaphase in both ESCs (Fig. 1E) and PEFs (Supplemental Fig. S1D), indicating a cycle of recruitment and loss that is similar to Xist RNA (Duthie et al. 1999). The smaller nucleus-wide foci remain qualitatively similar throughout ESC differentiation but are excluded from chromosomes in late metaphase (Supplemental Fig. S1D; Greaves et al. 2012).

Superresolution three-dimensional structured illumination microscopy (SR 3D-SIM) of endogenous CIZ1 together with endogenous Xist RNA in diploid female somatic C127I cells confirmed their adjacent localization, similar to that for SAF-A (Fig. 2A; Supplemental Fig. S2B; Smeets et al. 2014).

NM association of CIZ1 at Xi

The NM is a biochemically defined fraction that resists extraction from the nucleus and is thought to anchor and spatially organize nuclear processes, including DNA replication and repair, transcription, and pre-mRNA splicing (Wilson and Coverley 2013). Serial extraction (Wilson et al. 2016) to reveal the fraction of CIZ1 that remains in cells after solubilization with (1) low-level detergent under physiological salt concentrations, (2) 0.5 M salt, or

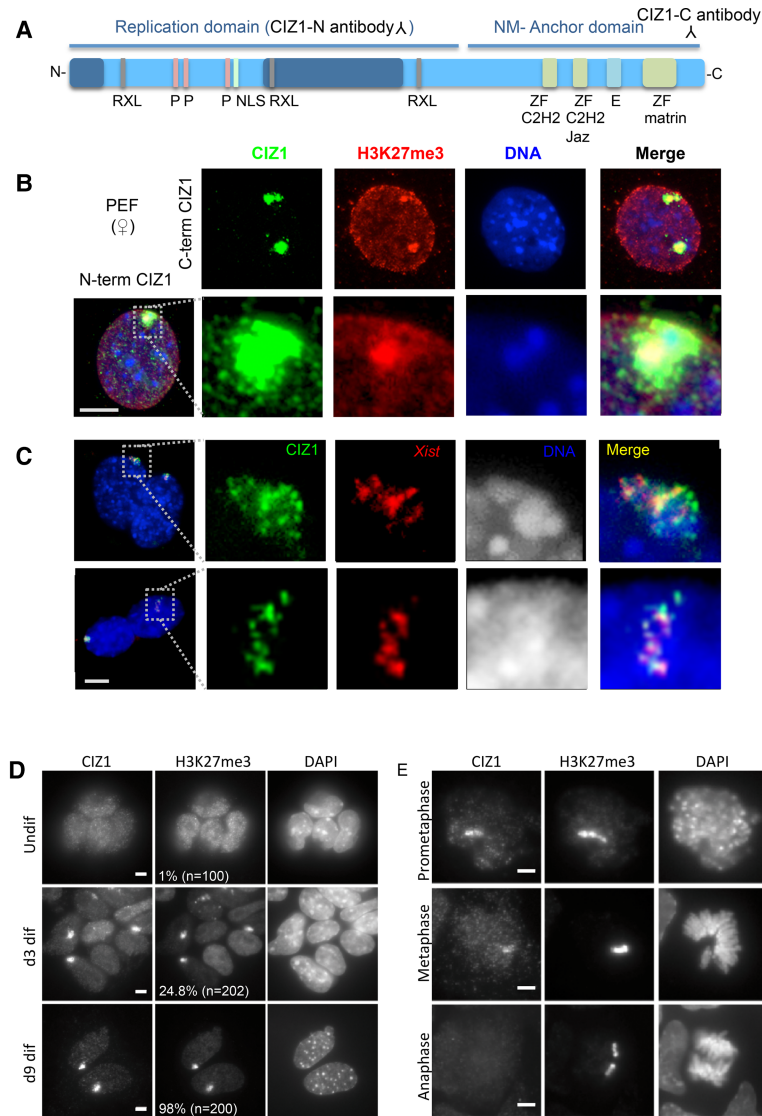


Figure 1. CIZ1 is enriched at Xi. (A) Schematic of CIZ1 indicating the replication domain (Copeland et al. 2015) with the nuclear localization signal (NLS, green), functional CDK phosphorylation sites (pink), and RXL cyclin-binding motifs (gray) and the NM anchor domain (Ainscough et al. 2007) with the locations of C2H2-type zinc fingers (green), the Matrin3-type RNA-binding zinc finger (Prosite PS50171), and the acidic domain (labeled E). The locations of epitopes recognized by N-terminal and C-terminal antibodies are indicated. (B) Immunodetection of CIZ1 (green) in female PEFs from wild-type mice using N-terminal and C-terminal antibodies. Colocalization of CIZ1 and histone H3K27me3 (red) was observed as a discrete domain in all cells with H3K27me3 staining. $n > 100$. DNA was stained with DAPI (blue). Additional cell lines are shown in Supplemental Figure S1; some have two domains, indicative of chromosomal duplication. (C) RNA-FISH for *Xist* (red) in female PEFs showing colocalization with CIZ1 protein (green) and the Barr body (gray). Bar, 5 μ M. (D) CIZ1 recruitment to Xi in differentiating XX ESCs correlates with *Xist*-mediated deposition of H3K27me3. d3 and d9 are days of differentiation after withdrawal of LIF. Additional time points are shown in Supplemental Figure S1C. (E) CIZ1 and H3K27me3 in differentiated XX ESCs during mitosis showing a reduction of CIZ1 in late metaphase and complete loss in anaphase but retention of H3K27me3. Bar, 5 μ m.

(3) nuclease (DNase or RNase) (Fig. 2B) revealed distinct populations in female 3T3 cells (Fig. 2C) and PEFs (Supplemental Fig. S2C). While the small nucleus-wide foci remained under all conditions (part of the core protein NM), the high-intensity domain at Xi was released by digestion with RNase but not DNase. This resistance to high salt and DNase defines CIZ1 at Xi as part of the NM, but release by RNase shows this to be the RNA fraction of the NM and is consistent with its close association with *Xist* RNA.

When 3T3 cells were treated with the protein-protein cross-linker DTSP prior to extraction (Fig. 2D), the CIZ1 domain at Xi was rendered resistant to digestion with RNase. This suggests that it is in close proximity to proteins in the core NM, possibly the resistant fraction of CIZ1. Thus, two qualitatively different populations of NM-anchored CIZ1 are present in the nucleus, but most of the CIZ1 at Xi is anchored by association with RNA (Fig. 2E).

Similar analysis of the NM proteins SAF-A and YY1 in 3T3 cells showed that they are not enriched at Xi, that a

NM-associated population can be revealed by removal of chromatin, and that both proteins are completely extracted by digestion of RNA (Supplemental Fig. S3). All three of these features are consistent with the published literature but distinguish SAF-A and YY1 from CIZ1.

Recruitment to Xi requires the CIZ1 C-terminal NM anchor domain

To ask whether recruitment of CIZ1 to Xi is mediated by the sequences that support attachment to the NM (Ainscough et al. 2007), the C-terminal anchor domain (GFP-C275, which includes C2H2-type zinc fingers and Matrin3-type RNA-binding zinc finger domains) and the N-terminal DNA replication domain (GFP-N572, which includes CDK phosphorylation sites and cyclin-binding motifs) were transiently transfected separately into 3T3 cells, and the frequency of accumulation at the Xi territory was scored after one cell cycle. N572 completely failed to accumulate at Xi, while C275 accumulated in large foci

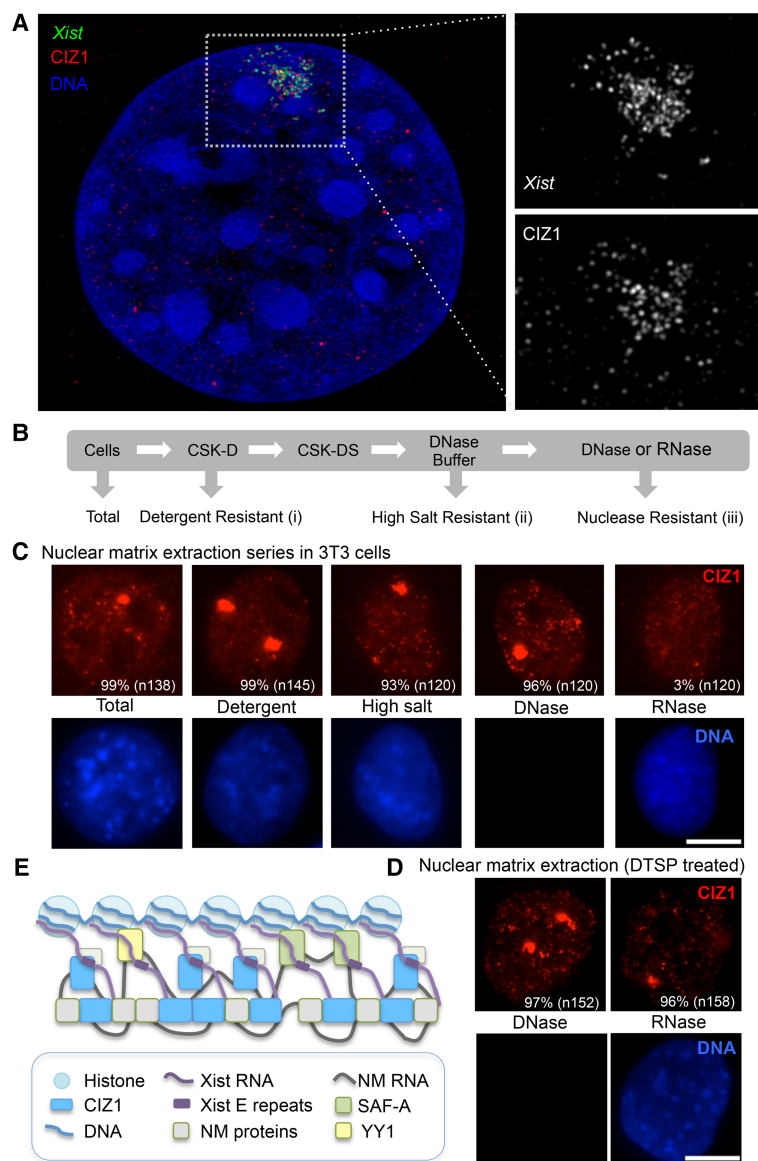


Figure 2. CIZ1 is part of the RNA-dependent NM at Xi. (A) Maximum intensity projection SR 3D-SIM image of a single C1271 cell nucleus showing adjacent localization of *Xist* foci (green) with CIZ1 foci (red) at Xi. Examples of individual Z sections from several cells are shown in Supplemental Figure S2. (B) Schematic showing the protocol for serial extraction with detergent, high salt, DNase, or RNase to reveal the protein-RNA NM fraction or the RNA-independent NM fraction (protein only). (C) Images show CIZ1 (red) after serial extraction of 3T3 cells. The proportion of cells with discrete CIZ1-Xi domains is indicated ($n > 100$ for each condition), some with two domains, indicating that a proportion of cells is tetraploids with two Xi. Similar results were obtained with PEFs (Supplemental Figure S2C). DNA (blue) shows the extent of nuclease treatment. Images were equally modified to allow direct comparison of fluorescence intensity across the different extraction conditions. Bar, 5 μ m. (D) As in C but with prior protein-protein cross-linking with DTSP. (E) Model showing two populations of CIZ1 in the NM (blue); RNA-dependent CIZ1 interacts with *Xist*, and RNA-independent CIZ1 is part of the core NM. SAF-A (Hasegawa et al. 2010) is also shown interacting with *Xist* E repeats and is depicted with YY1 in the RNA-dependent NM.

that colocalized with endogenous CIZ1 at Xi (Fig. 3A). However, compared with GFP-full-length CIZ1, the frequency of C275-marked X chromosomes was significantly reduced despite their presence in the nucleus. Thus, while sequences encoded in N572 are not sufficient to specify recruitment to Xi on their own, they do increase the efficiency of targeting. Live-cell imaging of a stably transfected P4D7F4 XY mESC line, which carries an mCherry-tagged inducible *Xist* transgene (Moindrot et al. 2015), revealed colocalization between C275 and *Xist* RNA domains, whereas N572 showed no *Xist* colocalization (Fig. 3B). Together, these findings support a key role for the C terminus of CIZ1 in binding at Xi.

Recruitment of CIZ1 by *Xist* requires the *Xist* E repeat region

In undifferentiated male MG-3E (XY) ESCs carrying an inducible *Xist* transgene, CIZ1 is recruited to the *Xist*

domain and shows an adjacent localization to *Xist* similar to that in female ESCs (Supplemental Fig. S4A). To define elements in *Xist* RNA required for CIZ1 recruitment, we analyzed a series of inducible transgenic *Xist* deletion constructs in XY ESCs (Fig. 3C; Supplemental Fig. S5) and a deletion of *Xist* exon IV from the endogenous *Xist* locus in female mouse embryonic fibroblasts (MEFs) (Supplemental Fig. S4B; Caparros et al. 2002). The deletions encompassed key elements, including six short tandem repeat regions (A–F) (Brockdorff et al. 1992; Brown et al. 1992; Nesterova et al. 2001), which are conserved and, in some cases, have been shown to be functionally important. CIZ1 recruitment was found to be independent of the A repeat region, which is required for *Xist*-mediated silencing (Wutz et al. 2002), and the XN region (repeats B and F), which is involved in the recruitment of PRC2 to Xi (Fig. 3C; da Rocha et al. 2014). However, a truncated *Xist* construct, which corresponds to the first 3 kb of *Xist*, does not recruit CIZ1, implicating regions further

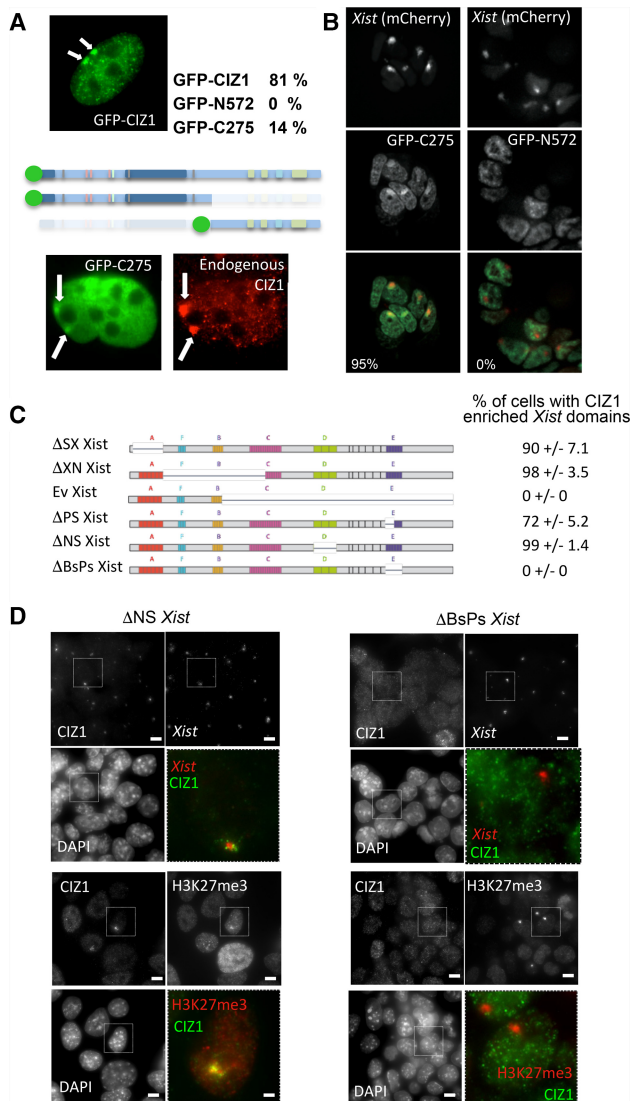


Figure 3. Delineation of the *Xist* and CIZ1 domains. (A) Recruitment of the indicated GFP-tagged CIZ1 constructs (Coverley et al. 2005) 24 h after transient transfection into cycling wild-type PEFs. The proportion of transfected cells with nuclear GFP, in which accumulation at Xi was observed ($n > 100$ for each construct), is indicated with representative images. For GFP-C275, endogenous CIZ1 (red) is also shown, detected via epitopes in the N-terminal end. (B) Accumulation of GFP-C275 but not GFP-N572 at Xi in stably expressing ESC lines that also carry inducible *Xist* RNA tagged with BglI stem-loops that bind a BglI-G-mCherry fusion protein (Moindrot et al. 2015). Almost all (58 out of 61) *Xist*-mCherry-expressing cells were positive for GFP-C275, while none ($n = 47$) was positive for GFP-N572. (C) Schematic of inducible *Xist* constructs transfected into XY ESCs to study CIZ1 recruitment. The result of CIZ1 localization studies in the transgenic cell lines is summarized. (D) Example images showing the lack of CIZ1 colocalization with H3K27me3 domains (bottom panels) or *Xist* (top panels) in a Δ BsPs *Xist* construct missing the E repeat. The absence of the *Xist* D repeat (Δ NS) does not affect recruitment of CIZ1. Data for all transgenes are shown in Supplemental Figure S5.

3'. Accordingly, deletion of the E repeats, encompassing a 1.5-kb span of *Xist* exon 7, entirely abolished CIZ1 recruitment by *Xist* RNA (Fig. 3D). Deletion of other 3' regions, the D repeats, or the highly conserved *Xist* exon 4 had no effect (Fig. 3C,D; Supplemental Fig. S4B). These findings demonstrate a requirement for *Xist* E repeats for recruitment of CIZ1 and together raise the possibility that the C terminus of CIZ1 might functionally interact with *Xist* RNA via the E repeat region.

Functional analysis of CIZ1 in XCI

Loss-of-function mutations affecting factors critical for XCI, including *Xist* RNA, result in female-specific lethality, usually during early or mid-stages of embryogenesis. To determine whether this is the case for CIZ1, targeted C57BL/6 ESCs generated using a gene trap strategy were used to produce heterozygous knockout mice (Supplemental Fig. S6A,B). Viable *ciz1*^{-/-} male and female F1 progeny were born at the expected ratio (Supplemental Table S1), showed no difference in growth rate, and had no overt developmental defects. Loss of *ciz1* transcript was confirmed in embryos (E12) and fibroblasts from 3-wk postnatal tail tip dermal tissue (tail tip fibroblasts [TTFs]) (Supplemental Fig. S6C). Loss of protein expression was confirmed in TTFs, lymphocytes (Supplemental Fig. S6D,E), and differentiating male germ cells, which normally express high levels of CIZ1 (Supplemental Fig. S6F; Greaves et al. 2012). Thus, the gene trap insertion abrogates expression from the CIZ1 locus in vivo and in vitro, demonstrating that CIZ1 is not essential for embryogenesis, early postnatal development, or cell viability ex vivo.

The absence of an embryonic phenotype suggests that CIZ1 is not required for the establishment of a transcriptionally quiescent, inactivated X chromosome despite recruitment during *Xist*-dependent initiation of X inactivation (Fig. 1D). Consistent with this, the transcriptome of CIZ1-null-derived female PEFs did not reveal widespread reactivation of Xi compared with wild-type controls (Fig. 4A; Supplemental Data Set S1). As expected, the *Ciz1* gene was silenced in null cells ($P = 5.00 \times 10^{-05}$; $q = 0.005$) (Supplemental Data Set S2), but comparison of all transcripts that map to the X chromosome of the *Mus musculus* C57BL/6 primary assembly GRCm38 (downloaded from <http://www.ensembl.org> on May 4, 2016) showed that most were not significantly altered and revealed little change in genes associated with the X inactivation center (XIC) (Fig. 4B). The lack of widespread reactivation is in line with similar analyses and the understanding that loss of *Xist* RNA or other factors does not significantly compromise the maintenance of XCI (Csankovszki et al. 1999).

However, deregulation at the single-gene level was significant ($P < 0.05$) for 62 X-linked transcription units dispersed across the X. This is 3.6% of those that are expressed in PEFs and includes a similar number of up-regulated and down-regulated genes and six where $q < 0.05$ [*Agtr2*, *Fhl1*, *Tmem164*, *Gpm6b*, XLOC3750, and XLOC830] (Supplemental Data Set S2). Induction of the

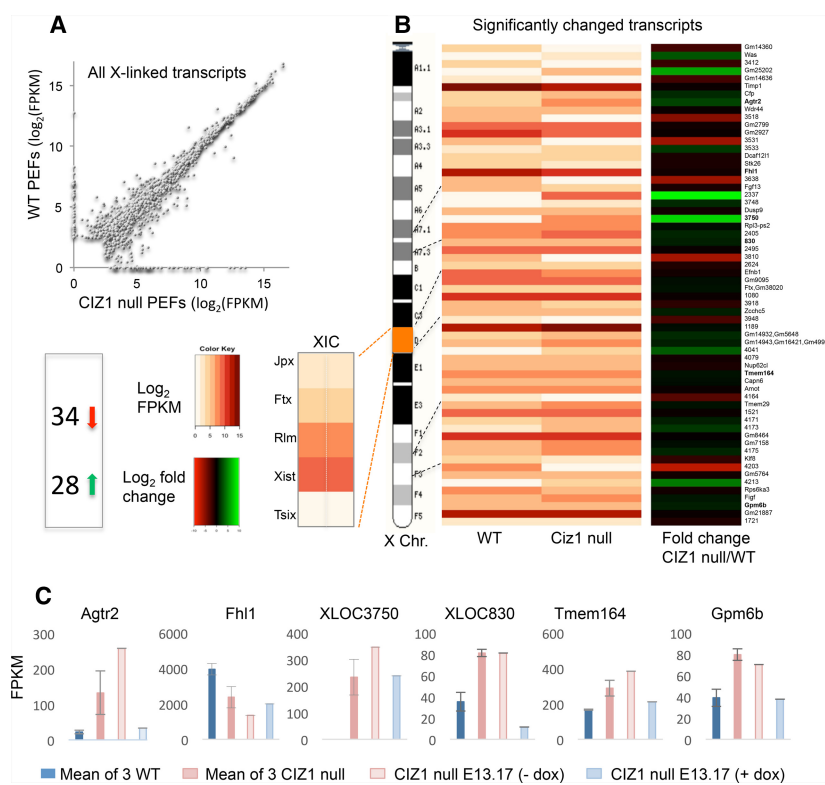


Figure 4. Expression of X-linked genes. (A) Scatter plot showing mean expression in three wild-type and three CIZ1-null PEF lines (log₂ FPKM [fragments per kilobase per million mapped fragments]) for all expressed X-linked transcription units. FPKM <0.99 were rounded to 1. The mean expression for all X-linked transcription units is in Supplemental Data Set S1. (B) Heat map showing (white to brown) expression levels in log₂ FPKM for 62 X-linked genes that are significantly changed in CIZ1-null PEFs. $P < 0.05$. Genes are listed in order against a schematic of the X chromosome. Unannotated transcripts are indicated by XLOC gene ID number (Supplemental Data Set S1), and predicted genes are indicated by the prefix Gm. A list of significantly changed transcription units is in Supplemental Data Set S2, of which 35 are annotated, and 23 have known functions. (Right) Fold change showing the 34 down-regulated (red) and 28 up-regulated (green) transcription units distributed across the chromosome. (Left of the X-chromosome schematic) Also shown are results for genes at the XIC. (C) Expression of six X-linked transcription units where $q < 0.05$, showing mean FPKM for three wild-type and three CIZ1-null cell lines as well as the effect of reinduction of CIZ1 in null-derived transgenic primary PEF line e13.17.

full-length GFP-CIZ1 transgene in PEFs derived from CIZ1-null line e13.17 harboring an inducible CIZ1 transgene and doxycycline-regulatable transactivator (see the Supplemental Material) rebalanced the expression of all six genes, modulating four of them back to wild-type levels (Fig. 4C). Thus, expression of full-length CIZ1 compensates for genetic ablation and presumptive loss of the full repertoire of *Ciz1* variant transcripts in the regulation of these genes.

Lymphoproliferative disorder in adult female mice

Although we observed no overt defects in embryogenesis or early postnatal development (Fig. 5), progressive infirmity was observed in female CIZ1-null mice from 9 mo onward. Eight females and an equivalent number of males were therefore evaluated for abnormalities between 9 and 19 mo. This revealed lymphoproliferative disorder in all eight females and none of the males. Detailed histological assessment was undertaken for six of the females and compared with six wild-type females. A summary assessment of abnormalities in the spleens, livers, lungs, and lymph nodes for individual CIZ1-null and wild-type females is in Supplemental Table S2, and pathology notes describing histological assessments are in Supplemental Tables S3 and S4. Notably, primary and secondary lymphoid tissues (spleen, lymph node, lung, and liver) were enlarged in all *Ciz1*^{-/-} adult females (Fig. 5C). Secondary lymphoid tissues are sites where B and T lymphocytes are directed in search of antigen, leading to the regulated turnover or amplification of subsets of cells within germinal

centers. This process is deregulated most notably in the spleen (Fig. 5C,D), which displayed a fivefold enlargement in *Ciz1*^{-/-} (181–3679 mg) compared with *Ciz1*^{+/+} (88–167 mg) mice. Histologically, lymph node and spleen architectures were abnormal, with effacement of normal follicles and significant infiltration of abnormal B (CD20 +ve) and reactive T (CD3 +ve) lymphocytes in all affected tissues (Fig. 5D,E). At the cellular level, the disorder resembled non-Hodgkin follicular-type lymphoma, with three showing evidence of high-grade transformation consistent with diffuse large B-cell lymphoma (Ward 2006). These data point to compromised XCI in lymphoid lineages and suggest that CIZ1 normally protects against tumor formation.

CIZ1 is required for *Xist* RNA localization in fibroblasts and splenocytes

While the viability of CIZ1-null embryos suggests that CIZ1 is not critical for the establishment of XCI, female-specific lymphoid hyperproliferation nevertheless implies an important lineage-restricted function. To further investigate this, we performed RNA-FISH to analyze *Xist* domains in PEFs and splenocytes derived from CIZ1-null mice. In independently derived CIZ1-null fibroblast cell lines, we observed a strikingly dispersed *Xist* signal that occupies 40% of the nuclear area compared with <5% in wild-type cells as well as loss of H3K27me3 (Fig. 6A). Dispersal cannot be attributed to increased *Xist* levels, as none of the three CIZ1-null PEFs showed any change in *Xist* transcript (Supplemental data set S1). X-chromosome

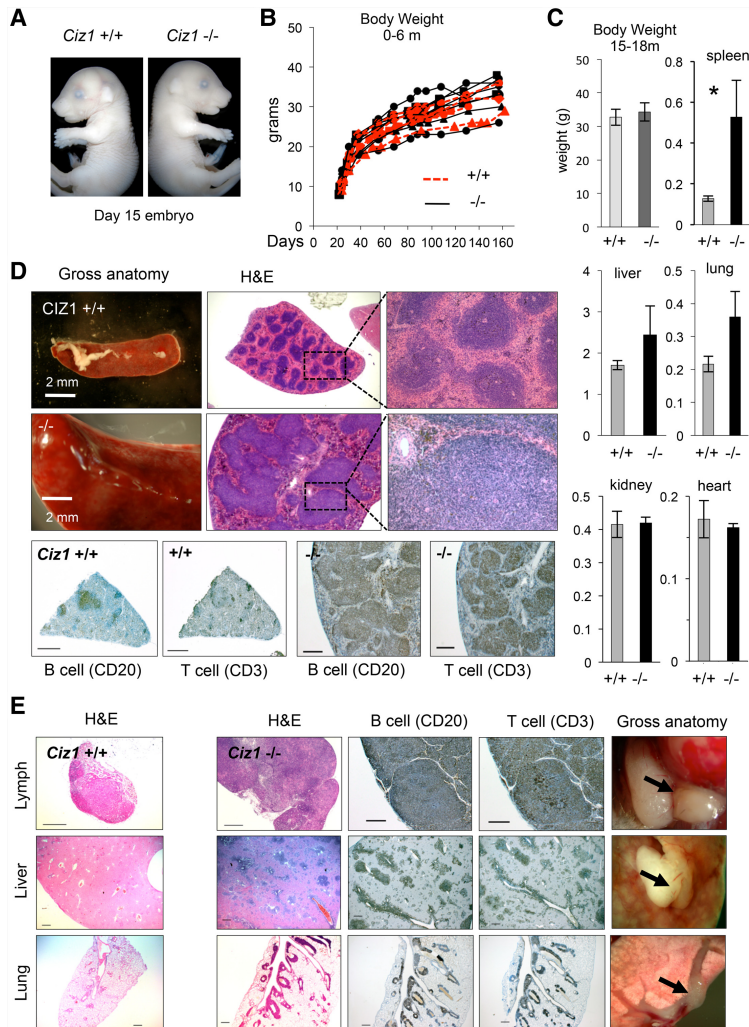


Figure 5. CIZ1-null mice develop normally but show gross lymphoid abnormalities in adult females. (A) *Ciz1*^{-/-} embryos are indistinguishable from wild-type littermates at E15. (B,C) Growth profiles of *Ciz1*^{+/+} (*n* = 5) and *Ciz1*^{-/-} (*n* = 8) mice between days 20 and 160 after birth (B) and at 15–18 mo old in *Ciz1*^{+/+} (*n* = 8) and *Ciz1*^{-/-} (*n* = 7) females (C). No significant differences were detected. However, *Ciz1*^{-/-} females had enlarged spleens (*n* = 8 *Ciz1*^{+/+}; *n* = 6 *Ciz1*^{-/-}), livers (*n* = 6 *Ciz1*^{+/+}; *n* = 4 *Ciz1*^{-/-}), and lungs (*n* = 5 *Ciz1*^{+/+}; *n* = 5 *Ciz1*^{-/-}). Other organs, including the kidney and heart, were not affected. (D) Representative image of gross spleen enlargement in *Ciz1*^{-/-} females, with histological sections stained with hematoxylin and eosin (H&E). Lymphoid cell nuclei (stain darkly) are highly organized into foci in *Ciz1*^{+/+} spleens but not *Ciz1*^{-/-} spleens. (Right) High-magnification images show morphology consistent with lymphoproliferative disorder in *Ciz1*^{-/-} mice. (Below) Immunohistochemical detection of CD20 and CD3 (B-cell-specific and T-cell-specific, respectively) suggests B-cell lymphoma with T-cell infiltration. Positive cells are stained dark gray and show overlapping distribution. Bar, 200 μ m. (E) Representative H&E staining of *Ciz1*^{+/+} and *Ciz1*^{-/-} female lymph nodes, livers, and lungs. Enlargement of secondary lymphoid tissues in *Ciz1*^{-/-} females correlates with excess proliferation of lymphoid cells, as in D. (Right) Examples of gross tissue anatomy in *Ciz1*^{-/-} females showing areas of lymphoproliferative disorder as pale outgrowths. Bar, 200 μ m.

paints showed no significant difference between wild-type and CIZ1-null PEFs (Fig. 6B), suggesting that there is a deficit in Xist RNA localization rather than Xi organization.

Further substantiating the conclusion that CIZ1 plays a role in Xist RNA localization, induction of the full-length GFP-CIZ1 transgene (Fig. 6C–E) fully reinstated the localization of Xist RNA over Xi domains (Fig. 6F,G). Prior to induction of CIZ1, Xist was dispersed in >80% of CIZ1-null e13.17 PEFs but became relocalized to discrete domains that overlap with GFP-CIZ1 domains within 20 h (Fig. 6H). Together, these observations demonstrate that CIZ1 plays a key role in Xist RNA localization in PEFs.

In light of the female-specific lymphoproliferative disorder observed in CIZ1-null animals, we went on to investigate the role of CIZ1 in X inactivation in hematopoietic lineages. We evaluated the impact of CIZ1 deletion in splenocytes from 6-wk-old females after stimulation of mixed populations of naïve B/T cells with either the B-cell activator lipopolysaccharide (LPS) or T-cell activator α CD3 antibody (α CD3). Consistent with a previous report (Wang et al. 2016), activation of both cell types from wild-type mice induced dramatic focal localization of Xist to Xi

within 24 h, and this was mirrored by accumulation of CIZ1 (Fig. 7). However, activated B and T lymphocytes from CIZ1-null mice failed to efficiently localize Xist RNA to the Xi territory (Fig. 7B). This finding identifies a transition point in the affected lineages that is compromised in CIZ1-null animals. To ask whether aberrant Xist localization leads to relaxed control over X-linked genes, we compared the transcriptomes of wild-type spleens (containing mostly naïve cells) and CIZ1-null spleens (containing hyperproliferative cell populations likely expanded from rare activated precursors) from adult mice. Overall, from the 2209 X-linked genes that returned test data (Supplemental Fig. S7A; Supplemental data set S2), 16.4% were up-regulated and 8.7% were down-regulated by greater than twofold (Supplemental Fig. S7B). As expected, whole-genome gene set enrichment analysis returned highly significant overlap with immunological processes and cell division gene ontology terms (Supplemental Fig. S7C; Supplemental data set S3). Comparison of the X-linked genes with the gene list reported to be up-regulated in blood cells of Xist mutant mice (Yildirim et al. 2013) showed that many of the same genes are

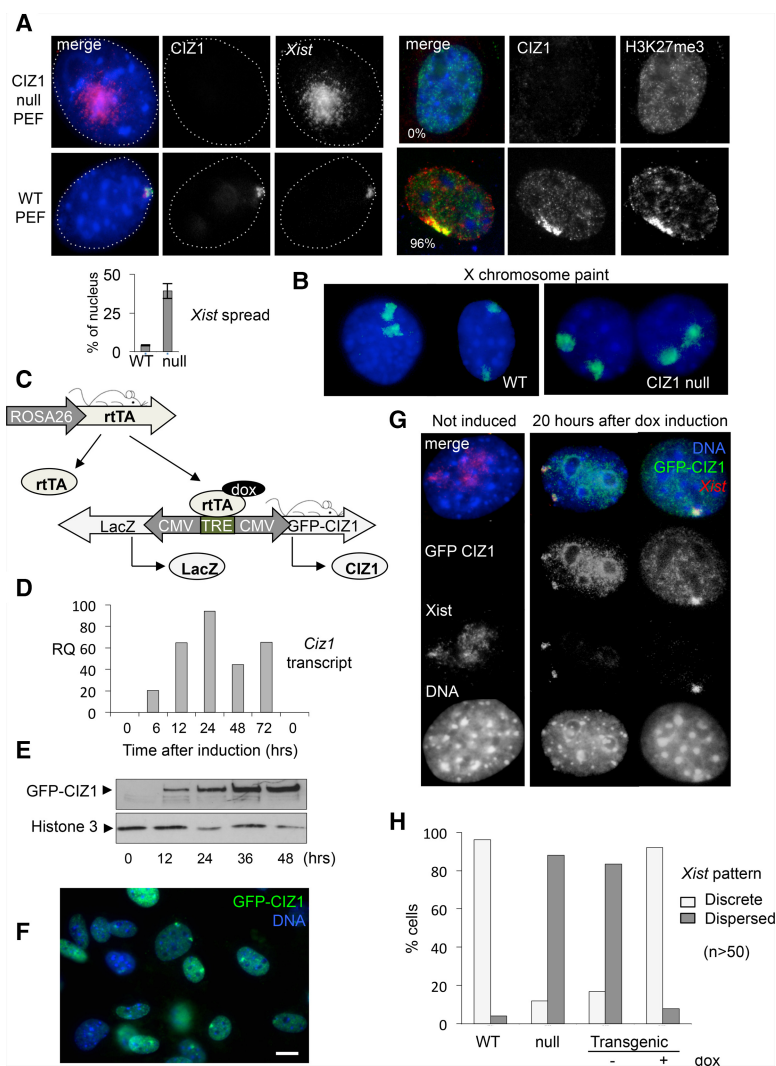


Figure 6. Loss and reinstatement of *Xist* localization at Xi is dependent on CIZ1. (*A, left*) Immunofluorescence showing CIZ1 and *Xist* RNA, which is delocalized in CIZ1-null (*ciz1*^{-/-}) PEFs. DNA was stained with DAPI (blue). (*Below*) Quantitation of the area of *Xist* FISH signal showing mean distribution over ~40% of the nucleus in CIZ1-null PEFs compared with <5% in wild-type cells. (*Right*) CIZ1 and H3K27me3 in wild-type and CIZ1-null cells showing the proportion of cells with marked Xis. *n* = 100. (*B*) X-chromosome paint shows similar X-chromosome territories in CIZ1-null and wild-type PEFs. (*C*) Double-transgenic female PEFs were derived from embryos harboring a tetracycline-responsive *Ciz1* responder transgene and reverse transactivator transgene on the CIZ1-null background. (*D,E*) *Ciz1* transcript (primers Mm00503766_m1) (*D*) and protein (N-terminal antibody in whole-cell lysates) (*E*) were detected after doxycycline induction. (*F*) Representative field view of GFP-CIZ1 expressed from the transgene on a CIZ1-null background 24 h after induction with doxycycline. Note the presence of two domains in some cells, indicating two Xis in tetraploid cells. Bar, 10 μ m. (*G*) Expression of full-length CIZ1 in *ciz1*^{-/-} PEFs leads to relocalization of *Xist* to the Xi territory. (*H*) Quantitation of the proportion of cells with *Xist* FISH signal that is “dispersed” (defined as occupation of >10% of the nuclear area). CIZ1 transgene induction reverts CIZ1-null cells to apparent normality for this criterion by 20 h.

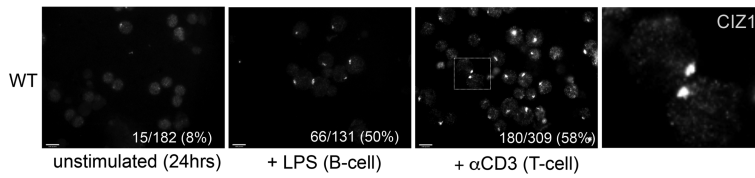
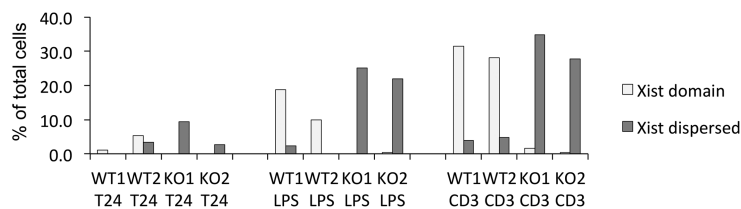
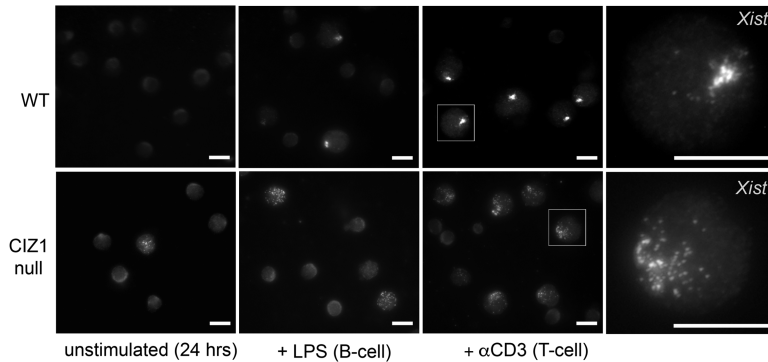
affected (Supplemental Fig. S7D; Supplemental data set S4); however, a similar proportion of genes was affected genome-wide (Supplemental Fig. S7B). Together, the data demonstrate that CIZ1 plays a key role in stabilizing *Xist* association with Xi in lymphoid lineages but that its effects are not limited to the X chromosome.

Discussion

Here we demonstrate that the NM protein CIZ1 is strongly enriched over the Xi territory. Although CIZ1 was not previously linked to XCI, one of four recent screens identified CIZ1 among 81 candidate *Xist* interactors (Chu et al. 2015). Based on our findings, a relationship between CIZ1 and *Xist* is clear, although whether localization of CIZ1 to Xi domains is attributable to a direct interaction with *Xist* RNA or an interaction with other *Xist* or Xi-bound factors remains to be seen. Several observations support a functional interaction. First, loss of CIZ1 results in the dispersal of *Xist* in somatic cells. Second, CIZ1 enrichment occurs rapidly at the onset of *Xist* RNA expres-

sion and is present in all observed cases in which *Xist* RNA domains are observed. Third, SR 3D-SIM analyses demonstrate that CIZ1 and *Xist* RNA lie in very close proximity and that CIZ1 is localized to the RNA-dependent NM compartment at Xi. Finally, the C terminus of CIZ1, which anchors it at Xi, encompasses known RNA-binding domains, notably the Matrin3-type zinc finger, suggesting a possible direct interaction. Interestingly, Matrin3 has also been identified as a candidate *Xist* interactor (Chu et al. 2015; Moindrot et al. 2015). Further studies are nevertheless required to determine whether there is a direct interaction between CIZ1 and *Xist* RNA, specifically the E repeat region, and, at this stage, we cannot rule out that CIZ1 recognizes other *Xist*-interacting protein; for example, PTBP1 and PTBP2, recently shown to bind the E repeat region (Chen et al. 2016).

Several studies have pointed to a role for the NM in anchoring *Xist* RNA within the Xi territory, and a number of other proteins that interact with the NM or with DNA sequences that interact with the NM (S/MARs) have been implicated in XCI, including SAF-A, YY1, and SATB1. SAF-A and CIZ1 are similar in that their ability to support

A CIZ1 localisation in mouse splenocytes**B** *Xist* localisation in mouse splenocytes

Xist-mediated gene silencing and recruitment of *Xist* to Xi, respectively, is dependent on *Xist* E repeats (Hasegawa et al. 2010), possibly identifying a common mechanism. Moreover, deletion of the E repeats was shown recently to result in dispersed localization of *Xist* RNA (Yamada et al. 2015). These findings were attributed to SAF-A; however, our results suggest that loss of interaction with CIZ1 might contribute to the observed phenotype. Notably, the sensitivity of both SAF-A and YY1 to digestion of the RNA component of the NM distinguish them from CIZ1, which is part of the core protein matrix throughout the nucleus, suggesting important differences in their roles.

Scaffolding the structural reorganization of Xi during XCI or the maintenance of the compacted structure are possible functions for the NM to which CIZ1 might contribute. Another possibility, suggested by the function of CIZ1 in DNA replication, is the regulatable recruitment of factors into or away from the Xi territory. During late G1 phase, CIZ1 supports the recruitment of cyclin A to the NM. Cyclin docking on CIZ1, but not CIZ1 recruitment to the NM, is switched off at S phase (Copeland et al. 2015). Thus, in the context of DNA replication, CIZ1 appears to be a cargo carrier or mediator that is sensitive to cell cycle stage, raising the possibility of a relationship with cell cycle regulators implicated in *Xist* retention (Hall et al. 2009).

Although our observations in PEFs and lymphocytes implicate CIZ1 in anchoring *Xist* to the NM, the fact

Figure 7. CIZ1 modulates *Xist* localization in splenocytes (A) CIZ1 localization in wild-type splenocytes before and after activation with LPS or α CD3. Stimulation causes the accumulation of CIZ1 at Xi in both B and T lymphocytes within 24 h. Bar, 10 μ m. (B) *Xist* RNA localization in splenocytes before and after activation with LPS or α CD3. Similar to CIZ1, stimulation causes accumulation of *Xist* at Xi of wild-type B and T lymphocytes within 24 h, whereas *Xist* RNA is not properly localized in CIZ1-null cells. Bar, 10 μ m. $n = 200$ –300 per group.

that CIZ1-null females are viable implies that the same relationship might not apply during early embryogenesis. Based on this, we hypothesize that CIZ1 functions redundantly with other anchoring factors (for example, SAF-A) and that these are sufficient during embryogenesis but insufficient in PEFs or lymphocytes. A recent report suggested that SAF-A is not essential for *Xist* localization in all lineages and that other factors may compensate for its loss (Kolpa et al. 2016). However, this is in contrast to genetic depletion of SAF-A in MEFs, which showed a requirement for SAF-A (Sakaguchi et al. 2016). Thus, the relationship between CIZ1 and SAF-A is not known, and there may be more factors capable of anchoring *Xist* at Xi.

Here we show that the compact localization of *Xist* upon the stimulation of lymphoid lineages (Wang et al. 2016) is dependent on CIZ1, describing a role for CIZ1 in a lineage-restricted transition that occurs throughout the lifetime of mice. An *Xist*-dependent silencing pathway was reported previously to be transiently activated during hematopoietic differentiation (Savarese et al. 2006), with pre-B and pre-T cells the most dependent. This aligns closely with our observation of B-cell and T-cell hyperplasia in CIZ1-null mice and is consistent with independent analysis that described a susceptibility to oncogene-induced transformation leading to leukemias in the absence of CIZ1, although no sex bias was reported in this study (Nishibe et al. 2013).

We observed widespread deregulation of gene expression in the absence of CIZ1 in affected lineages, although this was not specific to the X chromosome. However, we cannot rule out that low-level reactivation of multiple X-linked genes together, considered collectively, confers the female-specific lymphoproliferation phenotype. Thus, it remains an open question whether the disorder observed in CIZ1-null mice is a consequence of compromised XCI. Proliferative disorders of the hematopoietic system have been associated with deletion or suppression of factors linked with XCI (Leong et al. 2013; Yildirim et al. 2013), and abnormalities of XCI are reported frequently in cancers, including duplication of the active X (Xa) in breast and ovarian cancers and leukemias (Spatz et al. 2004; Lee and Bartolomei 2013). However, X chromosomes carry proportionally more immune-related genes than the rest of the genome (Bianchi et al. 2012), which means that a more general failure of the control of gene expression might manifest preferentially in females. Transcriptome analysis in PEFs identified candidate X-linked drivers of hematopoietic malignancies; *Gpm6b* overexpression is linked with B-cell lymphoma (Charfi et al. 2014), and *Figf* (VEGF-D) is implicated in the metastatic spread of tumors via lymph nodes (Bardelli et al. 2007; Pazgal et al. 2007). However, we interpret this with caution because changes elsewhere in the genome, initiated directly or indirectly as a consequence of the loss of CIZ1, are also likely to play a role.

In conclusion, we defined a novel component of the X inactivation pathway: the NM protein CIZ1. Our results indicate that CIZ1 functionally interacts with Xist RNA via the E repeat region and, moreover, that CIZ1 facilitates in *cis*-localization of Xist RNA, functioning as an anchor to the NM in somatic cell lineages.

Materials and methods

Further details are available in the [Supplemental Material](#).

Animals and genotyping

All animal work was carried out under a UK Home Office license. CIZ1-null mice were generated from C57BL/6 ES clone IST13830B6 (TIGM) harboring a neomycin resistance gene trap inserted downstream from exon 1. The absence of *Ciz1*/CIZ1 in homozygous progeny was confirmed by quantitative PCR, immunofluorescence, and immunoblot with CIZ1 N-terminal antibody. Inducible *GFP-Ciz1*-Tg mice, generated by pronuclear injection of a GFP full-length *Ciz1* construct into CBA/C57BL6 fertilized eggs, were crossed with ROSA26-rtTA mice (Jackson Laboratories). All primers used for the characterization of *Ciz1* targeting and the detection of transactivator and responder transgenes and sex are in the [Supplemental Material](#).

Cell lines

All stable cell lines were grown following standard procedures. Mouse PEFs were derived from individual embryos at days 13–14 of gestation. Primary TTFs were generated from individual 3-wk-old mice. Genotype and sex were confirmed after explant culture using primers listed in the [Supplemental Material](#). For in-

ducible cells harboring ROSA26-rtTA and *GFP-Ciz1*-Tg transgenes, the addition of 5 µg/mL doxycycline to medium was used to induce detectable GFP-CIZ1 within 6 h. ESCs were grown on feeders with the addition of LIF. Where applicable, *Xist* expression was induced with doxycycline at 1.5 µg/mL. Male XY P4D7 ESCs, derived from the cross between *Mus castaneus* and 129+*Ter*/SvJcl and containing an rtTA cassette in the Gt(ROSA)26Sor locus (Moindrot et al. 2015) were used to generate stable autosomal integrants of *Xist*-inducible deletion variants.

Splenocyte isolation and activation

Spleens isolated from 6-wk-old wild-type and *Ciz1*-null females were pressed through 70-µm nylon filters to dissociate naïve B and T lymphocytes into medium (RPMI 1640; Invitrogen) supplemented with 10% fetal calf serum, 100 µ/mL penicillin, 10 µg/mL streptomycin, and 2 mM L-glutamine. The cells were pelleted at 450 g for 5 min and then resuspended in red blood cell lysis solution (Sigma) for 3 min before being pelleted and resuspended in 2 mL of medium. The cell suspensions were counted with Trypan blue to determine viability and adjusted to 10×10^6 cells per milliliter. One-hundred microliters (1×10^6 cells) was transferred into individual wells of a 96-well plate and supplemented with 100 µL of (1) medium for unactivated control, (2) 1 µg/mL LPS (Sigma) for B-cell activation, or (3) 1 µg/mL αCD3 (BioLegend) for T-cell activation. After 24–48 h, the cells were processed for RNA-FISH, immunofluorescence, and protein isolation.

Whole-genome RNA sequencing and bioinformatics

In brief, cell lines (detailed in the [Supplemental Material](#)) were grown to 80% density before RNA extraction and DNase I treatment. Libraries, optimized for 250- to 400-base-pair inserts, were prepared using NEBNext Ultra (Illumina), enriched for mRNA using NEBNext poly(A) mRNA magnetic isolation module, and sequenced to generate $\sim 5 \times 10^7$ reads per sample. STAR software was used to align reads to the C57BL/6 X chromosome. Transcriptome assembly and expression quantification were performed using Cufflinks and Cuffdiff. Of 85 differentially expressed X-linked transcription units ($P < 0.05$), 23 were excluded due to differential expression between biological replicates. Heat maps and gene enrichment analysis were carried out as described in the [Supplemental Material](#).

Histology

Following dissection, tissues were transferred immediately into histological grade formalin and processed after 24–48 h. Immunostaining for CD antigens was performed using rabbit anti-CD3 for T cells (Abcam, ab16669) at 1:200 and goat anti-CD20 for B cells (Santa Cruz Biotechnology, sc7735) at 1:500.

ESC differentiation

Female PGK12.1 ESCs were grown in ES medium with LIF on gelatin without feeders. To induce differentiation, 1×10^6 cells were plated onto nongelatinized dishes without LIF. On day 3, differentiating colonies were replated onto bacterial dishes to stimulate embryoid body formation. On day 7, embryoid bodies were transferred to nongelatinized dishes to reattach. Fibroblast outgrowths were passaged as required.

NM extraction

Cells were serially extracted with (1) detergent to reveal soluble factors, (2) salt to reveal loosely bound chromatin-associated factors, (3) DNase I to reveal tightly attached chromatin-associated factors, and (4) RNase to reveal RNA-associated factors, as described (Wilson et al. 2016), with improvements detailed in the Supplemental Material. Coverslips were then fixed and processed for immunofluorescence.

Immunofluorescence

Cells were washed in PBS and either (1) fixed in 4% PFA to reveal total protein or (2) treated with detergent prior to fixing to reveal chromatin- and NM-associated factors prior to incubation with primary antibody for 2 h, and then secondary antibody for 1 h. For the ESC differentiation course, PGK12.1 cells were fixed in 2% formaldehyde for 15 min prior to permeabilization. The antibodies used were α -H3K27me3 mAb (Abcam, ab6002; Active Motif, ab61017), α -CIZ1 N-terminal (1794), α -CIZ1 C-terminal (Novus, NB100-74624), SAFA anti-HNRNP-U (Abcam, ab10297), and anti-YY1 (SC7341). Coverslips were costained with limiting concentrations of Hoechst 33258 (10 ng/mL; Sigma) for quantitative detection of chromatin.

RNA-FISH

Female cultured cells were processed for the detection of *Xist* transcript (red) by RNA-FISH followed by immuno-FISH for CIZ1 (green) using N-terminal antibody 1794. An 11-kb *Spe*I–*Sal*I mouse *Xist* fragment was fluorescently tagged with Cy3-dUTP (GE Healthcare) using BioPrime labeling kit (Invitrogen). Samples were incubated with probe overnight at 37°C. For subsequent detection of CIZ1, antibody 1794 was applied for 1 h followed by secondary anti-rabbit FITC (Sigma) for 1 h. Cells were imaged and processed using Adobe Photoshop CS4 to enhance signal definition. Prior RNA-FISH processing resulted in reduced CIZ1 signal intensity throughout the nucleus. For SR 3D-SIM and RNA-FISH on splenocyte *Xist*, cDNA was labeled with Spectrum green-dUTP or Spectrum red-dUTP by nick translation (Abbott Molecular). Following fixation and permeabilization, cells were incubated with primary antibody for 1 h and then with Alexa fluor goat anti-rabbit 594 for 30 min and then washed and post-fixed before detection of *Xist* overnight. After extensive washing, the cells were incubated with 2 μ g/mL DAPI and mounted with VectorShield.

Chromosome paints

FITC-conjugated X-chromosome paint (AMP 0XG) was used as instructed (Cytocell Ltd.). Labeled cells were mounted in VectorShield with DAPI and imaged.

Microscopy

Images were collected using a Zeiss Axiovert 200M and AxioCam and Openlab image acquisition software and quantified using ImageJ (National Institutes of Health) using raw images acquired under identical conditions. Images for publication were enhanced using Adobe Photoshop or Affinity Photo 1.4 by applying identical manipulations to test and control samples so that image intensities reflect actual relationships. Live images were collected on a PE Ultraview spinning-disk confocal microscope. SR 3D-SIM was performed on a DeltaVision OMX V3 Blaze system (GE Healthcare) equipped with a 60 \times /1.42 NA plan apo oil immersion objective (Olympus), sCMOS cameras (PCO), and 405-,

488-, and 593-nm lasers. 3D-SIM image stacks were acquired as described in the Supplemental Material.

Acknowledgments

X-chromosome paints were a kind gift from Cytocell. We are grateful to James Hewitson, Dimitris Lagos, Mike Shires, and Matthew Wiseman for advice or assistance, and Sally James, Richard Randle-Boggis, Katherine Newling, and Peter Ashton of York Technology Facility Genomics laboratory. This work was supported by a Radhika Sreedhar Scholarship to R.R.-F., University of York priming funds, Biotechnology and Biological Sciences Research Council PhD training scholarships to E.R.S. and R.W., Genetics Society training funds to R.R., Wellcome Trust grants to N.B. (081385,091911), and the Micron Advance Imaging Initiative (Wellcome Trust 103768).

References

- Ainscough JF, Rahman FA, Sercombe H, Sedo A, Gerlach B, Coverley D. 2007. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. *J Cell Sci* **120**: 115–124.
- Bardelli M, Leucci E, Schurfeld K, Bellan C, Passiatore G, Rocchigiani M, Bartolommei S, Orlandini M, Zagursky J, Lazzi S, et al. 2007. VEGF-D is expressed in activated lymphoid cells and in tumors of hematopoietic and lymphoid tissues. *Leuk Lymphoma* **48**: 2014–2021.
- Bianchi I, Lleo A, Gershwin ME, Invernizzi P. 2012. The X chromosome and immune associated genes. *J Autoimmun* **38**: J187–J192.
- Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**: 515–526.
- Brown CJ, Willard HF. 1994. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* **368**: 154–156.
- Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**: 527–542.
- Caparros ML, Alexiou M, Webster Z, Brockdorff N. 2002. Functional analysis of the highly conserved exon IV of *XIST* RNA. *Cytogenet Genome Res* **99**: 99–105.
- Charfi C, Edouard E, Rassart E. 2014. Identification of GPM6A and GPM6B as potential new human lymphoid leukemia-associated oncogenes. *Cell Oncol (Dordt)* **37**: 179–191.
- Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M. 2016. *Xist* recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**: 468–472.
- Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. 2015. Systematic discovery of *Xist* RNA binding proteins. *Cell* **161**: 404–416.
- Clemson CM, McNeil JA, Willard HF, Lawrence JB. 1996. *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132**: 259–275.
- Copeland NA, Sercombe HE, Ainscough JF, Coverley D. 2010. Ciz1 cooperates with cyclin-A–CDK2 to activate mammalian DNA replication in vitro. *J Cell Sci* **123**: 1108–1115.

- Copeland NA, Sercombe HE, Wilson RH, Coverley D. 2015. Cyclin-A-CDK2-mediated phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. *J Cell Sci* **128**: 1518–1527.
- Coverley D, Marr J, Ainscough J. 2005. Ciz1 promotes mammalian DNA replication. *J Cell Sci* **118**: 101–112.
- Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R. 1999. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet* **22**: 323–324.
- da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, et al. 2014. Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol Cell* **53**: 301–316.
- den Hollander P, Rayala SK, Coverley D, Kumar R. 2006. Ciz1, a novel DNA-binding coactivator of the estrogen receptor α , confers hypersensitivity to estrogen action. *Cancer Res* **66**: 11021–11030.
- Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM, Brockdorff N. 1999. Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in *cis*. *Hum Mol Genet* **8**: 195–204.
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, et al. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**: 1237973.
- Greaves EA, Copeland NA, Coverley D, Ainscough JF. 2012. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. *J Cell Sci* **125**: 2466–2477.
- Guo B, Odgren PR, van Wijnen AJ, Last TJ, Nickerson J, Penman S, Lian JB, Stein JL, Stein GS. 1995. The nuclear matrix protein NMP-1 is the transcription factor YY1. *Proc Natl Acad Sci* **92**: 10526–10530.
- Hall LL, Byron M, Pageau G, Lawrence JB. 2009. AURKB-mediated effects on chromatin regulate binding versus release of XIST RNA to the inactive chromosome. *J Cell Biol* **186**: 491–507.
- Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S. 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**: 469–476.
- Heard E, Distechi CM. 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**: 1848–1867.
- Helbig R, Fackelmayer FO. 2003. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma* **112**: 173–182.
- Higgins G, Roper KM, Watson IJ, Blackhall FH, Rom WN, Pass HI, Ainscough JF, Coverley D. 2012. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc Natl Acad Sci* **109**: E3128–E3135.
- Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**: 119–133.
- Kolpa HJ, Fackelmayer FO, Lawrence JB. 2016. SAF-A requirement in anchoring XIST RNA to chromatin varies in transformed and primary cells. *Dev Cell* **39**: 9–10.
- Lee JT, Bartolomei MS. 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**: 1308–1323.
- Lee JT, Strauss WM, Dausman JA, Jaenisch R. 1996. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* **86**: 83–94.
- Leong HS, Chen K, Hu Y, Lee S, Corbin J, Pakusch M, Murphy JM, Majewski IJ, Smyth GK, Alexander WS, et al. 2013. Epigenetic regulator Smc4d1 functions as a tumor suppressor. *Cancer Res* **73**: 1591–1599.
- Makhoulouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C. 2014. A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat Commun* **5**: 4878.
- McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, et al. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**: 232–236.
- Mitsui K, Matsumoto A, Ohtsuka S, Ohtsubo M, Yoshimura A. 1999. Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. *Biochem Biophys Res Commun* **264**: 457–464.
- Moindrot B, Cerase A, Coker H, Masui O, Grijzenhout A, Pintacuda G, Schermelleh L, Nesterova TB, Brockdorff N. 2015. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep* **12**: 562–572.
- Monfort A, Di Minin G, Postlmayr A, Freimann R, Arieti F, Thore S, Wutz A. 2015. Identification of *Spen* as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep* **12**: 554–561.
- Nesterova TB, Slobodyanyuk SY, Elisaphenko EA, Shevchenko AI, Johnston C, Pavlova ME, Rogozin IB, Kolesnikov NN, Brockdorff N, Zakian SM. 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res* **11**: 833–849.
- Nishibe R, Watanabe W, Ueda T, Yamasaki N, Koller R, Wolff L, Honda Z, Ohtsubo M, Honda H. 2013. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. *FEBS Lett* **587**: 1529–1535.
- Pazgal I, Boycov O, Shpilberg O, Okon E, Bairey O. 2007. Expression of VEGF-C, VEGF-D and their receptor VEGFR-3 in diffuse large B-cell lymphomas. *Leuk Lymphoma* **48**: 2213–2220.
- Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for Xist in X chromosome inactivation. *Nature* **379**: 131–137.
- Pollex T, Heard E. 2012. Recent advances in X-chromosome inactivation research. *Curr Opin Cell Biol* **24**: 825–832.
- Rahman FA, Aziz N, Coverley D. 2010. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. *BMC Cancer* **10**: 482.
- Sakaguchi T, Hasegawa Y, Brockdorff N, Tsutsui K, Tsutsui KM, Sado T, Nakagawa S. 2016. Control of chromosomal localization of Xist by hnRNP U family molecules. *Dev Cell* **39**: 11–12.
- Savarese F, Flahndorfer K, Jaenisch R, Busslinger M, Wutz A. 2006. Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. *Mol Cell Biol* **26**: 7167–7177.
- Sheardown SA, Duthie SM, Johnston CM, Newall AE, Formstone EJ, Arkell RM, Nesterova TB, Alghisi GC, Rastan S, Brockdorff N. 1997. Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell* **91**: 99–107.
- Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**: 465–469.
- Smeets D, Markaki Y, Schmid VJ, Kraus F, Tattermusch A, Cerase A, Sterr M, Fiedler S, Demmerle J, Popken J, et al.

2014. Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin* **7**: 8.
- Spatz A, Borg C, Feunteun J. 2004. X-chromosome genetics and human cancer. *Nat Rev Cancer* **4**: 617–629.
- Wang J, Syrett CM, Kramer MC, Basu A, Atchison ML, Anguera MC. 2016. Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for increased expression from the inactive X. *Proc Natl Acad Sci* **113**: E2029–E2038.
- Ward JM. 2006. Lymphomas and leukemias in mice. *Exp Toxicol Pathol* **57**: 377–381.
- Wilson RH, Coverley D. 2013. Relationship between DNA replication and the nuclear matrix. *Genes Cells* **18**: 17–31.
- Wilson RH, Hesketh EL, Coverley D. 2016. Preparation of the nuclear matrix for parallel microscopy and biochemical analyses. *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot083758.
- Wutz A, Jaenisch R. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* **5**: 695–705.
- Wutz A, Rasmussen TP, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**: 167–174.
- Xiao J, Vemula SR, Xue Y, Khan MM, Kuruvilla KP, Marquez-Lona EM, Cobb MR, LeDoux MS. 2016. Motor phenotypes and molecular networks associated with germline deficiency of Ciz1. *Exp Neurol* **283**: 110–120.
- Yamada N, Hasegawa Y, Yue M, Hamada T, Nakagawa S, Ogawa Y. 2015. Xist exon 7 contributes to the stable localization of Xist RNA on the inactive X-chromosome. *PLoS Genet* **11**: e1005430.
- Yildirim E, Kirby JE, Brown DE, Mercier FE, Sadreyev RI, Scadden DT, Lee JT. 2013. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* **152**: 727–742.

Appendix B

PDF of Swarts DRA, **Stewart ER**, Higgins GS, Coverley D. CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours. *Cell Cycle*. 2018;17(18):2268-83.

RESEARCH PAPER



CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours

Dorian R.A. Swarts , Emma R. Stewart , Gillian S. Higgins, and Dawn Coverley 

Department of Biology, The University of York, York, UK

ABSTRACT

CIZ1 promotes cyclin-dependent DNA replication and resides in sub-nuclear foci that are part of the protein nuclear matrix (NM), and in RNA assemblies that are enriched at the inactive X chromosome (Xi) in female cells. It is subjected to alternative splicing, with specific variants implicated in adult and pediatric cancers. CIZ1-F is characterized by a frame shift that results from splicing exons 8–12 leading to inclusion of a short alternative reading frame (ARF), excluding the previously characterized C-terminal NM anchor domain. Here, we apply a set of novel variant-selective molecular tools targeted to the ARF to profile the expression of CIZ1-F at both transcript and protein levels, with focus on its relationship with the RNA-dependent and -independent fractions of the NM. Unlike full-length CIZ1, CIZ1-F does not accumulate at Xi, though like full-length CIZ1 it does resist extraction with DNase. Notably, CIZ1-F is sensitive to RNase identifying it as part of the RNA-fraction of the NM. In quiescent cells CIZ1-F transcript expression is suppressed and CIZ1-F protein is excluded from the nucleus, with re-expression not observed until the second cell cycle after exit from quiescence. Importantly, CIZ1-F is over-expressed in common solid tumors including colon and breast, pronounced in early stage but not highly-proliferative late stage tumors. Moreover, expression was significantly higher in hormone receptor negative breast tumors than receptor positive tumors. Together these data show that CIZ1-F is expressed in proliferating cells in an unusual cell cycle-dependent manner, and suggest that it may have potential as a tumor biomarker.

ARTICLE HISTORY

Received 20 June 2018
Revised 29 August 2018
Accepted 9 September 2018

KEYWORDS

CDKN1A-interacting zinc finger protein 1; nuclear matrix; alternative splicing

Introduction

CDKN1A interacting zinc finger protein 1 (CIZ1) is subject to extensive alternative splicing to yield multiple transcript variants [1–5]. Functional analysis of discrete protein fragments has identified domains and regulatory sites in the N-terminal half of the protein that are involved in DNA replication [6,7] and interaction with cell cycle regulators [7], (Figure 1(a,b)). Separate sequences in the C-terminus of CIZ1 support interaction with nuclear structures, identifying a salt and DNase-resistant nuclear matrix (NM) interaction domain [8], which may restrict its function to specific sub-nuclear locations. Recently, we and others showed that CIZ1 localizes to the inactive X chromosome (Xi) and is required to ensure retention of Xist long non-coding RNA (lncRNA) at Xi in fibroblasts [9,10]. Alternative splicing affects both the

replication domain (RD) and NM anchor domain (AD) of CIZ1, and several splicing events have been implicated in cancer, including mutation driven exclusion of exon 4 (CIZ1-Δ4) in Ewing's Tumor cell lines [2] and exclusion of 24 nucleotides of exon 14 in lung cancer (CIZ1-B) [5]. A unique splice-junction epitope from CIZ1-B is a circulating biomarker for patients with small cell and non-small cell lung cancer [5,11]. CIZ1 has also been reported to be involved in the development of breast cancer [12], the most common type of cancer in women [13]. Alternative splicing of CIZ1 is implicated in other chronic diseases including upregulation of a form in which part of exon 8 is excluded (CIZ1-S) in Alzheimer's disease [3], and in cervical dystonia where mutations in an exonic splicing enhancer may affect splicing patterns [14].

CONTACT Dawn Coverley  dawn.coverley@york.ac.uk

 Supplemental data for this article can be accessed [here](#).

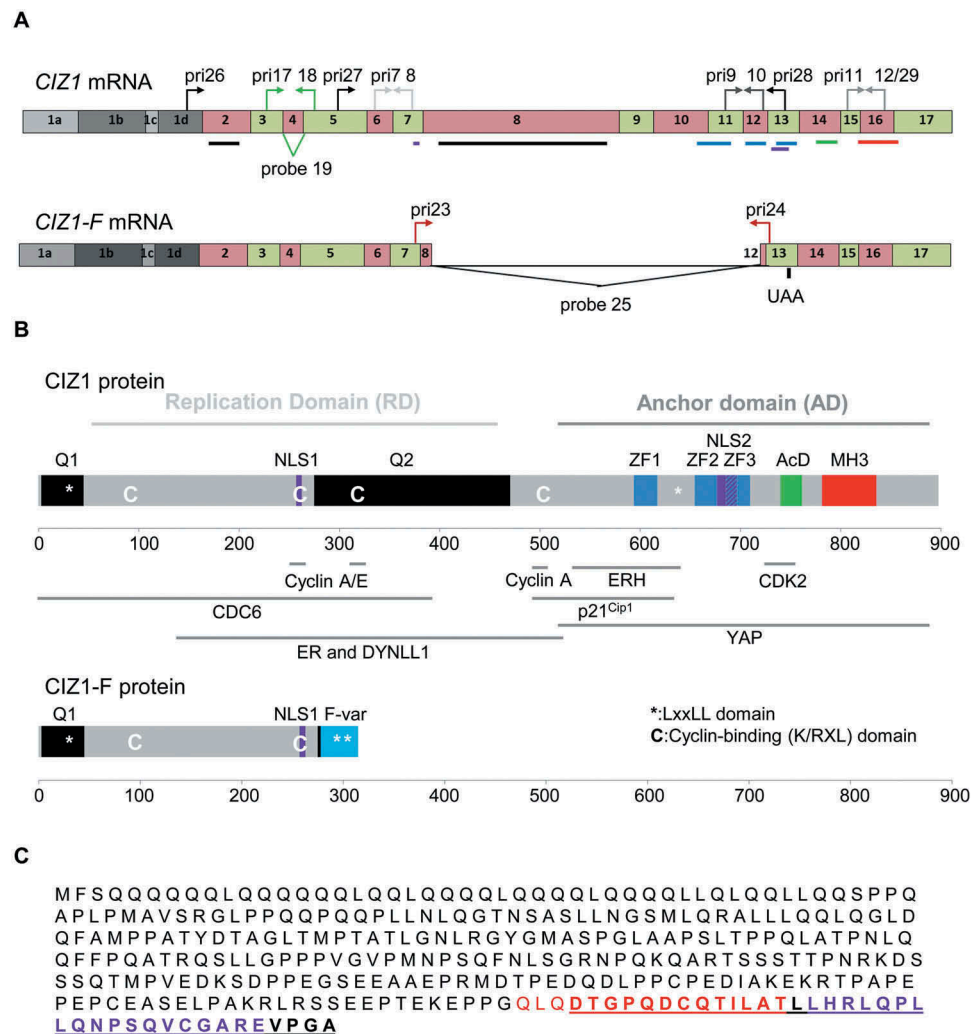


Figure 1. Human CIZ1 and variant CIZ1-F.

(A) Exonic sequence in messenger RNA of full-length and *CIZ1-F*, with location of primers (pri) and Taqman probes indicated. Sequences are provided in Supplementary Table 2. Lines indicate the location of sequences that encode the protein domains illustrated in (B). Four alternative exon 1s are indicated; *CIZ1-F* has been detected in combination with exons 1b, 1c, and 1d. (B) Full-length and *CIZ1-F* protein showing annotated protein domains including glutamine-rich regions (Q), nuclear localization signals (NLS), zinc finger domains (ZF), matrin-3 domain (MH3), and acidic domain (AcD). Characterized functional domains and interaction sites for full-length CIZ1 are also shown, including replication domain (RD) which contains all sequences required for replication activity [6], and anchor domain (AD) which contains sequences that are sufficient to mediate attachment to the nuclear matrix [8]. The *CIZ1-F* specific sequence, encoded by an alternative reading frame (ARF) of exon 12–13, is highlighted in light blue. Also shown are the location of K/RXL cyclin-binding motifs implicated in replication [7] ('C'), LXXLL ER-interaction motifs (asterisks), and sequences reported for full-length CIZ1 to interact with CDC6 [18], cyclin-dependent kinase 2 (CDK2) [15], cyclins A and E [7], dynein light chain (DYNLL1) [15], estrogen-receptor (ER) [12], enhancer of rudimentary homolog (ERH) [42], p21 [16], and YAP [17]. Note that several of these interactions were discovered in mouse CIZ1, and that the reported interaction between ER and CIZ1 is not via a domain that includes LXXLL [12]. X-axis shows amino acid number. (C) Amino acid sequence of the *CIZ1-F* protein, with the sequence specified by its ARF bold and underlined. The peptides in red and purple were used to generate *CIZ1-F*-specific antibodies.

Despite functional studies that link full-length murine CIZ1 and embryonic variant murine CIZ1 (ECIZ1) with DNA replication [6,7], little is known about the function of other variants. Alternative splicing of *CIZ1* may have important consequences for the cell and is known to affect

the sub-nuclear localization of the protein (Supplementary Table 1). For example, exclusion of exon 4 changes the sub-nuclear distribution of CIZ1 from focal to diffuse [2]. Thus alternative splicing of *CIZ1* may influence the spatial organization of DNA replication, and suggests that some

CIZ1 splice variants have the potency to act as dominant-negative controllers of other variants [2].

We previously designed an exon-junction microarray that identified a number of novel variants of CIZ1, including one common cancer-associated *CIZ1*-variant characterized by a continuous deletion of part of exon 8, exon 9–11 and part of exon 12, here referred to as CIZ1-F (Figure 1(a)) [1]. CIZ1-F was initially identified in Ewing's Tumor cell lines and found to be overexpressed in primary lung tumors but not matched normal tissues [1]. Here, we characterize CIZ1-F mRNA and protein expression patterns by exploiting a short unique peptide encoded by expression of an alternative reading frame (ARF) of exon 12–13. We show that CIZ1-F is specifically expressed in G1 phase following mitosis but not G1 phase following quiescence. Despite lacking the C-terminal NM-attachment domain [8] and several protein interaction sites that are important for replication function, CIZ1-F resists extraction with DNase1. It is however sensitive to RNase, which contrasts with the behavior of full-length CIZ1. Most importantly, *CIZ1*-F is overexpressed in early stage primary human tumors and is associated with ER-negative status in breast tumors.

Results

Human *CIZ1*-F is characterized by an 1181 base pair deletion (c1038_2218del1181 in reference sequence NM_012127.2), which excludes part of exons 8 and 12 and all of exons 9–11 (Figure 1(a), Supplementary Figure 1(a)). The exon 8–12 junction causes a frame-shift in the *CIZ1* reading frame, leading to a premature translational stop codon in exon 13, despite the presence of downstream exons in *CIZ1*-F mRNA (Figure 1(a)). *CIZ1*-F protein therefore lacks previously characterized functional domains, including the MH3-domain which anchors CIZ1 to the NM [8], and cyclin-binding motifs through which CIZ1 interacts with cyclin A, and which are essential for its DNA replication activity (Figure 1(b)) [6,7]. Other characterized sites of interaction with CDK2, p21^{cip1} and YAP are also excluded [15–17], while interaction with CDC6 [18], ER α [12], and dynein light chain [15] might be retained. The interaction

sites of the CIZ1 binding factors PDRG1 [19], SH3BP4 [20], and TCF4 [21] are not yet characterized. Importantly, the ARF created by the 8–12 junction encodes 37 amino acids of unique sequence not present in other forms of CIZ1, or in other proteins, and results in inclusion of two additional LXXLL estrogen receptor (ER)-interaction motifs (Figure 1(b,c)) [22]. Thus *CIZ1*-F transcript appears to encode a 315 amino acid protein with a unique C-terminal end. To study *CIZ1*-F mRNA expression we designed a *CIZ1*-F junction-specific gene expression probe (Supplementary Table 2, Supplementary Figure 1(b-c)). Initially, *CIZ1*-F mRNA levels were measured by qRT-PCR in a set of cell lines (Supplementary Figure 1(d)), revealing variable levels, with high expression in bladder and cervical cancer lines EJ and HeLa, as well as in breast cancer cell lines MCF-7, BT474 and MDA-MB-231. Evaluation of cultured tumor-derived cell lines showed that *CIZ1*-F transcript is elevated in MCF7 breast cancer cells compared to normal MCF-10A breast cells (Supplementary Figure 1(d)), and is in fact one of the main *CIZ1* PCR-products (Supplementary Figure 1(a)), identifying MCF7 cells as a model cell line to study the function of CIZ1-F in cell proliferation. However, despite the reported functional interaction between CIZ1 and ER α [12], we were unable to detect a reproducible effect of estrogen on the expression of *CIZ1* or *CIZ1*-F transcript in MCF7 cells (Supplementary Figure 1(e)).

***CIZ1*-F protein is part of the RNA-dependent NM**

In order to confirm that *CIZ1*-F transcript encodes a protein we generated an affinity-purified anti-peptide antibody against sequences in the unique C-terminal ARF of predicted CIZ1-F protein (Figure 1(c)). Specificity of the antibody was confirmed using ectopically expressed EGFP-full-length CIZ1 and EGFP-CIZ1-F for western blot and immunofluorescence applications (Supplementary Figure 2(a-b)). Further confirmation of the specificity of our CIZ1-F antibody was obtained using peptide blocking, which yielded two CIZ1-F-specific products of approximately 42kD and 71kD in total cell lysates (Supplementary Figure 2(c)). The 42kD monomeric CIZ1-F product is larger than the predicted molecular weight of CIZ1-F (34kD)

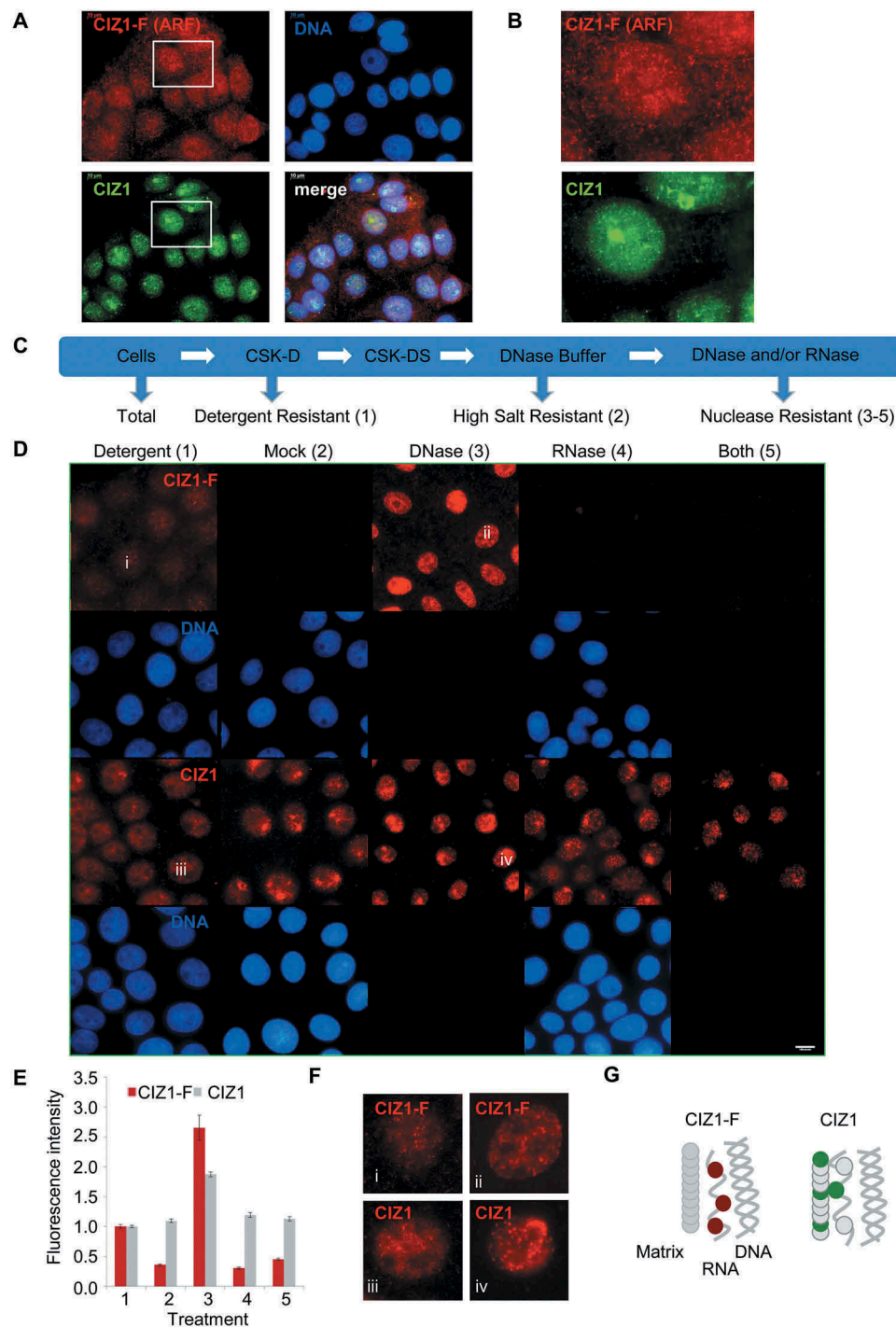


Figure 2. Nuclear CIZ1-F resists extraction of chromatin but not RNA.

(A-B) CIZ1-F does not localize to the inactive X chromosome (Xi). Immunofluorescence with a monoclonal antibody for a peptide located in the CIZ1 anchor domain (AD; green) shows both diffuse nuclear expression and specific localization to a region previously shown to be the Xi [9]. Immunofluorescence with purified antibody raised to the CIZ1-F ARF (red) shows absence of Xi-localization. DNA is stained with Hoechst33258 (blue). Nuclei in white boxes are enlarged in (B). (C) Overview of nuclear matrix (NM) extraction procedure. MCF-7 cells were serially extracted with 1) detergent-containing cytoskeletal buffer (CSK), 2) detergent-containing CSK supplemented with 0.5 M NaCl (mock extraction), 3) DNase 1, or 4) RNase, or 5) both enzymes (see methods). (D) MCF-7 cells subjected to the treatments described under (C) were counterstained with Hoechst33258 to control for removal of chromatin (blue), and probed with purified polyclonal anti-CIZ1-F or replication domain (CIZ1-RD) antibody 1794 (red). CIZ1-RD resists all treatments, while CIZ1-F is sensitive to extraction of RNA, and is also dramatically revealed by removal of chromatin. Bar is 10 microns. (E) Quantification of the fluorescence intensities in (D), shown after subtraction of background signal and expressed relative to detergent treated cells. Number of cells quantified ≥ 70 per condition. Numbers refer to the treatment conditions listed above. (F) Enhanced and enlarged images of the indicated nuclei i-iv from (D), showing detergent and DNase-resistant nuclear fractions. (G) Interpretation of the data showing the dependency of CIZ1-F on RNA for nuclear retention, compared to full resistance of CIZ1-RD.

but in agreement with the mobility of EGFP-CIZ1-F after subtraction of EGFP (45kD; Supplementary Figure 2(a)).

Since full-length CIZ1 localizes to the Xi and interacts with Xist lncRNA, we first determined the relationship between CIZ1-F and the Xi in MCF-7 cells. Full-length CIZ1, detected by a monoclonal antibody to a C-terminal epitope, clearly localizes to the Xi in MCF-7 cells, though in line with reports that Xi in breast cancer-derived cells is a less discrete (compact) entity than in 'normal' cells [23] Xi staining is markedly irregular, and also accompanied by the nucleus wide signal that is typical of CIZ1 [9]. In contrast, CIZ1-F shows a different pattern including both nuclear and cytoplasmic foci, but no enrichment at Xi (Figure 2(a-b)). We next examined the relationship of full-length CIZ1 and CIZ1-F with the NM. Full-length CIZ1 is part of the core protein NM (resistant to extraction with DNase and RNase), as evident by immunodetection of endogenous CIZ1 with RD-specific antibodies (Figure 2(c-f)). Analysis of attachment to this core protein NM can be recapitulated using EGFP-full-length CIZ1 and derived fragments, allowing the domain responsible for attachment to the NM to be located at the C-terminal end, and classified as 'anchor domain' or AD [8]. AD is missing from CIZ1-F, which would therefore not be expected to resist extraction. Importantly, under identical conditions to those used to profile endogenous CIZ1, the extraction profile for CIZ1-F indicates attachment to the RNA component of the NM (RNA-protein NM) because it resists extraction with DNase, but not when used in conjunction with RNase (Figure 2(c-f)). Notably, direct comparison of immunofluorescence signal in nuclei at different stages in the extraction process shows that the majority of the CIZ1-F epitope in the nucleus is in fact masked by chromatin (Figure 2(e)). The signal seen before extraction appears to be not the same population as that revealed when chromatin is removed, because pre-extraction signal is lost in mock treated nuclei and only revealed if DNase is also included in the reaction (Figure 2(e)). Again, despite the reported association between ER α and CIZ1 [12], we did not observe co-localization of CIZ1-F with nuclear ER α in

immuno-detection experiments (Supplementary Figure 2(d)).

In conclusion, CIZ1-F resists extraction of chromatin and is therefore a NM protein, however it is not part of the core protein NM like full-length CIZ1 as it can be fully extracted by digestion of nuclear RNA, in the presence or absence of co-extraction of DNA, identifying it as part of the RNA-protein NM (Figure 2(g)). The differential extraction profiles of full-length CIZ1 and CIZ1-F epitopes suggest that they may have distinct functions within the NM.

CIZ1-F expression in cycling and quiescent cells

We next focused on CIZ1-F expression in cycling and quiescent cell populations, since CIZ1-F retains most regions involved in DNA-replication (Figure 1(b)). In detergent-treated cell populations, endogenous CIZ1-F protein expression patterns changed on entry to quiescence, switching from nuclear localization in cycling cells, to nuclear exclusion with cytoplasmic signal at confluence, to nearly undetectable in quiescence (Figure 3(a-b)). Importantly, when quiescent cell populations were subjected to full NM-extraction, all signal was lost when cells were treated with DNase, demonstrating that quiescent cell populations do not retain any CIZ1-F on the NM (Figure 3(c)).

CIZ1-F transcript levels were also assessed in cycling, contact-inhibited and serum-deprived cells, and compared to amplicons in CIZ1-RD and CIZ1-AD (Figure 3(d-e)). Overall, CIZ1-F transcript is approximately 50-fold lower than 'total' CIZ1 mRNA in cycling cells, which is broadly consistent with the frequency of expressed sequence tags in NCBI UniGene (accessed 03/03/2018), where 11 of 875 sequences (all from cancers) are CIZ1-F. CIZ1-RD and CIZ1-AD levels remain relatively stable during proliferative growth, and slightly increase during growth rate decline caused by serum starvation. However, CIZ1-F mRNA expression fell dramatically as proliferation rate dropped (Figure 3(d); Supplementary Figure 3(a)), and recovered again upon addition of fresh serum. In contrast, contact-inhibited normal fetal lung fibroblast MRC5 cells did not resume proliferation or restore CIZ1-F

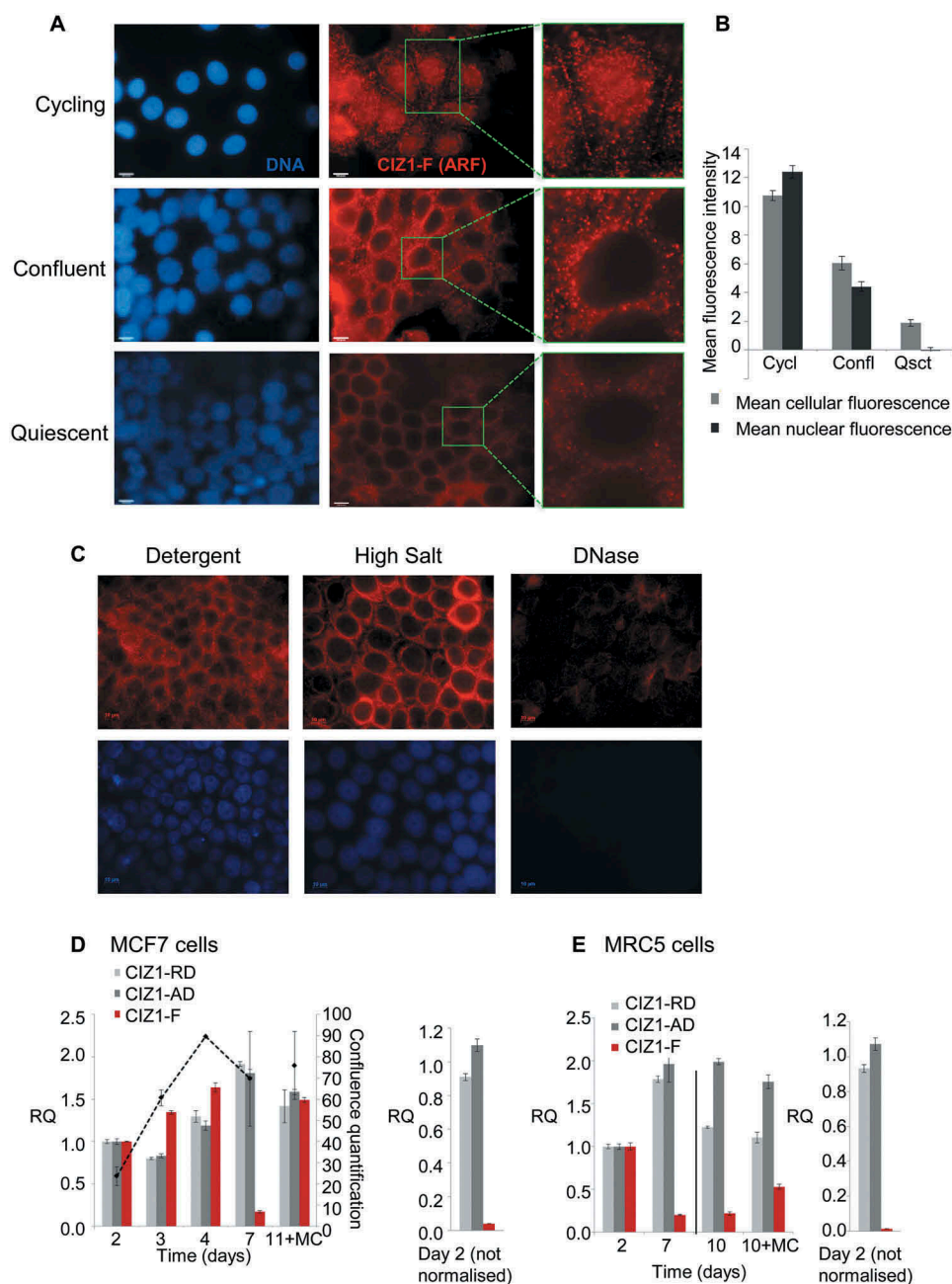


Figure 3. CIZ1-F transcript is suppressed and CIZ1-F protein is excluded from the nuclear matrix in quiescent cells.

(A) Immunodetection of CIZ1-F in cycling (day 2), confluent (day 4) and quiescent (day 7 with medium change) populations of MCF-7 cells, after washing with detergent-cytoskeletal buffer to remove soluble protein. Enlarged images of example cells are shown and DAPI-stained images (left) show cell nuclei in the same fields. Bar is 10 microns. (B) Quantification of the average immunofluorescence signal of cells depicted in (A) after subtraction of background. Abbreviations: Cycl, cycling; Confl, confluent; Qsct, quiescent. (C) NM-extraction of quiescent MCF-7 cell populations showing no CIZ1-F on the NM after treatment with DNase. For details on the method see Figure 2(a). (D) Transcript levels in MCF-7 breast epithelial carcinoma cells at the indicated number of days post-plating, measured by quantitative RT-PCR (qPCR). A parallel culture that received regular changes of media (indicated as 'MC'), harvested at 11 d, is shown for comparison. Histograms show the mean of three technical replicates \pm SEM of a representative experiment (experiments were repeated at least 3 times). Data is expressed as relative quantification (RQ) after normalization to *ACTB* and *CYPA*, and is calibrated to levels at day 2. Cell counts at the time of harvesting are provided \pm SEM (black dotted line; average of three counts expressed as proportion of a confluent population). Right, comparison between replication domain (*CIZ1*-RD), anchor domain (*CIZ1*-AD) and *CIZ1*-F in MCF-7 cells at day 2 \pm SEM, with no normalization to show relative levels. (E) As in (D) for MRC5 normal fetal lung fibroblast cells. Right, comparison between *CIZ1*-RD, *CIZ1*-AD and *CIZ1*-F at day 2. SEM for three technical replicates is shown. Primer and probe sequences are in Supplementary Table 2.

expression when boosted with serum (Figure 3(e)). This suggests that stable contact-induced growth arrest suppresses expression of *CIZ1-F*. In the same MCF-7 cell populations we observed additional effects on *CIZ1* alternative splicing, most notably increased *CIZ1-S* (Supplementary Figure 3(a)) an isoform reported to be upregulated in Alzheimer's disease (Supplementary Table 1), and changes in *CIZ1-Δ4* (Supplementary Figure 3(b)) [2,3]. Together the data show that *CIZ1-F* mRNA and protein are expressed in proliferating cells and inhibited by both serum-deprivation-induced proliferation arrest (MCF-7), and contact-inhibition-induced proliferation arrest (MRC5).

***CIZ1-F* mRNA is expressed in G1 following mitosis (M-G1) but not G1 following quiescence (Q-G1)**

To understand better the role of *CIZ1-F* in proliferation we analyzed its cell cycle expression profile. In order to generate synchronized populations, serum-deprived MCF-7 cells were released back into cycle in the presence of fresh serum and the nucleotide analog EdU (Figure 4(a)). The frequency of EdU-positive cells indicates that the population passes through S-phase between 10 and 20 h after exiting quiescence, peaking at 16 h (49% of nuclei; Figure 4(a)), and this is preceded as expected, by an increase in expression of cyclin E mRNA. However, *CIZ1-F* transcript does not begin to increase until after the majority of cells have passed through S-phase, recovering only at 31 h (Figure 4(a)). In contrast, *CIZ1* AD and RD expression remains relatively stable. When MCF-7 cells were exposed to the DNA synthesis inhibitors aphidicolin or thymidine, or arrested in mitosis by nocodazole upon release from quiescence, *CIZ1-F* expression was strongly suppressed compared to uninhibited cells (Figure 4(b)). Similar results were obtained with MRC5 cell populations treated with thymidine or nocodazole (Figure 4(c)). Thus *CIZ1-F* mRNA levels do not accumulate in either S-phase or mitosis and *CIZ1-F* is not expressed in Q-G1. However, when MCF7 cells were released from arrest with nocodazole in the M-phase following quiescence, *CIZ1-F* mRNA was expressed within 2 h, which is at least 4 h before cyclin E rises (Figure 4(d)). Thus, *CIZ1-F* mRNA is expressed in early G1 phase in the second cell cycle

following release from quiescence, or M-G1 (Figure 4(e)).

***CIZ1-F* mRNA expression in primary tumors and cancer cell lines**

To confirm that *CIZ1-F* is overrepresented in cancer tissue, and to extend our analysis of lung tumors [1], we analyzed *CIZ1* mRNA in colon and breast cancer. Almost without exception colon cancer samples expressed elevated *CIZ1-F* compared to matched normal control tissue ($P < 0.001$, Related-Samples Wilcoxon Signed Rank Test; Figure 5(a); see Supplementary Figure 4 for individual sample data), and surprisingly the highest levels were in early stage tumors ($P = 0.017$ for early versus late stage, Mann-Whitney U-test). In contrast *CIZ1-RD* does not increase as dramatically in early stages, nor diminish significantly at later stages ($P = 0.44$ for the same comparison, Mann-Whitney U-test; Figure 5(a)). Thus the data indicate upregulation of *CIZ1-F* specifically in early stage colon cancer. Similar results were obtained with breast tumors with higher *CIZ1-F* expression in tumor tissue compared to normal breast tissues ($P = 0.016$, Mann-Whitney U-test), whereas expression of *CIZ1-RD* was not significantly different (Figure 5(b); see Supplementary Figure 5 for individual data) [1]. Again *CIZ1-F* levels were most elevated in early stage tumors ($P = 0.006$ when compared to normal tissues, Mann-Whitney U-test). Thus the profile of *CIZ1-F* expression across tumor stage is similar for both breast cancer and colon cancer, and further preliminary analysis of lung and urinary bladder cancers returned similar trends (Supplementary Figure 6), suggesting that the stage profile may be common to other tumor types.

When grouped by tumor grade rather than stage, the greatest elevation was at lowest grade for colon cancer (Figure 5(c)), but for breast cancer *CIZ1-F* levels increased with grade ($P = 0.029$, Kruskal-Wallis Test; Figure 5(d)). This is not surprising however since, in this series, most low-grade breast cancers were high-stage, and vice versa (see Figure 5(d)).

To evaluate the relationship with proliferation, the transcript level of the proliferation marker *MKI67* was measured in the same breast and colon samples (Figure 5). For colon, although

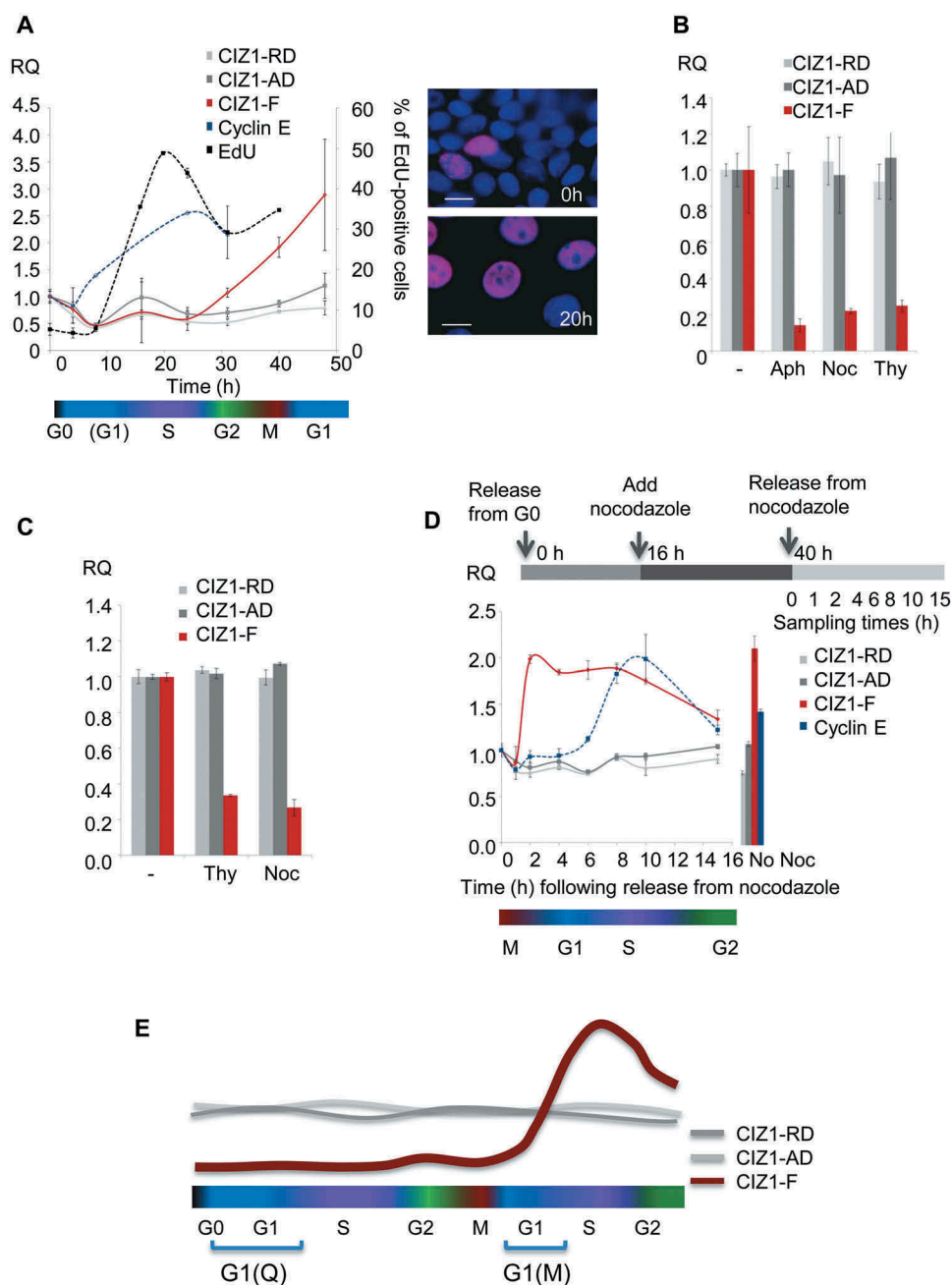


Figure 4. *CIZ1-F* transcript is expressed in G1-phase.

(a) Quantitative RT-PCR showing *CIZ1-F* transcript (red) in MCF-7 cells during the first cell cycle following release from contact-inhibition and serum-deprivation induced cell cycle arrest, relative to *ACTB* and *CYPA* housekeeping genes, and calibrated to unreleased cells. *CIZ1* anchor domain (*CIZ1-AD*) and replication domain (*CIZ1-RD*) expression (gray lines), as well as cyclin E1 (blue) are shown for comparison. Data show means of four experiments \pm SD. The percentage of cells that incorporated EdU into newly synthesized DNA during a 30 min pulse at the indicated times is shown for comparison (black line; two biological repeats \pm SD). Note that error bars for cyclin E1 are very small (range, 0.00001–0.008). Cell cycle stages (estimates based on EdU-incorporation and cyclin-expression) are indicated below. Images on the right show EdU (purple) and total DNA (blue) before (0 h) and after (20 h) release. Bar is 10 microns. (b) *CIZ1-AD*, *CIZ1-RD* and *CIZ1-F* transcript in MCF-7 cells 31 h post-release from quiescence, in the presence and absence of aphidicolin (Aph), nocodazole (Noc) and thymidine (Thy), showing mean of three biological replicates \pm SEM. (c) As in (b), but for MRC5 normal lung fetal fibroblast cells. Mean RQ of three technical replicates is expressed relative to cycling samples (\pm SEM). (d) As in (a), but showing release of MCF-7 cells from a 24 h nocodazole-arrest applied 16 h after release from quiescence. Synchronization strategy is indicated above the graph. Graph shows transcript levels (mean of three technical replicates \pm SEM), at the indicated hours after release from arrest, with cell cycle stage (based on cyclin E expression) illustrated below. Right, control sample without nocodazole (40 h post-release from quiescence). (e) Schematic summarizing *CIZ1-RD*, *CIZ1-AD* and *CIZ1-F* mRNA expression levels during the first cell cycle following release from cell cycle arrest (Q-G1) and mitotic arrest (M-G1). Primer and probe sequences are in Supplementary Table 2.

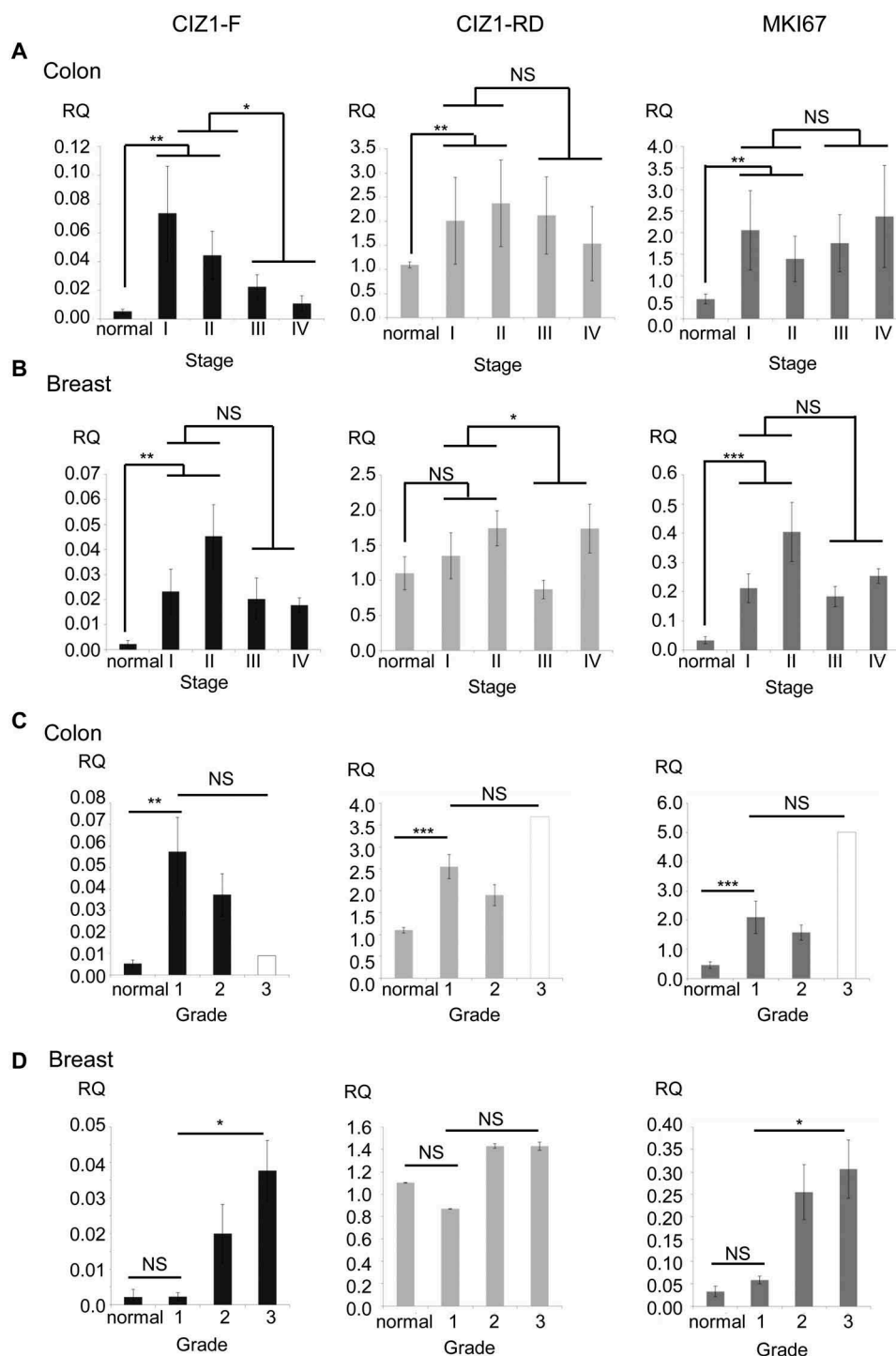


Figure 5. *CIZ1-F* is overexpressed in early stage human breast and colon cancer.

(a) *CIZ1-F* and replication domain (*CIZ1-RD*) expression in 24 primary colon tumors and matched normal samples (Origene colon cancer cDNA array HCRT103), showing mean mRNA expression levels for the average of all matched normal tissues and the four cancer stage classifications. *CIZ1-RD* and *MKI67* are shown for comparison. Individual data per case can be found in Figure 4, and individual classifications and pathology notes accesses at www.origene.com. (b) As in (a), showing 5 normal samples and 43 primary breast cancer samples from the indicated stages (Origene breast cancer cDNA array BCRT102). Data for individual cases can be found in Figure 5. (c) As in (a) for the same 24 colon cancer samples and matched normal samples, showing mean mRNA expression grouped by grade. Grade 1: 50% late stage; grade 2: 38% late stage; grade 3: 100% late stage (1 case, white bar). Late stage refers to stage III and stage IV. (d) As in (b) for the same 5 normal samples and 43 primary breast cancer samples, showing mean mRNA expression grouped by grade. Grade 1: 100% late stage tumors; grade 2: 43% late stage tumors; grade 3: 27% late stage tumors. Late stage refers to stage III and stage IV. Mean RQ's to the mean *CIZ1-RD*-expression of normal samples \pm SEM are shown. Significant differences are indicated (NS, not significant). Primer and probe sequences are in Supplementary Table 2.

MKI67 and *CIZ1-F* levels correlated closely when the matched normal tissues were included in the analysis ($P = 0.0020$; Pearson Correlation), no correlation was found between levels across the cancer samples only, whereas *CIZ1-RD* was strongly correlated to *MKI67* levels by both analyzes ($P \leq 0.00043$; Pearson Correlation). Although less obvious for breast tumors, again no significant correlation across the full set for *CIZ1-F* and *MKI67* could be identified ($P = 0.058$; Pearson Correlation), compared to correlation with RD ($P \leq 0.0050$; Pearson Correlation). When performing linear regression with both *CIZ1-RD* and *CIZ1-F* in colon or breast samples, only *CIZ1-RD* significantly relates to *MKI67* expression in all analyzes ($P \leq 0.0050$). Thus, unlike *CIZ1-RD*, *CIZ1-F* does not mirror expression of *MKI67* and therefore does not seem to have a simple relationship with cell proliferation.

Interestingly, *CIZ1-F* levels correlated with hormone receptor status in breast tumors. ER and PR negative tumors have significantly higher *CIZ1-F* transcript levels than ER or progesterone receptor (PR) positive tumors ($P = 0.005$, Mann-Whitney U-test; Figure 6(a-b)), while HER2 status was not correlated (Figure 6(c)). This is not the case for *CIZ1-RD*, arguing for a relationship between *CIZ1-F* and ER α and PR in primary tumors.

Discussion

Here we demonstrate that *CIZ1-F* is part of the RNA-dependent nuclear-matrix and elevated in early-stage cancers including hormone receptor negative, but not positive breast tumors. We show that *CIZ1-F* is distinct from other forms of *CIZ1* and may have potential as a biomarker of early-stage disease.

CIZ1-F is part of the RNA-dependent NM and distinct from full-length CIZ1

CIZ1 has multiple functions, both within and outside the context of DNA replication and the cell cycle, and its complexity is increased by the existence of multiple splice variants (Supplementary Table 1) [1–7,18]. We and others have shown that *CIZ1* is a NM protein, defined by its resistance to extraction of chromatin [3,8], which suggests that its function is connected to nuclear architecture,

possibly contributing to the spatial regulation of nuclear processes [6–8,18,24]. Although the NM appears to be corrupted in certain types of cancer [25] previous work [26–28] and our own experiments indicate that MCF-7 cells retain a NM structure to some extent. In MCF-7 cells both *CIZ1-RD* and *CIZ1-F*, as well as ER α , resist co-extraction with chromatin, indicating the presence of a non-chromatin stabilizing structure with which they are associated. This is of interest because *CIZ1-F* lacks the previously characterized NM anchor domain encoded by the C-terminal end of *CIZ1*. Moreover *CIZ1-F* is actually revealed by removal of chromatin, suggesting that the ARF-encoded epitope is normally in close proximity to chromatin, and masked by it. Further distinguishing *CIZ1-F* from full-length *CIZ1* [8] is its behavior in quiescent cells. Here, *CIZ1-F* is excluded from the nucleus and not evident even after removal of chromatin, suggesting that its role is exclusively in cycling cells. Finally, though *CIZ1-F* is co-extracted with the RNA fraction of the NM (Figure 2(g)), it does not normally reside at the Xi implying that, unlike full-length *CIZ1* this is not mediated by interaction with Xist lncRNA [9,10]. Future research should now focus on assigning a function to *CIZ1-F*. Since the unique junction sequence is a very limited target, we were not able to achieve specific *CIZ1-F* knockdown using an siRNA approach (data not shown).

Biomarker potential for early stage cancers

Previous reports have shown overexpression of *CIZ1* in colon [29,30], gallbladder [21], lung [31], and prostate cancer [32], as well as ependymomas, gliomas, and medulloblastomas [33], and it has also been reported that *CIZ1* knockout mice develop leukaemias and lymphomas [9,34]. Moreover, splice variants of *CIZ1* appear to have lineage-specific functions [4] some of which are over-represented in human tumors; notably *CIZ1-Δ4* in Ewings tumor and *CIZ1-B* in lung tumors [2,5]. However, for most published analyzes it is not clear which variants are reported on, and in some cases transcript detection tools that are reported to reflect *CIZ1* levels may actually reflect shifts in variant expression. Here we use validated and specific tools to demonstrate that *CIZ1-F* is

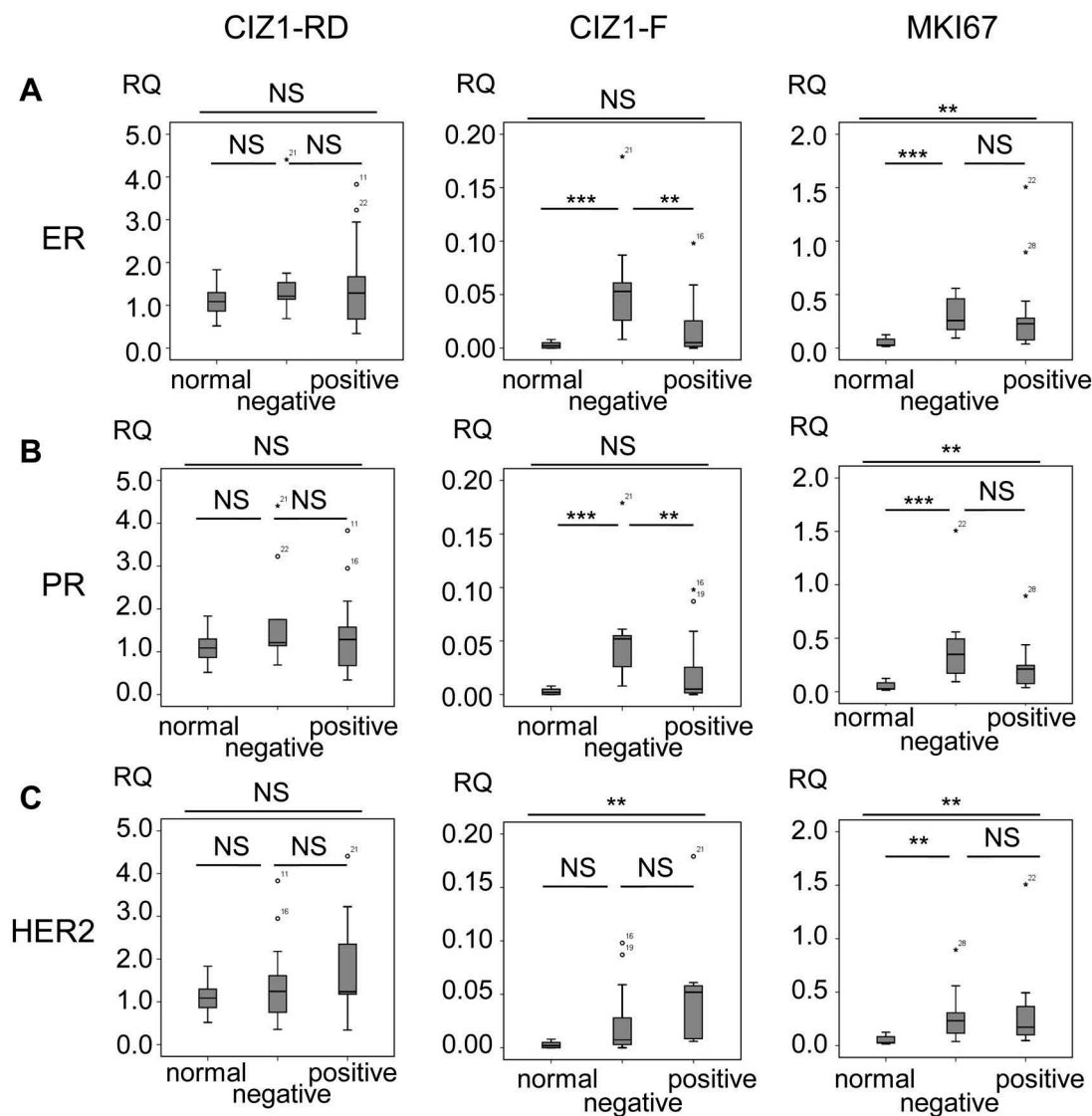


Figure 6. *CIZ1*-F expression is increased in ER-negative tumors.

Box-plots showing *CIZ1* replication domain (*CIZ1*-RD; left), *CIZ1*-F (middle) and *MKI67* (right) expression in (a) normal samples and estrogen receptor (ER)-positive and -negative tumors, (b) normal samples and progesterone receptor (PR)-positive and -negative tumors, and (c) normal samples and HER-2-positive and -negative tumors. RQ's are expressed relative to mean *CIZ1*-RD-expression of normal samples. Significant differences between subgroups are indicated (Mann Whitney U-tests; NS, not significant). Primer and probe sequences are in Supplementary Table 2.

upregulated in tumors of the colon and breast, that upregulation is most pronounced in early stage (I and II) disease, and is not directly correlated with proliferation. We also note that the association of *CIZ1*-F with tumor grade varies between tumor types, so although disrupted nuclear architecture and altered NM is a common feature of poorly differentiated and aggressive cancers [25,35], we cannot at this stage draw a correlation with suppression of *CIZ1*-F. In addition, our analyzes have currently been limited to *CIZ1*-F mRNA levels. Future studies should determine *CIZ1*-F protein expression patterns in primary tumor samples.

Perhaps most useful is the clear correlation with hormone receptor status of breast tumors. Previous analysis has identified *CIZ1* as an estrogen-responsive gene with estrogen-response elements in its promoter [12]. Moreover, the N-terminal domain of *CIZ1* protein can also interact with ER, conferring hypersensitivity to estrogen in animal models and enhancing the tumorigenicity of breast cancer cells [12]. Though alternative splicing was not addressed in this study, a contribution of *CIZ1*-F to the cells response to estrogen is likely because the ER interaction domain is partially retained in *CIZ1*-F, and

it has two additional LXXLL nuclear receptor binding motifs [22] encoded by its unique C-terminal ARF. Contrary to published results [12] we did not observe induction of *CIZ1* upon stimulation with estrogen for 24 h. A possible explanation for the apparent discrepancy is that different *CIZ1* primers were used previously [12], which may well have reported on alternative splicing of *CIZ1* exon 8 rather than overall levels. In our study we detected *CIZ1*-RD and *CIZ1*-AD amplicons in regions unrelated to exon 8 in order to avoid this highly variable region of *CIZ1*. *CIZ1* binding to (PR) has not been previously studied; though it has been shown that, in the context of NCoA-1/SRC-1, PR also requires two LXXLL domains [36]. Thus, while we could find no functional response to estrogen in MCF-7 cells, the presence of these domains and the correlation with breast tumor receptor status supports a relationship *in vivo*.

Different *CIZ1* splice variants at different cell proliferation stages

Transcript expression profile also distinguishes *CIZ1*-F from other variants of *CIZ1*. *CIZ1*-F has a distinct cell cycle-specific profile with low levels in G₀, and delayed upregulation beyond the first cell cycle, which is at variance with the relatively constant expression of *CIZ1*-RD and *CIZ1*-AD amplicons (Figure 4(e)). In these studies we also noted proliferation state dependent changes in the *CIZ1*-S and *CIZ1*-Δ4 variants. In fact in MCF-7 cells, they seem to be inversely related to *CIZ1*-F, with *CIZ1*-S and *CIZ1*-Δ4 being more highly expressed in arrested cells (Supplementary Figure 3). It is worth noting that both *CIZ1*-F and *CIZ1*-S variants are the product of exon 8 splicing events, and that exon 8 splicing patterns have been implicated in Alzheimer's disease and cervical dystonia [3,14].

In conclusion, this exploratory study has convincingly distinguished *CIZ1*-F from *CIZ1*, revealed association with the RNA-dependent NM, and highlighted an unusual cell cycle expression pattern. It also supports a relationship with estrogen responsive cancers and suggests that *CIZ1*-F may have potential as a biomarker.

Materials & methods

Cell culture and tumor samples

MCF-7 breast cancer cells were cultured in minimal essential medium (cat. 21090-022; Gibco, Life Technologies, Thermo Fisher Scientific [TFS] Inc) in the presence of 10% fetal calf serum (FCS; cat. FB-5815 Biosera, Boussens, France) and 1X penicillin-streptomycin-glutamine (PSG; cat. 10378-016; Gibco, Life Technologies, TFS Inc). MRC5 normal lung fetal fibroblast cells were grown in Dulbecco's Modified Eagle Medium (cat. 21885-025; Gibco, Life Technologies, TFS Inc) in the presence of 10% FCS and 1X PSG. MCF-10A normal breast cells were cultured as previously described [37]. The cell lines were tested and authenticated in August 2017 using PCR-single-locus-technology (Eurofins Medigenomix Forensik GmbH). TissueScanTM Cancer Survey cDNA array (CSRT103; Origene) was used to assess *CIZ1*-F levels in different human cancer types. The breast cancer cDNA array BCRT102 (Origene) was used to screen 43 breast cancer samples and 5 normal breast tissues. The colon cancer cDNA array HCRT103 (Origene) was used to screen 24 colon cancer samples with matched normal tissues.

Quantitative real-time PCR (qRT-PCR)

RNA was isolated using Trizol[®] Reagent (cat. 15596-026; Ambion RNA, Life Technologies, TFS Inc) and reverse transcribed into cDNA using Superscript[®] III First-Strand Synthesis System (cat. 18080-051; Life Technologies, TFS Inc). Quantitative RT-PCR was performed using SYBR green (FastSYBR[®] Green Mastermix, cat. 4385612; Applied Biosystems[®] [AB], TFS Inc) or Taqman (Taqman[®] Fast Universal PCR Master Mix, cat. 4352042; AB, TFS Inc) reagents using the AB StepOnePlusTM Real-Time PCR Systems (TFS Inc). The following program was used: 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. Data was analyzed using the StepOne Software v. 2.3 (AB, TFS Inc). Origene[®] cancer cDNA array plates were processed with the AB 7300 system (TFS Inc) using the TaqMan[®] Universal Master Mix II (cat. 4440040; AB, TFS Inc) using the following program: 50°C for 2 min,

95°C for 10 min, followed by 50 cycles of 95°C for 15 s and 60°C for 1 min.

Primers were designed using Primer 3 Plus (<http://primer3plus.com/>), except for the *ACTB* primers [5] and *CYPA* primers [38], which were previously described, and cyclin D1 and cyclin E1 primers, which were retrieved from qPrimerDepot (<http://primerdepot.nci.nih.gov>). All primers are listed in Supplementary Table 2. *CIZ1*-RD was detected with primers in exon 6/7 and *CIZ1*-AD with primers in exon 15/16 (Figure 1(a), Supplementary Table 2). For the Origene cDNA array plates, *CIZ1*-F was detected with primers 23–25, *CIZ1*-RD with primers 20–22 and *MKI67* using a Taqman Assay (assay no. Hs01032443_m1). All expression levels have been normalized to the mean expression of *ACTB* and *CYPA* unless stated differently.

Constructs and transfection

Human *CIZ1*-F was cloned from cDNA from MCF-7 cells using PCR with DreamTaq (cat. EP1701; TFS Inc) and T/A cloning into the pGEM[®]-T Easy Vector System I (cat. A1360; Promega) and verified by sequencing. *CIZ1*-F was subsequently transferred using restriction enzymes (PmlI and SanDI, cat. ER0361 and FD2164, respectively; TFS Inc) into the pEGFP-C3 vector, allowing direct comparison with full-length *CIZ1*, containing all translated exons (2–17), also cloned into pEGFP-C3 as part of a previous study [5]. Both plasmids were transiently transfected into MCF-7 cells using the *TransIT*[®]-3T3 and *TransIT*[®]-LT1 Transfection Kits (cat. MIR 2180 and MIR 2300, respectively; Mirus Bio LLC).

Antibodies

CIZ1-F specific antibody was made by immunizing rabbits with two peptides (Figure 1(c)) from the unique *CIZ1*-F ARF-encoded peptide (Covalab UK Ltd). Antisera from two rabbits were pooled and immunopurified against peptide 2 (which had no significant similarity to any other protein) and used for Western Blot and immunofluorescence (dilutions 1:100–1:500 and 1:50, respectively). *CIZ1*-F blocking peptides were used in a 1:500 dilution. Mouse monoclonal antibody (hC221a)

against *CIZ1* was raised by Fusion Antibodies against a recombinant human *CIZ1* C-terminal fragment encoded by amino acids 678–898 in reference sequence UniProt Q9ULV3. In addition, purified *CIZ1* 1794 polyclonal antiserum [6] was used to detect *CIZ1* without discriminating *CIZ1*-F. For immunofluorescence goat polyclonal anti-mouse-AlexaFluor488 conjugate or goat polyclonal anti-rabbit-AlexaFluor568 conjugate (cat. A11001 and A11011, respectively, Life Technologies, TFS) were used as appropriate. For Western Blot *CIZ1* exon 5 antibody (cat. HPA-020380; Sigma-Aldrich) was used to detect *CIZ1* without discriminating *CIZ1*-F. Actin (cat. A3853, Sigma Aldrich) was used as a loading control.

Immunodetection

Protein extracts were obtained by washing cells with cold PBS and harvesting directly into SDS sample buffer (100 mM DTT, 2% SDS, 60 mM Tris pH 6.8, 0.001% bromophenol blue) at 95°C in the presence of 2 mM PMSF. Western blot was performed as previously described [6]. Bands were visualized using EZ-ECL (cat. 20-500-120; Biological Industries Ltd).

For immunofluorescence, cells grown on glass coverslips were transferred into PBS, washed with cytoskeletal (CSK) buffer (containing EDTA-free protease inhibitor tablets [cat. 05056489001; F Hoffmann-La Roche Ltd, Sigma Aldrich] and 1 mM DTT) with or without 0.1% Triton X-100 [39] for 30 s, then transferred to PBS and fixed for 20 min in 4% paraformaldehyde in PBS. Primary antibodies were incubated for 1.5 h at 37°C and secondary antibodies for 1 h at 37°C in the dark in antibody buffer (PBS, 10 mg/ml BSA, 0.2% SDS, 0.1% Triton X-100). Nuclei were counterstained with DAPI (VECTASHIELD Antifade Mounting Medium with DAPI, cat. H-1200; Vector Laboratories).

Fluorescent image data was collected with a Zeiss Axiovert microscope (Carl Zeiss) with a 63/1.40 oil immersion objective and an AxioCam camera (Carl Zeiss Vision) with Openlab software (version 4.0.2, Improvion). Images were processed using Adobe PhotoShop to reduce background to black and to overlay multicolor images. Where quantitative data is derived from the images, all capture parameters and image

processing was identical for samples and controls within an experiment.

Nuclear matrix extractions

Cells were fractionated as previously described [39,40]. In brief, coverslips were washed with detergent (0.1% Triton-X-100 in CSK), followed by the same buffer with 0.5 M NaCl. Cells were then incubated at 37°C for 1 h in DNase Buffer with or without DNase I (0.5 U/μl, cat. 000000004716728001; F Hoffmann-La Roche Ltd, Sigma Aldrich), RNase (cat. 000000011579681001; F Hoffmann-La Roche Ltd, Sigma Aldrich) or both nucleases in combination, then fixed for 20 min with 8% paraformaldehyde. Vanadyl Ribonucleoside Complex (cat. S1402S; New England Biolabs) was added at 1:80 to all buffers with the exception of the RNase containing treatment step and its preceding wash steps. Coverslips were counterstained with Hoechst33258 (1:100,000, cat. H3569; TFS) to verify removal of chromatin, and immunostained as indicated.

Cell cycle synchronization

MCF-7 cells were synchronized by release from quiescence as described [41]. Cells were grown to confluence, then the medium was changed, and cells cultured for a further four days without further media changes and then released in serum-containing medium by 1:4 splitting. Alternatively, cells were incubated in the presence of nocodazole (0.04 μg/ml), aphidicolin (5 μg/ml) or thymidine (2.5 mM) for 24 h, and released from arrest by washing twice in PBS followed by addition of fresh medium. In order to verify synchrony, incorporation of EdU into newly synthesized DNA was measured using Click-iT® EdU Alexa Fluor® 555 (cat. C10338; Invitrogen, Life Technologies, TFS Inc).

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows (version 21.0.0.0). Associations between expression levels and clinical parameters were calculated using the Mann Whitney U-test or the Kruskal-Wallis test, and

for related samples using the Related-Samples Wilcoxon Signed Rank Test, as indicated. Relationships between expression levels amongst different genes were determined using Pearson Correlation and linear regression. All statistical tests were two-sided and all *P*-values were considered significant when < 0.05 . Significance levels were indicated as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Abbreviations:

AB	Applied Biosystems
AD	anchor domain
ARF	alternative reading frame
CIZ1	CDKN1A-interacting zinc finger protein
CSK	cytoskeletal
ERα	estrogen receptor alpha
FCS	fetal calf serum
NM	nuclear matrix
PR	progesterone receptor
PSG	penicillin-streptomycin-glutamine
RD	replication domain
RQ	relative quantity
SD	standard deviation
SEM	standard error of mean
qRT-PCR	quantitative real-time PCR
TFS	Thermo Fisher Scientific
Xi	inactive X chromosome
Xist lncRNA	X-inactive specific transcript long non-coding RNA

Acknowledgments

We thank Dr Justin Ainscough for critical comments on the manuscript, Hannah Screeton for her contribution to cell studies and quantitative PCRs, and Dr William Brackenbury for providing cell line RNA.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was funded by a project grant from Yorkshire Cancer Research (YCR grant no Y258 to DC) and Biotechnology and Biological Sciences Research Council doctoral training funds to ES (White Rose DTP grant no BB/M011151/1).

Author's contribution

DS and ES conducted experiments. GH provided technical assistance. DS and DC wrote the manuscript.

ORCID

Dorian R.A. Swarts  <http://orcid.org/0000-0002-7927-7613>

Emma R. Stewart  <http://orcid.org/0000-0003-0189-9348>

Dawn Coverley  <http://orcid.org/0000-0001-8262-7023>

References

- [1] Rahman FA, Aziz N, Coverley D. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. *BMC Cancer*. 2010;10:482.
- [2] Rahman F, Ainscough JF, Copeland N, et al. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. *Hum Mutat*. 2007;28:993–1004.
- [3] Dahmcke CM, Buchmann-Moller S, Jensen NA, et al. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol Cell Neurosci*. 2008;38:589–594.
- [4] Greaves EA, Copeland NA, Coverley D, et al. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. *J Cell Sci*. 2012;125:2466–2477.
- [5] Higgins G, Roper KM, Watson IJ, et al. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc Natl Acad Sci USA*. 2012;109:E3128–35.
- [6] Coverley D, Marr J, Ainscough J. Ciz1 promotes mammalian DNA replication. *J Cell Sci*. 2005;118:101–112.
- [7] Copeland NA, Sercombe HE, Ainscough JF, et al. Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. *J Cell Sci*. 2010;123:1108–1115.
- [8] Ainscough JF, Rahman FA, Sercombe H, et al. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. *J Cell Sci*. 2007;120:115–124.
- [9] Ridings-Figueroa R, Stewart ER, Nesterova TB, et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev*. 2017;31:876–888.
- [10] Sunwoo H, Colognori D, Froberg JE, et al. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc Natl Acad Sci USA*. 2017;114:10654–10659.
- [11] Coverley D, Higgins G, West D, et al. A quantitative immunoassay for lung cancer biomarker CIZ1b in patient plasma. *Clin Biochem*. 2017;50:336–343.
- [12] Den Hollander P, Rayala SK, Coverley D, et al. Ciz1, a Novel DNA-binding coactivator of the estrogen receptor alpha, confers hypersensitivity to estrogen action. *Cancer Res*. 2006;66:11021–11029.
- [13] Siegel R, Ma J, Zou Z, et al. Cancer statistics. *CA Cancer J Clin*. 2014;64:9–29.
- [14] Xiao J, Uitti RJ, Zhao Y, et al. Mutations in CIZ1 cause adult onset primary cervical dystonia. *Ann Neurol*. 2012;71:458–469.
- [15] Den Hollander P, Kumar R. Dynein light chain 1 contributes to cell cycle progression by increasing cyclin-dependent kinase 2 activity in estrogen-stimulated cells. *Cancer Res*. 2006;66:5941–5949.
- [16] Mitsui K, Matsumoto A, Ohtsuka S, et al. Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. *Biochem Biophys Res Commun*. 1999;264:457–464.
- [17] Lei L, Wu J, Gu D, et al. CIZ1 interacts with YAP and activates its transcriptional activity in hepatocellular carcinoma cells. *Tumour Biol*. 2016;37:11073–11079.
- [18] Copeland NA, Sercombe HE, Wilson RH, et al. Cyclin A/CDK2 phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. *J Cell Sci*. 2015;128:1518–1527.
- [19] Jiang L, Luo X, Shi J, et al. PDRG1, a novel tumor marker for multiple malignancies that is selectively regulated by genotoxic stress. *Cancer Biol Ther*. 2011;11:567–573.
- [20] Thalappilly S, Suliman M, Gayet O, et al. Identification of multi-SH3 domain-containing protein interactome in pancreatic cancer: a yeast two-hybrid approach. *Proteomics*. 2008;8:3071–3081.
- [21] Zhang D, Wang Y, Dai Y, et al. CIZ1 promoted the growth and migration of gallbladder cancer cells. *Tumour Biol*. 2015;36:2583–2591.
- [22] Heery DM, Kalkhoven E, Hoare S, et al. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*. 1997;387:733–736.
- [23] Chaligne R, Popova T, Mendoza-Parra MA, et al. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Res*. 2015;25:488–503.
- [24] Wilson RH, Coverley D. Relationship between DNA replication and the nuclear matrix. *Genes Cells*. 2013;18:17–31.
- [25] Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. *Nat Rev Cancer*. 2004;4:677–687.
- [26] Spencer VA, Samuel SK, Davie JR. Nuclear matrix proteins associated with DNA in situ in hormone-dependent and hormone-independent human breast cancer cell lines. *Cancer Res*. 2000;60:288–292.
- [27] Samuel SK, Minish TM, Davie JR. Nuclear matrix proteins in well and poorly differentiated human breast cancer cell lines. *J Cell Biochem*. 1997;66:9–15.
- [28] Khanuja PS, Lehr JE, Soule HD, et al. Nuclear matrix proteins in normal and breast cancer cells. *Cancer Res*. 1993;53:3394–3398.
- [29] Wang DQ, Wang K, Yan DW, et al. Ciz1 is a novel predictor of survival in human colon cancer. *Exp Biol Med*. 2014;239:862–870.
- [30] Yin J, Wang C, Tang X, et al. CIZ1 regulates the proliferation, cycle distribution and colony formation

- of RKO human colorectal cancer cells. *Mol Med Rep.* **2013**;8:1630–1634.
- [31] Zhou X, Liu Q, Wada Y, et al. CDKN1A-interacting zinc finger protein 1 is a novel biomarker for lung squamous cell carcinoma. *Oncol Lett.* **2018**;15:183–188.
- [32] Liu T, Ren X, Li L, et al. Ciz1 promotes tumorigenicity of prostate carcinoma cells. *Front Biosci (Landmark Ed).* **2015**;20:705–715.
- [33] Warder DE, Keherly MJ. Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0-2)YYAC. *J Biomed Sci.* **2003**;10:406–417.
- [34] Nishibe R, Watanabe W, Ueda T, et al. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. *FEBS Lett.* **2013**;587:1529–1535.
- [35] Verheijen R, van Venrooij W, Ramaekers F. The nuclear matrix: structure and composition. *J Cell Sci.* **1988**;90(Pt 1):11–36.
- [36] McInerney EM, Rose DW, Flynn SE, et al. Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev.* **1998**;12:3357–3368.
- [37] Tait L, Soule HD, Russo J. Ultrastructural and immunocytochemical characterization of an immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **1990**;50:6087–6094.
- [38] Swarts DR, Henfling ME, Van Neste L, et al. CD44 and OTP are strong prognostic markers for pulmonary carcinoids. *Clin Cancer Res.* **2013**;19:2197–2207.
- [39] Wilson RHH, Hesketh EL, Coverley D. The nuclear matrix: fractionation techniques and analysis. In: Pryor PP, editor. *Subcellular fractionation - a laboratory manual.* New York: Cold Spring Harbor Laboratory Press; **2015.** p. 223–233.
- [40] Hesketh EL, Knight JR, Wilson RH, et al. Transient association of MCM complex proteins with the nuclear matrix during initiation of mammalian DNA replication. *Cell Cycle.* **2015**;14:333–341.
- [41] Coverley D, Laman H, Laskey RA. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. *Nat Cell Biol.* **2002**;4:523–528.
- [42] Lukasik A, Uniewicz KA, Kulis M, et al. Ciz1, a p21 cip1/Waf1-interacting zinc finger protein and DNA replication factor, is a novel molecular partner for human enhancer of rudimentary homolog. *FEBS J.* **2008**;275:332–340.

Appendix C

PDF of **Stewart ER**, Coverley D. Visualization of Hidden Epitopes at the Inactive X Chromosome. *Methods Mol Biol.* 2018;1861:103-12.

Visualization of hidden epitopes at the inactive X chromosome

Authors: Emma R Stewart and Dawn Coverley

Affiliation: Department of Biology, University of York, Wentworth Way, YO10 5DD

Email: dawn.coverley@york.ac.uk

Running Head: Unmasking proteins at the Xi

Abstract

Immuno-detection of nuclear antigens is often complicated by epitope masking, so that proteins known to function in the nucleus are sometimes not easily detected at their sites of action. Moreover, protein populations that are detected before unmasking can be very different to those seen after removal of nucleic acids. This is particularly true for components of the nuclear matrix, including those known to function at the inactive X chromosome. Here we describe an unmasking protocol that reveals previously undetected proteins at the inactive X chromosome in mouse fibroblasts.

Keywords: Unmasking nuclear proteins, Nuclear matrix, Immunofluorescence, SAF-A, CIZ1, Inactive X chromosome

1.0 Introduction

Enrichment of proteins that are known to be involved in X chromosome inactivation (XCI) is sometimes only seen in the location of the inactive X chromosome (Xi) when using fluorescently tagged exogenously expressed forms. This is true for hnRNP U (SAF-A), a nuclear matrix (NM) protein that binds to both RNA and DNA that is required for XCI and mitotic spindle organisation [1-3]. SAF-A is not apparent within the Xi territory in fixed whole cells, but is revealed if cells are treated to remove chromatin and associated proteins before fixation [4]. An overview of our extraction procedure is shown in Figure 1A, and is described in detail in this chapter. Similar procedures can be used to address the class of interaction partner responsible for supporting accumulation of epitopes at Xi or elsewhere in the nucleus.

Example 1: Revealing masked epitopes at Xi

Immuno-staining of unextracted, rapidly cycling female mouse embryonic fibroblasts (MEFs) with anti-SAF-A antibodies returns a range of staining patterns, always

nuclear but with highly variable levels. In the experiment shown here, 22% of MEFs have a very bright SAF-A signal, but this is lost if cells receive a brief wash with detergent-containing, isotonic buffer prior to fixation (Figure 1B). This shows that most of the SAF-A in the cell is not bound to insoluble components of the nucleus. Moreover, neither the pre or post detergent SAF-A signal indicates enrichment at Xi, despite several reports of accumulation of exogenously expressed GFP tagged SAF-A at Xi [5,6] and a close functional relationship between SAF-A and Xist RNA [2,7-10]. In fact, line intensity scans show that SAF-A is apparently excluded from the Xi territory (Figure 1C,D). We obtained similar results for other NM proteins that are reported to function in XCI, including PTBP1 [11] and YY1 [12,13]. However, when nuclei are further processed by digestion with DNaseI to remove chromatin under controlled conditions, immunofluorescence signal is revealed at Xi (Figure 1C-E). This leads to detection of signal where there was absence from the Xi (SAF-A), and in some cases a signal peak (PTBP1).

Example 2: Nuclear matrix interactions

The nuclear matrix (NM) can be biochemically defined as the material that is retained in the nucleus after harsh extraction with detergent, high salt and nucleases [14]. It includes an RNA component that is retained after controlled extraction of chromatin, and a core protein fraction that is retained when both chromatin and RNA are removed. Differential extraction can be used to identify how XCI-related proteins are anchored in the nucleus. To illustrate this an extraction series of female mouse fibroblast cells is shown after immuno-staining for the CIZ1 protein (Fig 1F), which was recently shown to be involved in *Xist* localisation [4]. While nucleus-wide CIZ1 speckles are resistant to extraction under all conditions (they remain in fraction C6) and are therefore part of the core protein nuclear NM, the CIZ1 population at Xi is

highly sensitive to RNase (lost in fraction C5). This identifies CIZ1 at Xi as being associated with the RNA component of the NM, consistent with its functional interaction with Xist RNA. Like CIZ1, SAF-A is also highly sensitive to RNase (Figure 1E). CIZ1 in particular illustrates the different types of interaction with the NM, and also highlights the need for rigorous preservation of RNA during chromatin removal protocols, by use of RNase-free buffers and RNase inhibitors.

Example 3: Nucleus-wide superficial and buried populations

To control for unwanted degradation of RNA and to illustrate the effects of the high-salt wash steps that are required to fully remove digested nucleic acid fragments, we routinely generate mock-digested control samples. For CIZ1 (Figure 1F), SAF-A (Figure 1B,E), and indeed most of the NM proteins we have looked at, this illustrates the existence of normally unseen populations. Mock digested ‘control’ samples show the effectiveness of incubation in high-salt buffers (C3) at removing the epitopes that are available in untreated or detergent treated samples (C1, 2), without revealing those that are obscured by chromatin. When DNaseI is included in the incubation, and chromatin removed (C4) an apparently separate population is revealed. Together these serial nuclear extraction profiles show that what we see, when looking at total (unextracted) nuclei by immunofluorescence, is not everything and may in fact not be the functional fraction.

2.0 Materials

Prepare all buffers and buffer components in DEPC-treated water (1mL DEPC (Sigma Aldrich, D5758) per 1L of water), and ensure all extraction reagents are certified nuclease free. Store pre-prepared buffers at -20°C.

2.1 Base Buffer recipes

1. Cytoskeletal Buffer (CSK): 10mM PIPES/KOH (pH6.8), 100mM NaCl, 300mM Sucrose, 1mM EGTA, 1mM MgCl₂. Store in 10-50mL aliquots at -20°C. Before use, add fresh DTT to a final concentration of 1mM, and protease inhibitor (PI) cocktail tablet as recommended (Roche 05056489001). Do not re-freeze.
2. 10X DNase Buffer: 400mM Tris-HCl, 100mM NaCl, 60mM MgCl₂, 10mM CaCl₂, pH 7.9. Store at -20°C in 5mL aliquots. Add 45mL of DEPC water before use.
3. Cell culture grade Phosphate Buffer Saline (PBS)

2.2 Reagent List

1. Cells attached to glass coverslips, grown to 50-80% confluent density in a 24 well plate.
2. 8% Paraformaldehyde
3. Prepare buffer **CSK-D** by adding Triton-X-100 to CSK to a final concentration of 0.1%. We use a pre-diluted 10% stock solution to enable rapid homogenisation.
4. Prepare buffer **CSK-DS** by adding 5M NaCl to CSK-D to a final total concentration of 500mM (CSK already contains 100mM NaCl).
5. Immediately before use, prepare nucleases by adding RNase-free DNaseI (Roche, 04716728001), or DNase-free RNase (Roche, 11119915001) or both nucleases at 1/20 dilution, to 1X DNase Buffer on ice (make up enough for 20µL per coverslip). Keep stock nucleases at -20°C until needed and return immediately after use. Nucleases may need to be titrated to achieve maximal extraction.
6. Where RNase inhibitors (RI) are required, add Vanadyl Ribonucleoside Complex (VRC) (NEB, S1402S) to a final concentration of 2.5mM.
7. Affinity purified primary antibodies, and relevant secondary antibody for detection. For example in Figure 1B we have used mouse anti-SAF-A (Abcam,

ab10297) primary antibody with goat anti-Mouse IgG Alexa Fluor® 488 (ThermoFisher Scientific, A-11001).

8. Hoechst 33258 (10 ng/ml in PBS), or titrated DAPI DNA counterstain.
9. Mounting medium (without DNA counterstain). We use Vectashield (Vector Laboratories H-1000).
10. For immunofluorescence, protease-free BSA (Jackson Laboratories, 001-000-162).

2.3 Equipment

1. 13 mm glass coverslips
2. 24 well cell culture plate
3. Fine forceps and needle for moving coverslips
4. Incubator at 37°C
5. Humidified chamber for nuclease treatments and antibody incubations
6. Parafilm for nuclease treatments and antibody incubations
7. Glass slides
8. Microscope with a 63x oil immersion objective, and appropriate fluorescent image capture system. We use a Zeiss Axiovert 200M, Axiocam and Axiovision image acquisition software with base exposures of 50ms (Blue), 300ms (Red) and 600ms (Green).
9. Image analysis software. We use Adobe Photoshop or Affinity Photo and apply identical manipulation to all samples, so that image intensities reflect actual relationships unless specifically stated otherwise.

3.0 Methods

1. Thaw reagents and prepare all buffers.

2. Fix coverslip 1 (C1) in 500 μ L 4% PFA with RNase inhibitors (RI), for 15 minutes at room temperature before moving to a fresh well with cold PBS.
3. Incubate C2-C6 in 500 μ L CSK-D with RI for 1 minute.
4. Fix C2 in 500 μ L 4% PFA with RI for 15 minutes at room temperature, before moving to cold PBS.
5. Incubate C3 and C4 in 500 μ L CSK-DS with RI, and incubate C5 and C6 in 500 μ L CSK-DS (without RI), for 1 minute.
6. Wash coverslips twice in 500 μ L DNase buffer with (C3 and C4) or without (C5 and C6) RI, for 1 minute each.
7. Move coverslips C3-C6 to parafilm in a humidified chamber and, after addition of the following reagents, incubate for 1 hour at 37°C:
 - i) 20 μ L DNase buffer with RI (C3)
 - ii) 20 μ L DNase buffer with RI, plus 1/20 DNaseI (C4)
 - iii) 20 μ L DNase buffer, plus 1/20 RNase (C5)
 - iv) 20 μ L DNase buffer, plus 1/20 DNaseI and 1/20 RNase (C6)
8. Incubate C3 and C4 in 500 μ L CSK-DS with RI, and C5 and C6 in 500 μ L CSK-DS without RI for 1 minute.
9. Fix C3-C6 in 500 μ L 4% PFA with RI for 15 minutes at room temperature, before moving to cold PBS.
10. Coverslips should be processed as soon as possible by standard immunofluorescence protocols and imaged by high-resolution fluorescence microscopy. In brief, our protocol is as follows, but should be optimised for individual antigens taking into account the range of intensities expected across an extraction series. As a starting point, we optimise new antibodies using detergent-

treated cells (C2), bearing in mind that subsequent extractions can both improve and reduce signal.

- i) Block the coverslips in 10mg/ml BSA in PBS (with 0.1% Triton-X-100 and 0.02% SDS) for 10 minutes.
- ii) Incubate with affinity purified primary antibodies to your protein of interest for 1 hour in a humidified chamber at 37°C (in Figure 1F we used rabbit anti-CIZ1 1794 at 1:1000).
- iii) Wash and block the coverslips with 3 washes of 10 minutes in BSA antibody buffer, as above.
- iv) Incubate with the relevant secondary antibody for 1 hour at 37°C in a humidified chamber (in Figure 1F goat anti-rabbit IgG Alexa Fluor® 568 (ThermoFisher Scientific, A-11011) at 1:1000).
- v) Wash the coverslips 3 times for 10 minutes in BSA antibody buffer.
- vi) Wash coverslips with PBS.
- vii) Incubate for 5 minutes with Hoechst 33258 (10 ng/ml) in PBS in the dark.
- viii) Dip coverslips in distilled water and mount cells facing down, onto slides with Vectashield.
- ix) Slides can now be imaged.

11. Image a sample series using identical parameters. Often this requires optimisation of image acquisition parameters on a treated sample of interest, followed by application to all samples. Line intensity scans on the resulting images can be performed in Image J by drawing a straight line across composite images of the nuclei as shown in Figure 1C and the intensity profile plotted for multiple colours from the same line.

Notes

1. When performing this protocol for the first time, include immunodetection of a protein with known behaviour. This will enable you to better identify the potential issues described below. Examples of proteins that are lost from normal, wild type early passage cultured cells at different stages of the protocol are:-
Detergent step: PCNA is lost from non-S phase cells, unfused exogenous GFP.
High-salt wash (mock incubation): Available SAF-A is lost. This is primarily done as a control for nuclease incubation to show any loss is due to the nuclease and not the incubation in nuclease buffer, or subsequent high-salt wash. It is not uncommon to see a decrease in signal following this treatment with proteins which also have epitopes buried by chromatin.
DNaseI: MCM2, Histones, DNA stains.
RNase: CIZ1 is lost from Xi, SAF-A is lost nucleus-wide, RNA stains.
Remaining: Lamins, CIZ1 (nucleus-wide but not enriched at Xi)
2. Unanticipated loss of a putative nuclear matrix protein upon mock extraction suggests a degree of RNase contamination in the process. We recommend carrying out all RNase-containing steps in a separate lab space with dedicated pipettes and humidified chamber. Lab spaces and equipment used for this protocol should be cleaned with RNase elimination reagents such as RNase *AWAY*® to minimise risk of contamination.
3. Unanticipated retention of an epitope thought to be RNase sensitive, through the extraction process to C5 and C6, can occur if RNase inhibitors from step 3 are not thoroughly washed out. The CSK-DS and DNase buffer washes prior to enzyme incubation can be lengthened, or the RNase inhibitors in the CSK-D treatment in step 3 for C5 and C6 can be left out.

4. Unanticipated retention of chromatin is a common problem, often the result of sub-optimal enzyme activity. High levels of salt remaining from the high salt wash can inhibit the DNase enzyme, and is remedied by increasing the time, volume and number of DNase buffer washes. Do not omit the pre-digestion high-salt wash step completely, as prior removal of salt-sensitive proteins improves DNaseI access and efficiency. Avoid using higher concentrations of DNaseI.
5. Titrate Hoechst/DAPI or other DNA counterstain to the minimum concentration required to visualise unextracted nuclei. If the level of DNA counterstain is high it may not be possible to visualise the effect of DNaseI, as no NM extraction method will remove 100% of DNA. S/MAR DNA is protected from digestion by association with the NM, and will return a fluorescence signal that can saturate an uncalibrated system.
6. An alternative strategy to assess how effective and selective DNaseI treatments have been is to stain with propidium iodide. This will shift from nucleus-wide stain in the presence of chromatin, to RNA-dependant signal concentrated in nucleoli in chromatin-depleted cells.
7. This protocol is not compatible with loosely adhered cells or suspension cultures (e.g. embryonic stem cells), though, an in-solution version suitable for analysis by western blot, is possible. This was previously described, without inclusion of RNA analysis steps [15].
8. The serial extraction protocols described here are particularly powerful when combined with other cell culture techniques such as cell synchrony and expression of exogenous proteins. The effect of transfection processes on the nuclear matrix must be considered, especially when using electroporation techniques.

9. It is imperative that image capture and analysis parameters, including any manipulation for presentation purposes, are identical across a series of coverslips to ensure accurate representation of relative protein levels. This allows direct comparisons to be made between treatments, based on fluorescence intensity. For each protein to be investigated, determine the minimum exposure required to image detergent-treated cells (C2) and apply identical parameters across the whole series. In some cases it may be necessary to supplement a controlled series of image acquisitions, with additional imaging to record extremely low or extremely high fluorescence depending on the capability of your camera.
10. Not all cells in a population respond similarly to extraction, in many cases reflecting biological differences within a population. All image analysis must be performed on statistically meaningful numbers of cells, and representative images accompanied by numerical data.

Acknowledgements

We are grateful to Barbora Kriukelyte for the PTBP1 antibody staining and Justin Ainscough for his comments. This work was funded by ERS PhD scholarship as part of BBSRC White Rose Doctoral Training Partnership grant BB/M011151/1.

References

1. Roshon MJ, Ruley HE (2005) Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res* 14 (2):179-192
2. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Nakagawa S (2010) The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* 19 (3):469-476. doi:10.1016/j.devcel.2010.08.006
3. Ma N, Matsunaga S, Morimoto A, Sakashita G, Urano T, Uchiyama S, Fukui K (2011) The nuclear scaffold protein SAF-A is required for kinetochore-microtubule attachment and contributes to the targeting of Aurora-A to mitotic spindles. *J Cell Sci* 124 (Pt 3):394-404. doi:10.1242/jcs.063347
4. Ridings-Figueroa R, Stewart ER, Nesterova TB, Coker H, Pintacuda G, Godwin J, Wilson R, Haslam A, Lilley F, Ruigrok R, Bageghni SA, Albadrani G, Mansfield W,

- Roulson JA, Brockdorff N, Ainscough JFX, Coverley D (2017) The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* doi:10.1101/gad.295907.117
5. Helbig R, Fackelmayer FO (2003) Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma* 112 (4):173-182. doi:10.1007/s00412-003-0258-0
 6. Smeets D, Markaki Y, Schmid VJ, Kraus F, Tattermusch A, Cerase A, Sterr M, Fiedler S, Demmerle J, Popken J, Leonhardt H, Brockdorff N, Cremer T, Schermelleh L, Cremer M (2014) Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin* 7:8. doi:10.1186/1756-8935-7-8
 7. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161 (2):404-416. doi:10.1016/j.cell.2015.03.025
 8. Yamada N, Hasegawa Y, Yue M, Hamada T, Nakagawa S, Ogawa Y (2015) Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome. *PLoS Genet* 11 (8):e1005430. doi:10.1371/journal.pgen.1005430
 9. Kolpa HJ, Fackelmayer FO, Lawrence JB (2016) SAF-A Requirement in Anchoring XIST RNA to Chromatin Varies in Transformed and Primary Cells. *Dev Cell* 39 (1):9-10. doi:10.1016/j.devcel.2016.09.021
 10. Sakaguchi T, Hasegawa Y, Brockdorff N, Tsutsui K, Tsutsui KM, Sado T, Nakagawa S (2016) Control of Chromosomal Localization of Xist by hnRNP U Family Molecules. *Dev Cell* 39 (1):11-12. doi:10.1016/j.devcel.2016.09.022
 11. Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M (2016) Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* 354 (6311):468-472. doi:10.1126/science.aae0047
 12. Jeon Y, Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146 (1):119-133. doi:10.1016/j.cell.2011.06.026
 13. Makhoulouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C (2014) A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat Commun* 5:4878. doi:10.1038/ncomms5878
 14. Verheijen R, van Venrooij W, Ramaekers F (1988) The nuclear matrix: structure and composition. *J Cell Sci* 90 (Pt 1):11-36
 15. Wilson RH, Hesketh EL, Coverley D (2016) Preparation of the Nuclear Matrix for Parallel Microscopy and Biochemical Analyses. *Cold Spring Harb Protoc* 2016 (1):pdb prot083758. doi:10.1101/pdb.prot083758

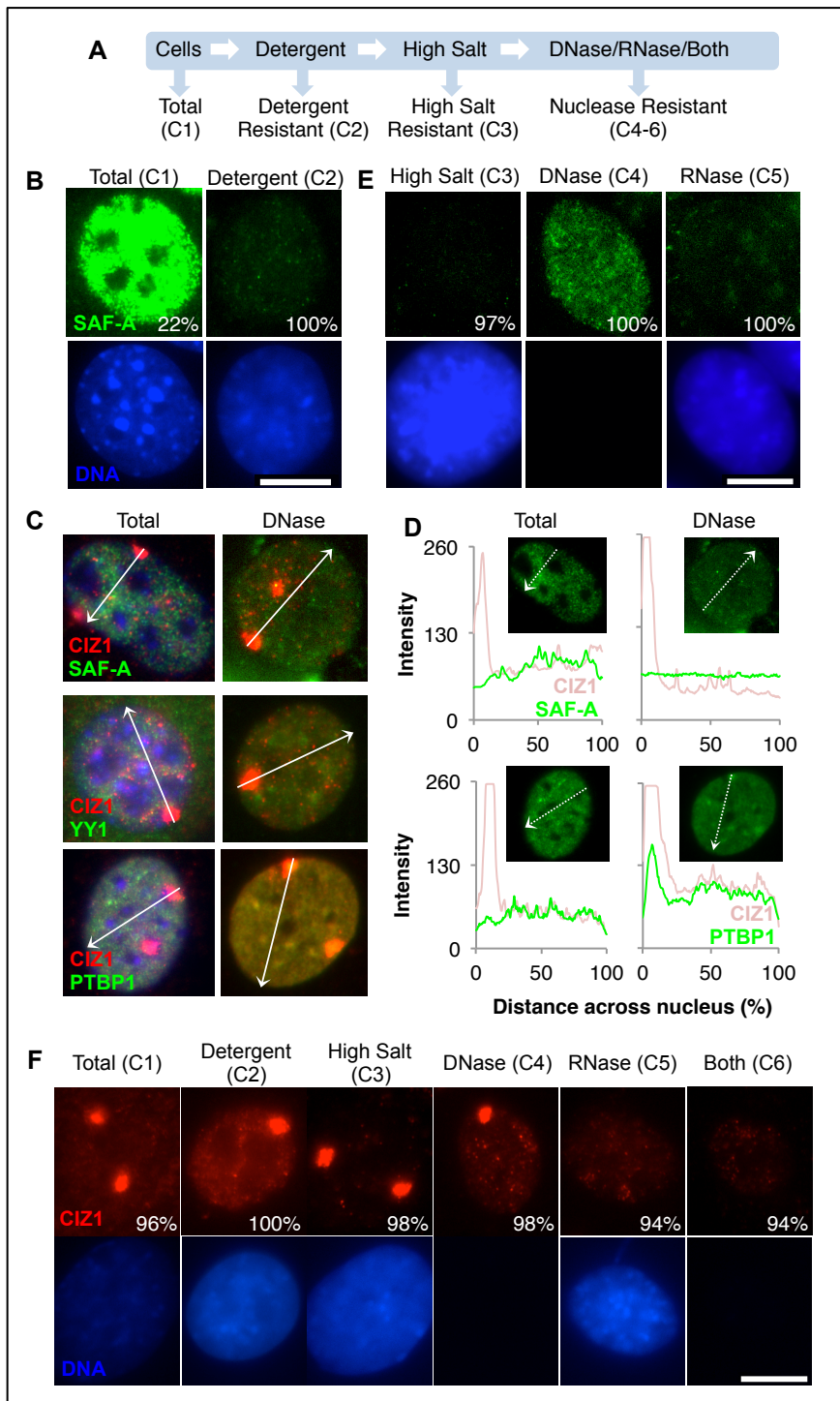


Figure Captions

Figure 1: Nuclear matrix extraction reveals epitopes that are masked by chromatin

- A) Overview of the nuclear matrix extraction protocol described here. Unextracted cells (C1) are serially treated to remove soluble proteins and lipids with detergent containing buffers (C2), loosely attached proteins with high salt containing buffers (C3) and chromatin or RNA bound factors (C4-6) with nucleases, to leave behind the nuclear matrix (NM). Absence of RNase reveals the RNA-protein NM, and its inclusion the core protein NM.
- B) Total (unextracted) and detergent-treated mouse embryonic fibroblast nuclei immuno-stained for SAF-A in green (ab10297). Images and modifications were performed identically. The percentage of cells within the cycling population with the staining pattern shown in the image is indicated ($n > 100$, scale bar is 10 microns). DNA is shown in blue.
- C) Images of total and DNaseI extracted mouse embryonic fibroblast nuclei immuno-stained for CIZ1 in red (1794, [4]) in combination with SAF-A (ab10297), YY1 (SC7341) or PTBP1 (ab5642) all in green. Images for different antibodies are not identically modified. Lines indicate the path of intensity scans performed in D.
- D) Line intensity scan data of SAF-A (upper) and PTBP1 (lower) showing total and DNaseI extracted nuclei from C. CIZ1 is shown in red to indicate the location of Xi, and SAF-A or PTBP1 in green. The green component of each nucleus that gave rise to the data is shown above each graph. Images are representative of the population of cells generated by extractions C1 and C4.
- E) Signal remaining after further extraction of mouse fibroblast nuclei from B with high salt, DNase or RNase. Images and modifications were performed identically across the series, but are different from those in B in order to capture a lower signal range. The percentage of cells within the cycling population with the staining pattern shown in the image is indicated ($n > 100$, scale bar is 10 microns). SAF-A is shown in green, DNA is shown in blue.
- F) A full extraction series applied to mouse fibroblasts and immuno-stained for CIZ1 N-terminal replication domain in red (1794, [4]). The percentage of cells within the cycling population with the staining pattern shown in the image ($n > 100$). DNA is shown in blue. All images were captured and processed under identical conditions. Scale bar is 10 microns.

Appendix D



PDF of **Stewart ER**, Turner RML, Newling K, Ridings-Figueroa R, Scott V, Ashton PD, et al. Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture. Nat Commun. 2019;10(1):460.

ARTICLE

<https://doi.org/10.1038/s41467-018-08072-2>

OPEN

Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture

Emma R. Stewart¹, Robert M.L. Turner¹, Katherine Newling ², Rebeca Ridings-Figueroa^{1,3}, Victoria Scott¹, Peter D. Ashton², Justin F.X. Ainscough ¹ & Dawn Coverley¹

The inactive X chromosome (Xi) serves as a model for establishment and maintenance of repressed chromatin and the function of polycomb repressive complexes (PRC1/2). Here we show that Xi transiently relocates from the nuclear periphery towards the interior during its replication, in a process dependent on CIZ1. Compromised relocation of Xi in CIZ1-null primary mouse embryonic fibroblasts is accompanied by loss of PRC-mediated H2AK119Ub1 and H3K27me3, increased solubility of PRC2 catalytic subunit EZH2, and genome-wide deregulation of polycomb-regulated genes. Xi position in S phase is also corrupted in cells adapted to long-term culture (WT or CIZ1-null), and also accompanied by specific changes in EZH2 and its targets. The data are consistent with the idea that chromatin relocation during S phase contributes to maintenance of epigenetic landscape in primary cells, and that elevated soluble EZH2 is part of an error-prone mechanism by which modifying enzyme meets template when chromatin relocation is compromised.

¹Department of Biology, University of York, York YO10 5DD, UK. ²York Bioscience Technology Facility, University of York, York YO10 5DD, UK. ³Present address: Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK. Correspondence and requests for materials should be addressed to D.C. (email: dawn.coverley@york.ac.uk)

The inactive X chromosome (Xi) is a discrete unit of facultative heterochromatin that is selected for repression early in the development of female mammals as a means of equalizing X-linked gene dosage between the sexes¹. *Xist* long noncoding RNA (LNCRNA) plays an essential role in the recruitment of chromatin modifying enzymes to Xi, and the progressive formation of a stable, heritable repressed state². Detailed analysis shows that *Xist*-dependent polycomb recruitment is initiated by PRC1, and is dependent on the RNA-binding protein hnRNPK through its interaction with sequences encoded by *Xist* repeat B³. Later steps in the polycomb cascade result in the accumulation of PRC1-mediated H2AK119ub1 and PRC2-mediated H3K27me3 on Xi chromatin, which is then maintained through subsequent rounds of cell division⁴. CIP1/CDKN1A-interacting zinc finger protein 1 (CIZ1) is recruited to Xi by *Xist* during the earliest stages of X-inactivation dependent on sequences encoded by *Xist* repeat E^{5,6}, though lack of overt embryonic phenotype in CIZ1 null mice suggest that there is no requirement for CIZ1 during these early stages of X-inactivation⁵. However, CIZ1 is required for retention of *Xist* at Xi in differentiated fibroblasts, and essential for its recruitment during lymphocyte activation in response to antigen stimulation in adult mice⁵, suggesting that it has a post-developmental function at Xi.

CIZ1 has been linked with the neurological disorders cervical dystonia⁷ and Alzheimer's disease⁸, and with both paediatric⁹, and adult common solid tumours including lung, colon, liver and breast^{10–13}, though no known underpinning molecular function convincingly links its role in these diverse human pathologies. Similarly, while a link with lymphocyte activation is established, the molecular mechanism that underpins its ability to guard against leukemias and lymphomas in mice is not understood^{5,11,14}. Moreover, while enrichment at Xi in female cells is striking (Xi-CIZ1), CIZ1 protein also occupies nucleus-wide foci in male and female somatic cells (focal-CIZ1)⁵, and is elevated in post-replicative male germ cells¹⁵ suggesting that it has additional functions unrelated to the inactive X-chromosome.

In the present study, Xi serves as a well-defined model to probe the mechanism of action of CIZ1, and shows that CIZ1 is required to support a change in the preferred location of Xi, between the nuclear periphery and the nuclear interior, during a brief window coincident with Xi replication. In CIZ1 null fibroblasts, failure to internalize is accompanied by the loss of PRC1/2-mediated modification of Xi chromatin, and relaxation of control over PRC1/2 target genes across the genome. Crucially, S-phase internalization of Xi is not observed in fibroblasts in long-term culture, even if CIZ1 is present, suggesting that the process in which CIZ1 normally functions is fragile, and corrupted at some level in cell lines. Moreover, the loss of function in cell lines is accompanied by up-regulation and increased solubility of PRC2 catalytic subunit EZH2, and in CIZ1 null cells, partial reinstatement of chromatin modification at Xi. This raises the possibility that the mechanism by which modifying enzyme and target chromatin meet is not the same in primary cells and derived cell lines. The data support the idea that chromatin relocation during S phase plays a role in the maintenance of epigenetic state in primary differentiated cells.

Results

Interaction between CIZ1 and nuclear matrix at Xi in S phase.

Enzymatic removal of chromatin (DNase1) or exposure to elevated non-physiological salt concentrations (500 mM NaCl) have little effect on either Xi-CIZ1 or focal-CIZ1^{5,16}, indicating that their location in the nucleus is not specified by association with chromatin. However, Xi-CIZ1 is sensitive to digestion with RNase in the majority of cells in a cycling population, indicating that

attachment at Xi is by association with RNA⁵, most likely *Xist*^{6,17}. We focus here on the small fraction of cells (3–7% depending on cell line and rate of cycling) that resist extraction with RNase so that CIZ1 remains anchored at Xi (Fig. 1a, b). This indicates additional or alternative interaction between CIZ1 and non-RNA, non-chromatin higher-order assemblies, consistent with a biochemically defined nuclear matrix (NM).

The low frequency of retention implies temporal control, which was supported by analysis in S phase cells. Incorporation of the nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU) identified Xis engaged in DNA replication and showed that Xi replication occurs in mid S phase in most cells in the fibroblast populations studied here, though the timing of Xi replication appears to vary with cell type^{18,19} (Supplementary Fig. 1). When combined with immuno-staining for CIZ1, it also showed that CIZ1 is significantly more likely to be resistant to extraction at the time of Xi replication (Fig. 1c, Supplementary Fig. 1). In fact, over half of Xi-replicating cells retain CIZ1 at Xi after RNase digestion (Fig. 1d). This was evident in triplicate independent populations of primary embryonic fibroblasts (PEFs, defined here as passage 5 or earlier), and remained unchanged in all derived populations adapted to long-term growth in culture (referred to as mouse embryonic fibroblasts, MEFs). Thus, during Xi replication, Xi-CIZ1 undergoes transient interaction with a structure that is independent of chromatin or RNA.

A shift in Xi location in primary cells but not cell lines.

The location of replicating CIZ1-marked Xis was not the same in PEFs and MEFs. In primary cells, it was significantly more likely to be seen at locations that did not contact the nuclear periphery (Fig. 1e). To confirm the preferred position independently of CIZ1, we monitored the location of EdU-labelled Xis at the time of Xi replication/EdU incorporation (pulse) compared to its location 30 min later (chase), and relative to the nucleolar marker fibrillar (Supplementary Fig. 1). This showed a greater likelihood of overlap with fibrillar-enriched regions at the time of the pulse, compared to after a chase period, indicating a change in preferred location during S phase. Although Xi is located at the nuclear periphery most of the time, anchored via lamin B receptor²⁰, these results are consistent with the very earliest descriptions of the Xi as occupying two distinct locations; the nuclear periphery or adjacent to nucleoli²¹.

Importantly, in three populations of primary cells isolated from independent WT embryos, this change in preferred location during S phase was readily monitored using a binary classification of peripheral or internal (P or I, Fig. 1f), but in culture-adapted lines derived from all three parent populations the shift in location was not evident (Fig. 1f), though replication of Xi took place in a similarly co-ordinated manner as a condensed EdU-labelled, CIZ1-marked entity. These data show that Xi replication is normally coincident with departure from the nuclear periphery, and suggest that the underlying mechanism for the change in preferred location is fragile and prone to degradation in culture-adapted cell lines. Moreover, it suggests that most cell lines likely do not represent the state of cells in the body with respect to chromatin dynamics, leading to questions about the functional relevance of the S-phase location change.

CIZ1 is involved in PRC1/2 target gene regulation. As might be expected, transcriptome analysis revealed extensive differences between primary parent cell lines and their culture-adapted derivatives, with approximately a thousand genes significantly affected (1038 at $q < 0.05$ (cuffdiff), Fig. 2, Supplementary data 1). Similar results were achieved using primary and culture-adapted populations from CIZ1 null mice (904 at $q < 0.05$ (cuffdiff), Fig. 2,

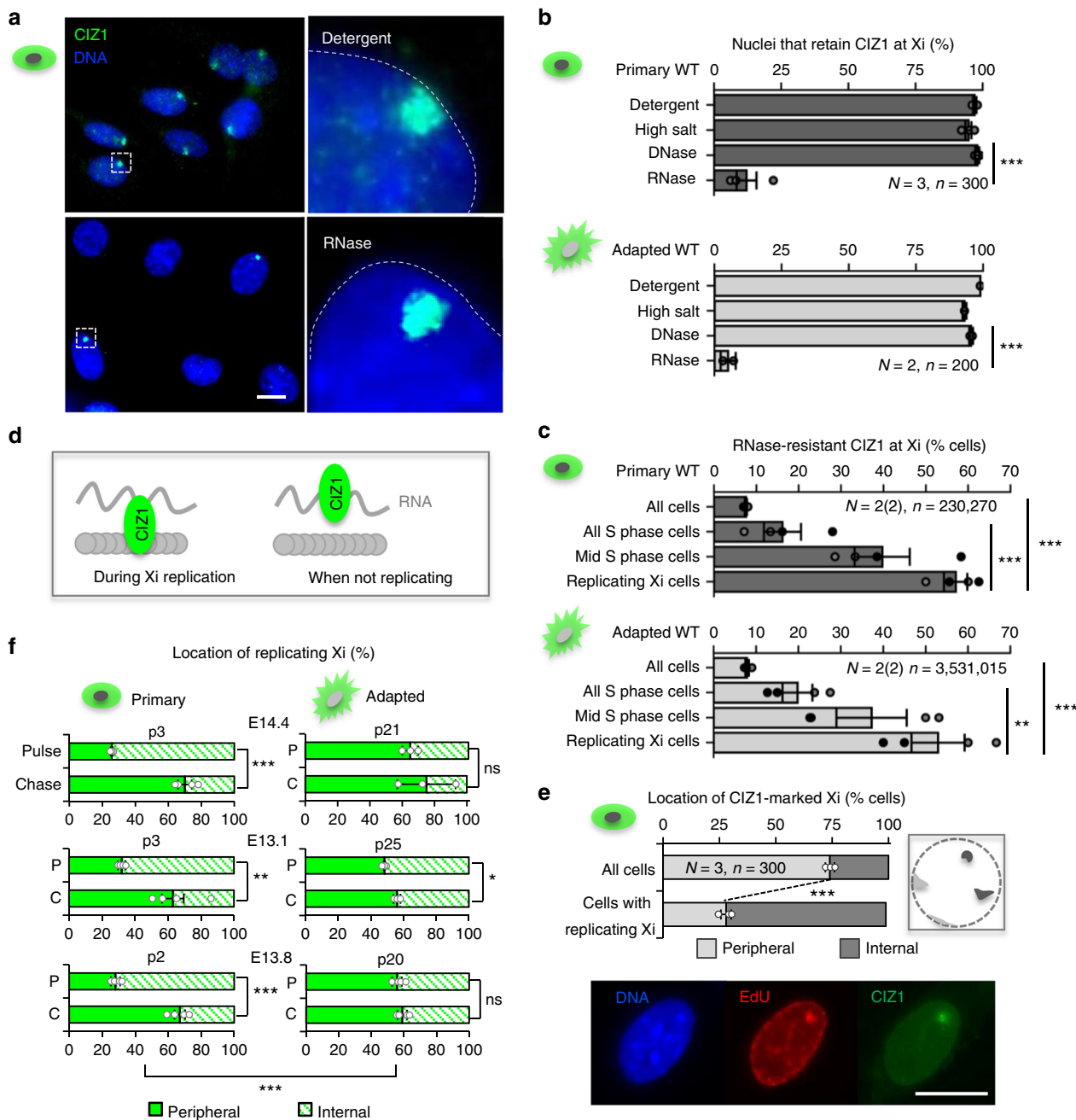


Fig. 1 Xi location during S phase. **a** CIZ1 at Xi in cycling WT primary embryonic fibroblasts (PEFs), with and without extraction of RNA and associated proteins, detected with anti-CIZ1 antibody 1794 (green). DNA is blue, scale bar 10 microns. **b** Proportion of cells that retain CIZ1 at Xi (Xi-CIZ1) at the indicated steps after extraction, for three independent populations of WT PEFs (means \pm SEM for 13.1, 13.8, 14.4 at passage 3–4, $n > 100$ for each step and each line. Significant differences are indicated (t -test). Similar results for culture-adapted derivative line 13.1 (passage > 20), and 3T3 cells is shown below, \pm SDEV. **c** Proportion of cells with RNase-resistant Xi-CIZ1 after pulse-labelling with EdU to classify cells by S-phase stage (Supplementary Fig. 1). Upper graph shows mean data for WT primary cells 13.1 and 13.8 (passage 3–4), each in duplicate, \pm SEM (t -test). Lower graph shows the same analysis for WT MEFs (13.1, passage > 20), and 3T3 cells, each in duplicate. **d** Illustration of transient attachment of Xi-CIZ1 (green) to non-chromatin, non-RNA assemblies (grey circles) during Xi replication. **e** Location of Xi-CIZ1 at the time of Xi replication (detected by 30 min pulse label with EdU) compared to the rest of the cell cycle. Location is classified as peripheral or internal (illustrated right). Means \pm SEM for three WT PEFs (13.1, 13.8, 14.4 p3–4) are shown ($n = 100$ for each), with significant changes indicated (t -test). Images show a cell engaged in replication of Xi (EdU, red), co-stained with anti-CIZ1 1794 (green), DNA is blue, scale bar 10 microns. **f** Xi location (peripheral or internal) at the time of incorporation of EdU (pulse), or 30 min later (chase), analysed by t -test. Graphs show data \pm SEM for three WT primary populations (13.1, 13.8, 14.4 p2–3), and three derived MEF lines ($p > 20$). EdU-Xi in primary WT cells shifts preferred location during the chase period, but remains unchanged in most adapted lines, compared by ANOVA. For all analyses $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. $N =$ independent cell lines, $n =$ total nuclei scored. Individual replicates contributing to each mean are overlaid on bar charts, shaded according to cell line where necessary

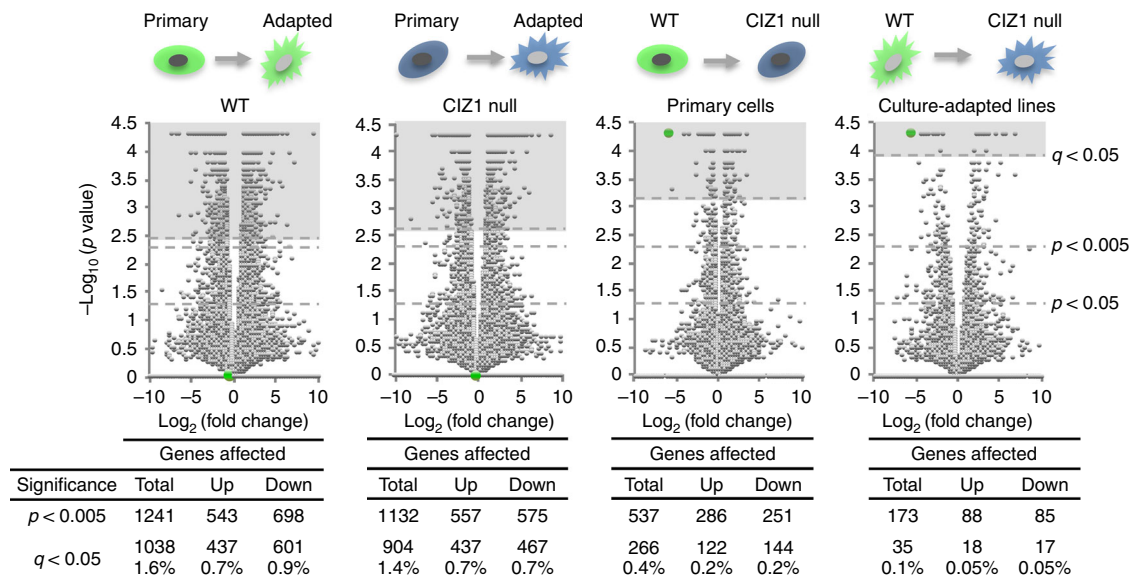


Fig. 2 Effect of loss of CIZ1 and culture adaption on gene expression. Volcano plots showing mean fold change in transcript level (FPKM) against significance (P -value, calculated by cuffdiff) for all 65530 annotated transcription units assigned to mouse genome assembly GRCm38, derived from whole transcriptome RNA-seq of three primary WT (13.1 p4, 13.8 p4, 14.4 p4) embryonic fibroblast populations and culture-adapted cell lines derived from the same primary cells, and three primary CIZ1 null (13.15 p3, 13.17 p3, 14.2 p4) embryonic fibroblast populations and culture-adapted cell lines derived from the same primary cells. CIZ1 is indicated in green. The populations of cells being compared are indicated above each plot. Below, the number of transcription units affected at the indicated significance thresholds. Significantly affected gene lists ($q < 0.05$, calculated by cuffdiff) are in Supplementary data 1 (primary to culture-adapted comparisons for both genotypes), Supplementary data 4 (CIZ1-dependent genes in primary cells), Supplementary data 5 (CIZ1-dependent genes in culture-adapted cells)

Supplementary data 1), approximately half of which (405) were the same genes as those affected in WT cells. For both sets, Gene Set Enrichment Analysis (GSEA)²² revealed significant overlap with genes regulated by PRC1 and PRC2 (Fig. 3a), as well as those responsive to Rb and TGF β among others (Supplementary data 2, tabs 4, 5, 6). PRC subunit expression was itself disrupted upon culture adaption, evident in both transcript level changes (Supplementary Fig. 2) and in splice variant diversity for some subunits (Supplementary Fig. 3, 4).

More informative is the effect of deletion of CIZ1 in primary cells. We showed previously that, despite enrichment at Xi, the loss of CIZ1 does not result in widespread reactivation of the inactive X-chromosome with only 28 X-linked genes significantly upregulated⁵. Here, we show that the majority of the 65,530 transcription units that align to the rest of mouse genome assembly GRCm38 are also unaffected (Supplementary data 3, Fig. 2). However, a sub-set of 266 transcription units (0.4%, $q < 0.05$ (cuffdiff)) consistently escape normal regulation in PEFs (Supplementary data 4, Fig. 2), with a similar proportion up- and down-regulated. Importantly, re-expression of full-length CIZ1 from an inducible integrated vector⁵ reverses this change back towards WT levels in 75% of affected genes within 24 h (Fig. 3b, c, Supplementary data 4) Thus, the expression of a distinct subset of genes is under the regulation of CIZ1 in primary fibroblasts, and these are distributed across all chromosomes, including the X. This is consistent with the interpretation that polycomb-mediated regulation is the primary pathway governing the expression of a subset of X-linked genes, and that for the majority inactivity is maintained by other levels of regulation.

GSEA with these CIZ1-dependent transcription units also revealed highly significant overlap with genes regulated by PRC1 and PRC2 (Fig. 3a, Supplementary data 2, tab 1), though unlike culture-adapted cells there was little change in PRC subunit transcript levels (Supplementary Fig. 2), and also no gross change in the overall immuno-staining pattern for SUZ12, RING1B or

BMI1 in primary cells (Supplementary Fig. 2). Oncogenic Signature sets that are up-regulated upon the loss of PRC subunits, or down-regulated upon the gain of PRC subunits, show a strong correlation (Fig. 3d), while polycomb-linked Curated Gene sets also support a functional interaction, evidenced by highly significant relationships with those affected by manipulation of EZH2 (M4196²³, $q < 3.99E-17$, where q is one-sided Fisher's Exact test with Benjamini-Hochberg false discovery rate (FDR) correction,) or SUZ12 (M2293²⁴, $q < 4.53E-22$) among others (Fig. 3a, Supplementary data 2 and Supplementary Tables 1 and 2).

Notably, a similar analysis of the effect of the loss of CIZ1 in culture-adapted populations returned only 35 CIZ1-dependent genes ($q < 0.05$, cuffdiff) and only 12 of these were among those identified in primary cells (Fig. 3e-g, Supplementary data 5). This is because many of those genes whose expression is normally dependent on CIZ1 are affected by culture adaptation of the WT state (39% $q < 0.05$, 56% $P < 0.05$). Again this argues that adaption to culture involves corruption of the process in which CIZ1 normally functions.

Deletion of CIZ1 mimics the effect of culture adaptation.

Because CIZ1 is highly enriched at Xi and undergoes specific interactions during Xi replication, we tested whether CIZ1 is required for Xi internalization in triplicate independent primary CIZ1 null populations. The data show that, unlike WT cells, Xi fails to change the preferred location within the test window, remaining primarily associated with the nuclear periphery in both the pulse and chase (Fig. 4a). It remains possible that a change in the rate of Xi movement, rather than complete loss of relocation, could account for the output in this assay. However, the data clearly show that CIZ1 plays a role in determining Xi location across the time window (and by implication the residency time at the nucleolus) thereby linking altered relocation with compromised polycomb function. Consistent with this, both PRC2-

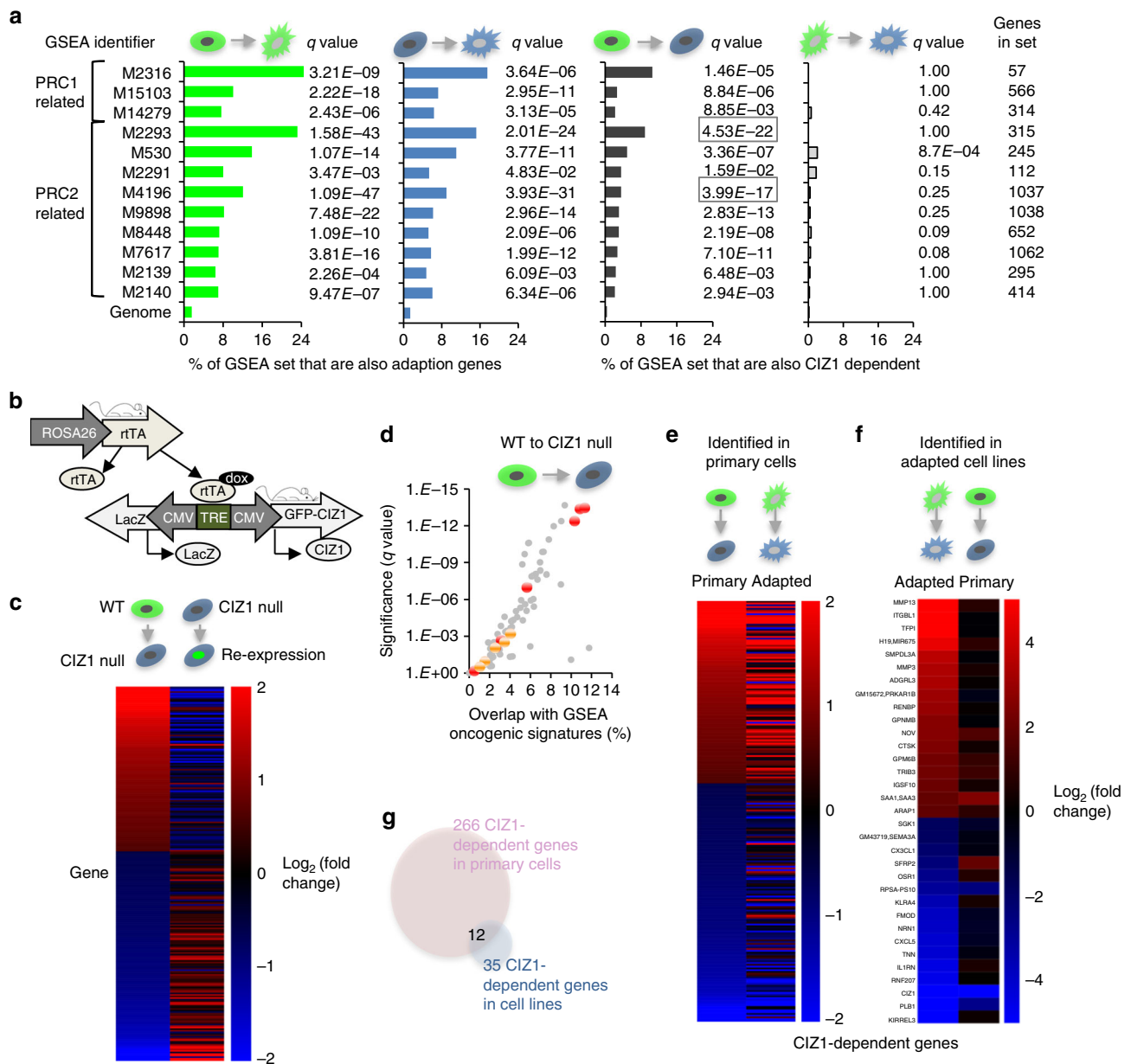


Fig. 3 Relaxation of polycomb-regulated gene expression. **a** Relationship between affected transcription units ($q < 0.05$, cuffdiff) in Fig. 2 and PRC1 or PRC2-related curated gene sets (GSEA MSigDB²²), with significance expressed as q value. Set identifiers and number of genes in sets are indicated. Overlap with genes affected by culture adaption of WT cells (green) and CIZ1-null cells (blue). Overlap with CIZ1-dependent genes in primary cells (dark grey) and culture-adapted cells (light grey). **b** Schematic of transgenes used to create doxycycline-inducible expression of full-length GFP-CIZ1 in CIZ1 null mice and derived cells⁵. **c** Heat map showing 266 transcription units ($q < 0.05$, cuffdiff) affected by the loss of CIZ1 in primary cells, organised by fold-change. Right column, mean fold change in the same transcription units in CIZ1 null cells (13.17 p3, 14.19p1), 24 h after induction of full-length CIZ1¹⁵ (Supplementary data 4, sheet 2). **d** Relationship between CIZ1-dependent transcription units in primary cells ($q < 0.05$, cuffdiff) and all 189 oncogenic signatures (GSEA MSigDB²²), showing overlap (% of genes that are affected by the loss of CIZ1) against significance. Grey dots indicate oncogenic signatures not related to PRC 1 or 2. Red, sets that are up-regulated when a PRC subunit is down-regulated, or vice versa. Yellow, sets that are upregulated when a PRC subunit is upregulated or vice versa. **e** Heat map showing fold-change in 266 CIZ1-dependent transcription units ($q < 0.05$, cuffdiff) in primary cells (left, same as Fig. 1c), compared to fold-change in the same genes in culture-adapted derivatives, upon loss of CIZ1 (right). Supplementary data 4, sheet 3. **f** Heat map showing the 35 transcription units that are CIZ1-dependent in adapted cells ($q < 0.05$, cuffdiff) (left), and fold change in these transcription units in primary cells (right). Supplementary data 5. **g** Venn diagram showing overlap between transcription units that are differentially expressed upon loss of CIZ1 in primary and derived lines. Highlighted in Supplementary data 5. All q values (false detection rate corrected P values) for overlap with GSEA MSigDB were calculated using one-sided Fisher's Exact tests with Benjamini-Hochberg false discovery rate correction. Heat maps are organised by fold-change from up (red) to down (blue)

mediated H3K27me³ and PRC1-mediated H2AK119ub1 are absent from Xi in primary CIZ1 null cells (Fig. 4b).

These data are consistent with the hypothesis that accurate receipt of polycomb-mediated chromatin marks is linked with the

delivery of template into sub-nuclear compartments enriched in spatially restricted modifying enzymes²⁵. In fact, EZH2, SUZ12 and Snf2h (part of the ACF1-ISWI chromatin remodelling complex involved in heterochromatin replication) have all been

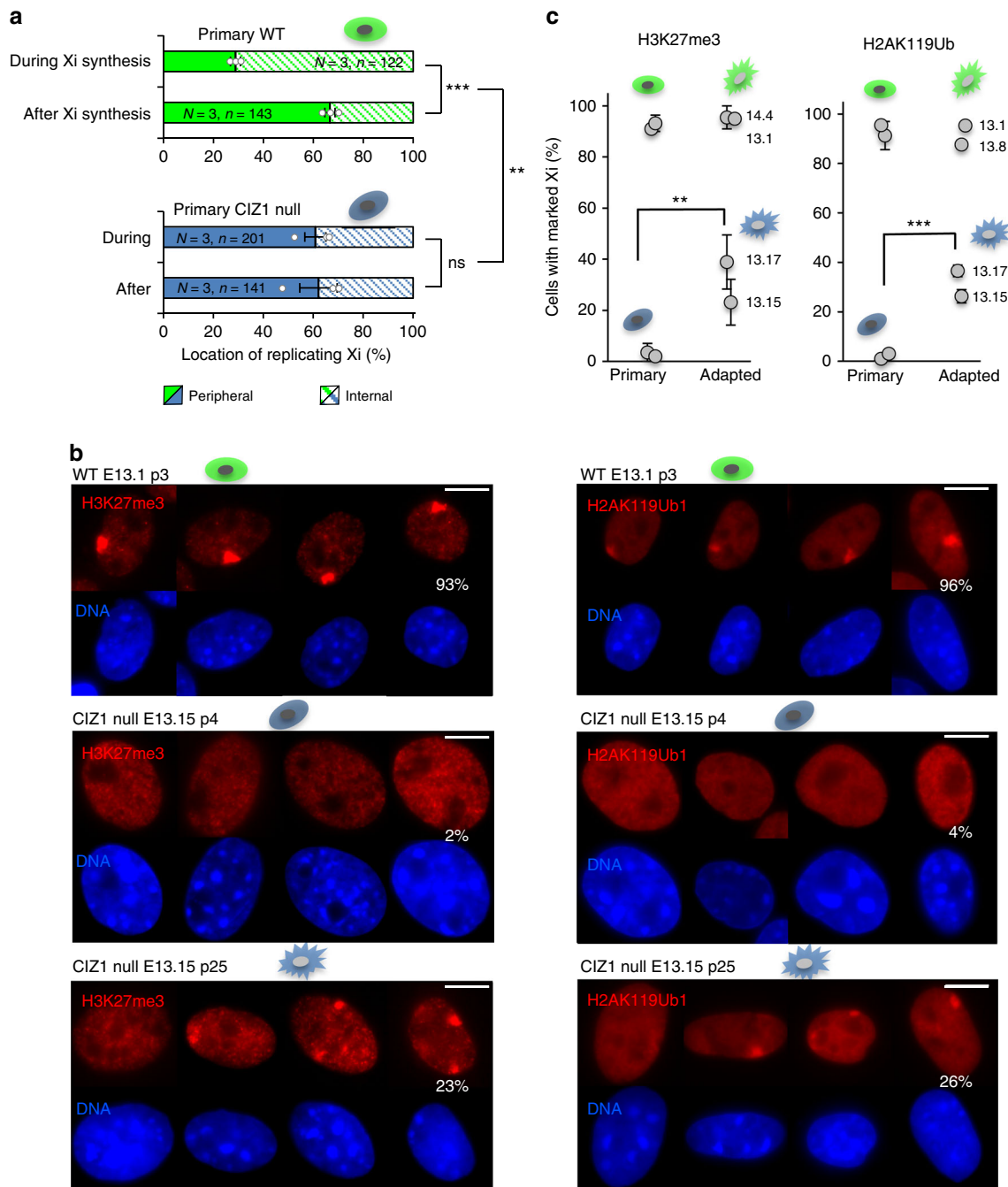


Fig. 4 CIZ1 is required for PRC-mediated chromatin modification in primary cells. **a** Binary classification of Xi location (peripheral or internal) at the time of incorporation of EdU (pulse) or after (chase), analysed by *t*-test. Graph shows mean data \pm SEM for three WT primary embryonic fibroblasts (green, 13.1, 13.8, 14.4 passage 2–3) and three CIZ1 null primary embryonic fibroblasts (blue, 13.15, 13.17, 14.2, passage 2–3). EdU-Xi in WT cells shifts preferred location during the chase period, but this remains unchanged in CIZ1 null cells, compared by ANOVA. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$. N = independent cell lines, n = nuclei scored. Individual data points contributing to the mean are overlaid. **b** Representative immuno-stain of PRC1-mediated H2AK119Ub1 and PRC2-mediated H3K27me3 in primary WT cells (13.1, p3), absence of these marks in primary CIZ1 null cells (13.15, p4), and re-emergence in culture-adapted CIZ1 null MEFs (13.15, p25). Percentage of cells with marked Xis is shown. Scale bar is 10 microns. **c** Graphs show the proportion of primary and culture-adapted cells with H3K27me3-marked Xis (left) and H2AK119Ub-marked Xis (right), in CIZ1 null (blue) and WT (green) genotypes, \pm SEM from triplicate analyses for each line. Significant changes upon culture adaptation are indicated, and compared by *t*-test

reported to be enriched at peri-nucleolar locations^{25–27}. In our analysis, peri-nucleolar enrichment of EZH2 was also detected, but varied greatly with both cell type and the EZH2 antibody used (see later).

Promiscuous polycomb activity in cell lines. Strikingly, enriched H3K27me3 and H2AK119ub1 re-emerges at Xi upon prolonged

culture of CIZ1 null cells (Fig. 4b, c), sometimes describing a more dispersed Xi territory than in WT cells. This suggests that the deficit caused by the loss of CIZ1 is overridden during prolonged culture, and raises the possibility that similar compensation might also be taking place in WT cells (in which chromatin relocation is compromised for reasons unrelated to CIZ1). It could in fact underpin the relaxed control over polycomb target

gene regulation illustrated in Fig. 3a, d, necessitating a direct comparison between the different cell states.

CIZ1 deletion or culture adaptation drives a shift in EZH2.

Protein level analysis of the PRC2 catalytic subunit EZH2 resolved two main isoforms in primary cells, though relative levels vary between WT and CIZ1 null cells (Fig. 5a, b, Supplementary Fig. 3). The upper form (designated EZH2p for primary) that is dominant in WT cells gives way to a lower form as cells are adapted to growth in culture, and this is increasingly up-regulated with passage (Fig. 5a, c). EZH2 transcript assemblies also identified differences in prevalence (Supplementary data 6, Supplementary Fig. 3). The most abundant assembly in adapted cells corresponds to UniProt Q61188 (equivalent to human Q15910-1, designated EZH2 α ²⁸), however this is vastly under-represented in primary cells and replaced by a form with additional 5' sequence. No canonical translational start codon is evident in the additional sequence which is currently annotated as 5' UTR. The relationship between EZH2p and the apparent 5' UTR sequence that is prevalent in the same cells (Supplementary Fig. 3, Supplementary data 6) is not clear at this time. Electrophoretic mobility suggests that the lower protein isoform is equivalent to canonical EZH2 α , so in primary fibroblasts EZH2 protein appears to include additional domains that are not part of the EZH2 that is expressed by cell lines. A similar analysis of transcript assemblies for the other subunits of both PRC1 and PRC2 identified additional changes linked with either deletion of CIZ1 or culture adaptation (Supplementary Fig. 4).

The relative solubility of EZH2 in primary and culture-adapted cells point to function-related differences. Both EZH2p and EZH2 α are fully resistant to salt extraction (500 mM) in primary cells, while half of the EZH2 α in culture-adapted cells is solubilized under the same conditions (Fig. 6a, b), possibly reflecting elevated levels and saturation of binding sites. Similar results were obtained by quantitative immunofluorescence in independent WT primary cells and derived cell lines (Fig. 6c, lower graph). Thus EZH2 shifts isoform, is dramatically elevated and considerably more soluble in cell lines.

Focusing on the primary cell state, we applied further extraction steps (Fig. 6a, e) to reveal proteins that are solubilised along with fragmentation and elution of chromatin, those that are part of assemblies that are dependent on nuclear RNA, and those that resist all extraction. This highlights differences in EZH2 that are driven by the loss of CIZ1. Unlike previous analyses of EZH2 in embryonic stem cells²⁹, the prevalent form in WT cells

(EZH2p) does not depend on either RNA or DNA for retention in the nucleus. However, the prevalent form in CIZ1 null cells (EZH2 α) is largely released by the removal of RNA (Fig. 6e, f). Consistent results were achieved by immuno-microscopy in two independent primary lines each for WT and CIZ1 null genotypes confirming distinct differences in extraction profile (Fig. 6c upper graph, d), and is in line with analyses of recombinant PRC2 which report promiscuous RNA-binding capability³⁰. Thus, EZH2 α lacks the ability to become anchored within the nucleus to non-chromatin, non-RNA protein assemblies (Fig. 6g).

Most of the analysis in this study was carried out using EZH2 antibody CSTD2C9 which recognises both EZH2p and EZH2 α in western blots (Fig. 5a), and gives relatively evenly distributed nuclear foci by immunofluorescence in either primary or adapted cells before extraction (Fig. 6d, Supplementary Fig. 3). In contrast, EZH2 antibody STJ112944 has higher affinity for EZH2p than EZH2 α in western blots (Supplementary Fig. 3) and reveals a distinctly different pattern in primary cells but not adapted cells (Supplementary Fig. 3). Comparison to the nucleolar marker fibrillar protein identified this as peri-nucleolar enrichment (Fig. 7a), and is consistent with the idea that EZH2p represents a functionally distinct and spatially restricted form of EZH2 in primary cells.

Both EZH2 antibodies were also applied in proximity ligation studies with CIZ1, to reveal locations in the nucleus where the two proteins are in close proximity (Fig. 7b). We did not see a large signal comparable to the enrichment of CIZ1 at Xi with either antibody, in any cell type, or at any stage in the cell cycle, which would be consistent with a stochastic interaction between EZH2 and Xi chromatin. However in WT cells, but not CIZ1 null cells, both EZH2 antibodies return a signal at discrete locations in the nucleus, indicating proximity to within 40 nm of each other at those sites. In line with the western blot data in Fig. 5a, for EZH2p (antibody STJ112944), the number of colocalizing sites is reduced in adapted WT cells compared to primary WT cells. Moreover, consistent with the observation that CIZ1-Xi is compromised in its ability to internalize in adapted cells, fewer of these colocalizing sites are located internally in culture adapted cells compared to primary cells.

Discussion

The data identify two drivers for the shift from EZH2p to EZH2 α ; the loss of CIZ1 in primary cells and culture-adaptation in both genotypes (Fig. 6g). Both drivers also result in the loss of observed S-phase location change and relaxed PRC-mediated control of

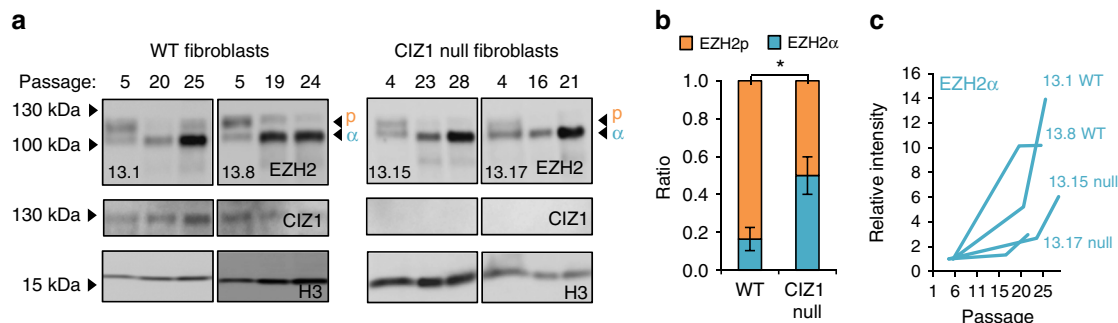


Fig. 5 Loss of CIZ1 drives changes in EZH2 isoform. **a** Western blot showing PRC2 catalytic subunit EZH2 (antibody D2C9) in representative primary embryonic fibroblasts and derived culture-adapted populations, of WT and CIZ1 null genotypes, showing two prominent isoforms in early passage cells, designated p and α . Relative affinity for these isoforms varies between commercially available anti-EZH2 antibodies (Supplementary Fig. 3). **b** Quantification of isoforms p and α in early passage WT and CIZ1 null cells, showing mean ratio for three independent populations of each type (WT; 13.1, 13.8, 14.4, CIZ1 null; 13.15, 13.17, 14.2). \pm SEM, significant changes are indicated (*t*-test). **c** Quantification of EZH2 α in primary cells and derived populations for the cell lysates shown in **a**, after normalization to histone (H3) and calibration to the early passage population for each line. All four, irrespective of genotype, show increasing EZH2 α with passage

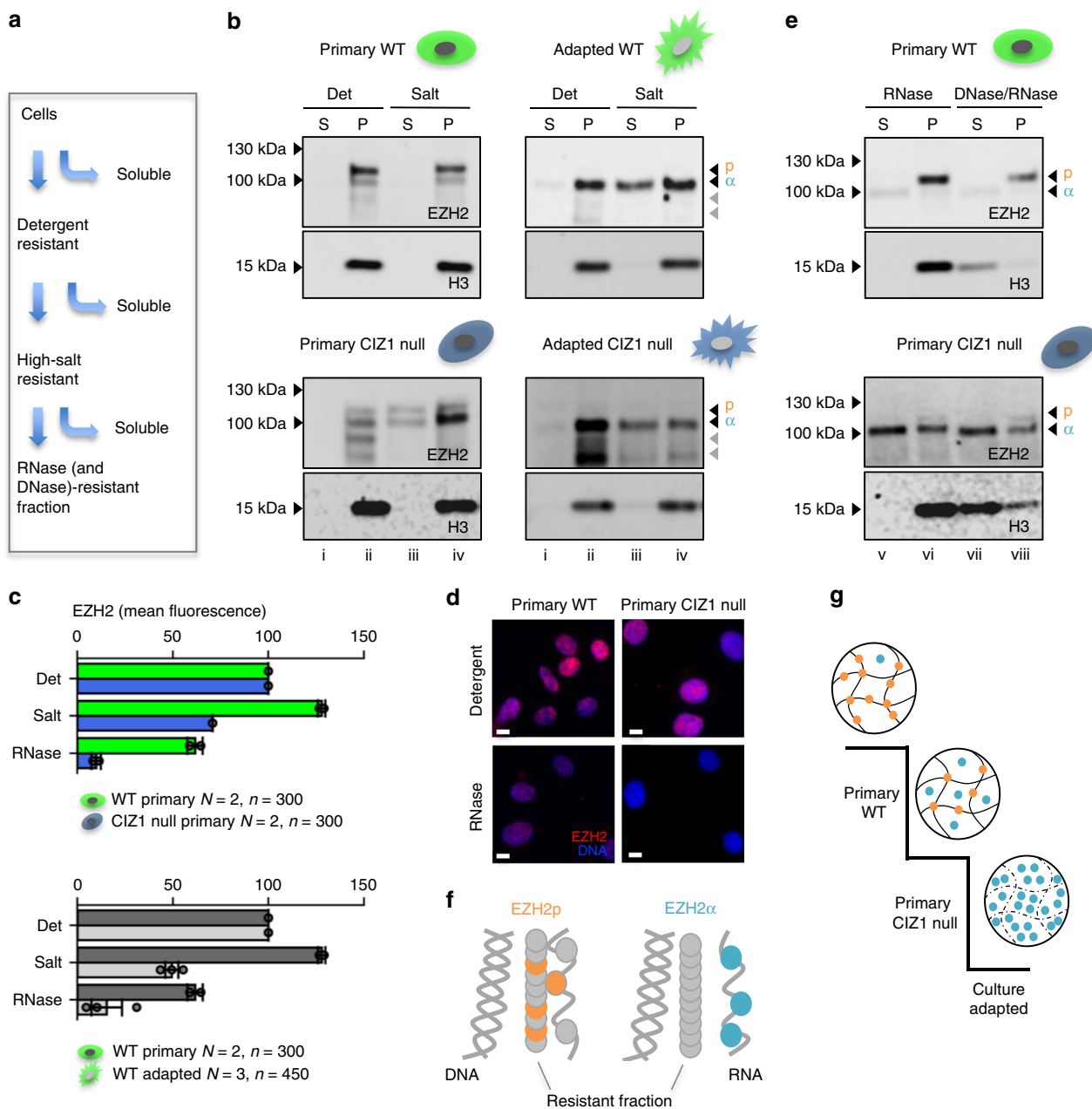


Fig. 6 Loss of CIZ1 drives changes in EZH2 binding. **a** Schematic showing sequential extraction with detergent-containing buffer (0.1% Triton X100), high-salt buffer (500 mM NaCl), and nucleases (RNase and/or DNase1), detailed in Methods. **b** Representative western blots showing partitioning of EZH2p and EZH2α (antibody D2C9) into soluble (S) and pellet (P) fractions, after incubation with detergent (Det) or with detergent plus 500 mM NaCl (Salt). All fractions within an extraction series are cell-equivalents. Results are shown for primary WT (13.8 p3) and CIZ1 null cells (13.17 p5) and culture-adapted derived populations (p28–29) of each genotype. Fraction numbers i–iv relate to the full serial extraction protocol in Methods. **c** Quantification of EZH2 immuno-stain in two primary cell populations of WT (13.1, 13.8) and CIZ1 null (13.17, 14.2) genotypes after the indicated extraction steps. Histogram shows mean fluorescence intensity normalized to detergent-treated samples, \pm SEM, where $n > 50$ for each genotype and condition, and $N = 2$. Below, the same analysis comparing two WT primary cells and three WT culture-adapted cells (13.1, 14.4, 3T3 cells). Individual data points contributing to the mean are overlaid. **d** Images show EZH2 (red, D2C9) after detergent-extraction and after RNase-extraction. DNA is blue, scale bar is 10 microns. **e** The salt-resistant pellet fraction (iv in **e**) from primary cells, was further partitioned by digestion to solubilize RNA, or both RNA and DNA as indicated, revealing complete nuclease-resistance of EZH2p in WT cells, despite efficient release of histone 3 (H3) into the soluble fraction. In contrast, in CIZ1 null cells (expressing EZH2α) more than half is sensitive to solubilisation with RNase. Fraction numbers v–viii relate to the full serial extraction protocol in Methods. **f** Illustration of the differing modes of attachment of EZH2α (released along with RNA), and EZH2p (immobilized by attachment to RNase and DNase-resistant assemblies). **g** Model illustrating two drivers of the shift away from immobilized EZH2p (primary WT cells) towards the expression of high levels of non-immobilized EZH2α (culture-adapted cells). CIZ1 null primary cells represent an intermediate state

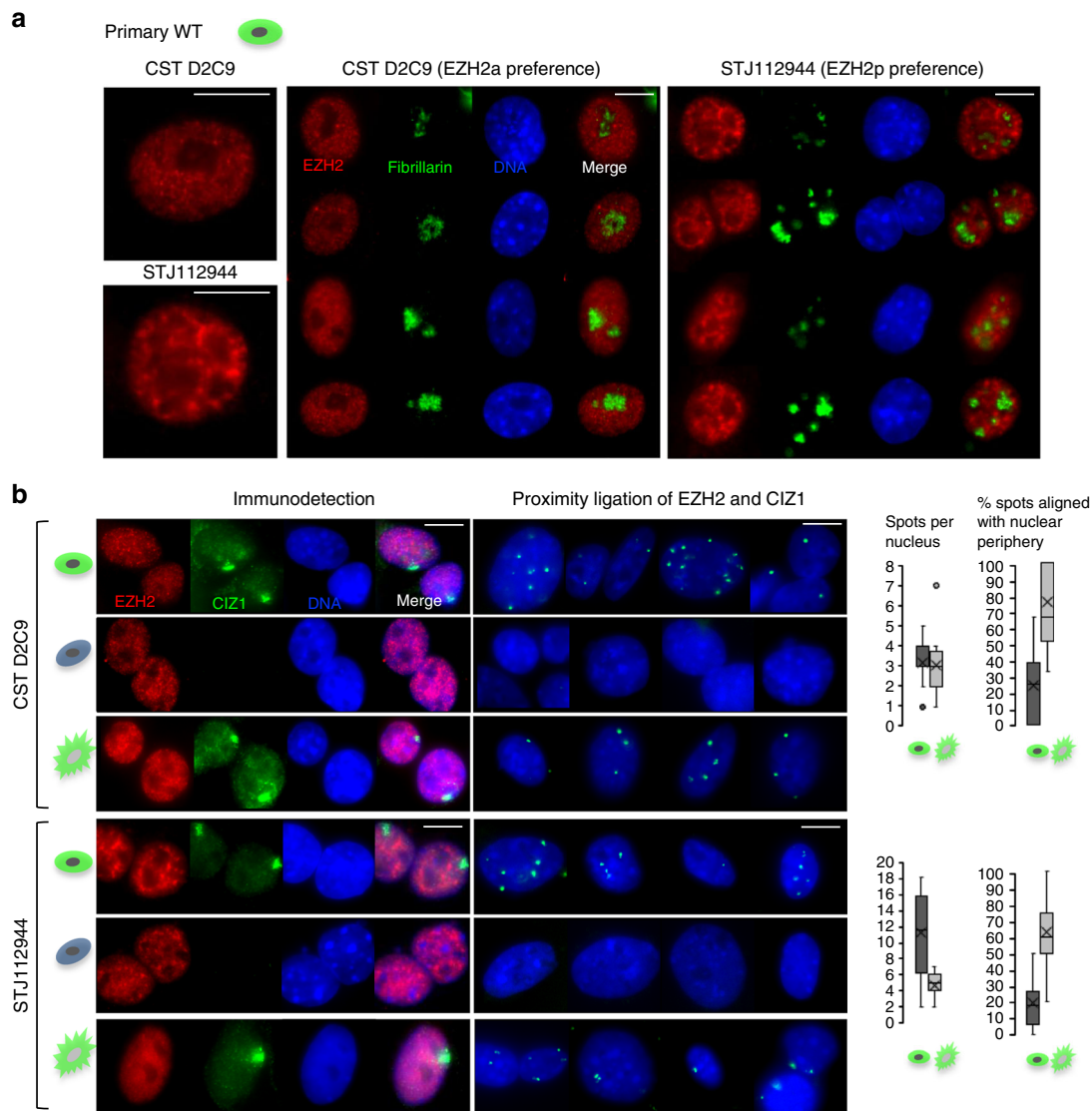


Fig. 7 Differential isoform detection with EZH2 antibodies. **a** Immuno-detection of EZH2 with two different EZH2 antibodies (CSTD2C9 and STJ112944), and of the nucleolar protein fibrillarin in primary mouse embryonic fibroblasts (14.4 p3) showing distinct EZH2 staining patterns, and peri-nucleolar enrichment with STJ112944. Scale bar is 10 microns. **b** Immuno-detection (left) and proximity ligation (middle) of two different EZH2 antibodies (CSTD2C9 and STJ112944) with CIZ1 (hC221) in primary WT (14.4 p3), primary CIZ1 null (13.15 p2) and culture adapted WT (13.8 p23) embryonic fibroblasts. For EZH2 image intensity is differentially adjusted to reveal pattern rather than relative signal intensity. Scale bar is 10 microns. Right, quantification of proximity ligation assay showing the number of spots per nucleus ($n = 20$) and the percentage of spots aligned with the nuclear periphery for the primary and culture adapted WT cells, for both EZH2 antibodies. X represents the mean, the mid-line shows the median, the top and bottom line of the box the 3rd and 1st quartiles, and the whiskers the minimum and maximum values, excluding outliers (points, greater than 1.5 times the interquartile range)

gene expression (summary model, Fig. 8). We propose here that these parameters are linked, and suggest that soluble EZH2 that is released from spatial constraint may be a compensatory response to the loss of chromatin relocation capacity, enabling reinstatement of essential regulatory events, including the observed reinstatement of H3K27me3 at Xi in CIZ1 null cells. By analogy with our previous analysis of the solubility transition of cyclin E in human, mouse and *Xenopus* differentiation models³¹, we further suggest that unconstrained enzyme may represent reversion, during culture, to a less differentiated state.

Notably, in addition to a shift in solubility, culture-adapted cells also experience the elevation of EZH2 protein. Promiscuous activity and gain-of-function mutations in EZH2 are documented in a wide range of human cancers prompting interest in its potential as a therapeutic target³². The data presented here illustrate distinctions between EZH2 in normal cells in the body

compared to EZH2 in transformed cells in culture, in terms of isoform, protein levels and solubility, which could inform design of selective interventions and should be considered in relation to the variable therapeutic response to inhibitors of EZH2³². Moreover, changes in polycomb function as a consequence of CIZ1 disruption, highlights a pathway by which the varied CIZ1-related diseases may arise. The switch in isoform expression, emergence of unconstrained wild enzyme and relaxed control over target selection that we show here, are consistent with the idea that compromised chromatin relocation may be a driver of epigenetic drift.

Moreover, the data strongly reinforce the message that culture-adapted fibroblast cell models do not fully represent the functional state of cells in the body, thereby restricting the application of somatic cell genetic modification strategies that include selection to derive clonal lines.

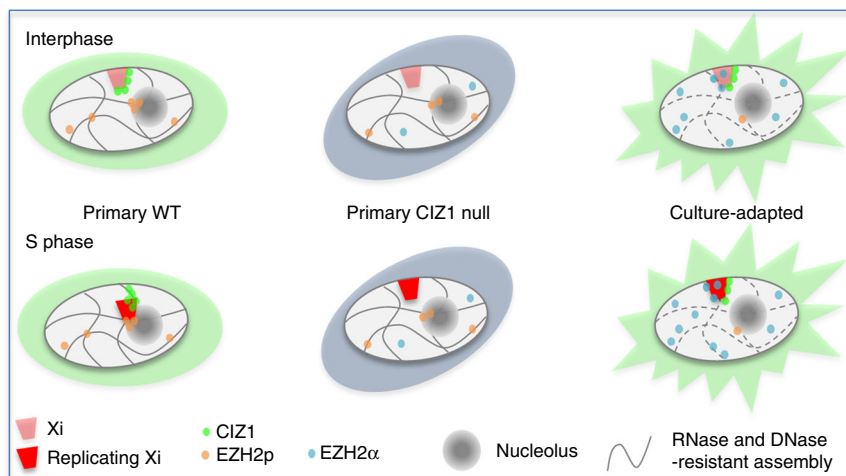


Fig. 8 Summary model. Summary model depicting two alternative pathways by which chromatin and modifying enzymes could meet. Left, anchored and spatially restricted EZH2p (orange) in WT primary cells, and CIZ1-dependent relocation of Xi during replication of Xi. Centre, compromised relocation of Xi in CIZ1 null PEFs and loss of marks from Xi. Right, elevated wild EZH2 α (blue) in culture-adapted cells is depicted diffusing to sites of action. We propose that this compensates for the loss of Xi relocation capability but could lower the stringency with which epigenetic state is maintained in cycling somatic cells

Long-range directional movement of chromatin has been studied in detail using tagged loci and live imaging approaches in CHO cells³³ or transformed human cell lines³⁴. The latter identified distinct nuclear substructures (nucleolus and nuclear periphery) that locally constrained the movement of a subset of loci but did not report evidence for transient relocation between those sites, while the former measured inducible unidirectional migration of an engineered locus from the periphery to the nuclear interior. This study, and others that focus on chromosome territory relocation during DNA repair³⁵ or entry to quiescence³⁶ in early passage human fibroblasts, have begun to implicate the chromatin-bound motor protein nuclear myosin 1 (NM1)³⁷, though little mechanistic detail has yet been uncovered.

The data presented here implicate CIZ1 in a mechanism underpinning Xi relocation in S phase, however a number of observations argue that this function extends beyond Xi to other loci. These are the presence of focal-CIZ1 in both male and female nuclei⁵, similar expression profiles in both male and female differentiating ES cells⁶, and the genome-wide effect of deletion of CIZ1 on polycomb-regulated gene expression. It should be noted that the present analysis does not directly address whether polycomb-regulated genes elsewhere in the genome visit the nucleolus. However, the behaviour of Xi, which can be readily monitored via locally high concentrations of CIZ1, appears to be a valuable and tractable native context in which to unpick the requirements of long-range chromatin relocation events.

Crucially, capacity to relocate Xi was compromised in all of the cell lines we studied, precluding full mechanistic analysis of the role of CIZ1 in cell lines. However, because cell-cycle-dependent recruitment of Xi-CIZ1 to RNA-independent assemblies is retained in these cells, we can infer that any role that this interaction plays in Xi relocation is not sufficient on its own, and that the relocation process is likely compromised at another level. To understand the mechanism by which relocation is normally achieved we now need to (i) define the transient S-phase specific interaction between CIZ1 and the NM (chromatin and RNA-independent assemblies), and (ii) in cells that are competent to support an S phase shift in location of Xi, ask what role this plays in the maintenance of epigenetic landscape in daughter cells. An important caveat is that we cannot yet be certain that the impact on Xi location in CIZ1

null cells is direct or indirect, but the above approaches will help to establish the relationship.

A pivotal role for LNCRNAs in long-range movement of chromatin is suggested by disruption of *Xist*²⁵ and *Firre*³⁸ (which also results in the loss of H3K27me3 from Xi), and by studies unrelated to Xi, which implicate Kcnq1ot1 in perinucleolar targeting of an episomal vector during mid-S phase^{39,40}. Again this was shown to contribute to silencing of linked genes by EZH2⁴¹. Thus, locus-specific relocation as a means of maintaining epigenetic state is an emerging function of LNCRNAs and is supported by independent studies. CIZ1 interacts with *Xist* via its repeat E and is dependent on repeat E for accumulation at Xi^{5,6}, apparently through most of the cell cycle. The S-phase switch in CIZ1 interaction profile leads us to consider whether a primary function of CIZ1 may be to transiently link *Xist*-coated Xi chromatin to machinery that supports Xi transit in S phase, thereby coupling replication to relocation. Such an S-phase-specific function for CIZ1 is consistent with previous analyses which have identified direct and sequential interactions with cyclin E and cyclin A⁴², and regulatory phosphorylation sites which impact on CIZ1s function in DNA replication⁴³, though at present are difficult to reconcile with data which suggest a link with pre-IC conversion⁴³.

While the role of CIZ1 in determining Xi location in primary cells is beginning to emerge, the drivers that lead to the loss of S-phase location change in WT cells in long-term culture and the molecules affected, remain opaque. It could reflect changes in *Xist*²⁵ or *Firre*³⁸, or other chromatin-associated RNAs⁴⁴, or a corrupted NM, or other factor such as NM1. However, all we can safely conclude at this stage is that corruption of the process in which CIZ1 normally functions is a recurrent feature of culture-adapted cells.

In summary, our data are consistent with the model that the selection of EZH2 target chromatin may be achieved by divergent pathways with differing fidelity. (i) A CIZ1-dependent, replication-linked pathway operating in normal WT somatic cells that supports timely and controlled meeting of spatially restricted enzyme and template via transient relocation of chromatin, and (ii) a CIZ1-independent pathway that is not dependent on chromatin relocation. Based on profiling the expression, location and binding characteristics of EZH2, we further hypothesize that release and up-regulation of EZH2 methyltransferase could

compensate for the loss of relocation capability, enabling chromatin maintenance (at the expense of fidelity) in rapidly cycling cells. While the evidence for this last point is currently circumstantial, based on the data presented here, we are able to safely conclude that CIZ1 is part of a mechanism that contributes to accurate maintenance of epigenetic landscape in primary cells, and suggest that this pathway more closely reflects the status of normal differentiated cells in the body.

Methods

Cell isolation, culture and derivation of lines. CIZ1 null mice were generated from C57BL/6 ES clone IST13830B6 (TIGM) harbouring a neomycin resistance gene trap inserted downstream of exon 1. The absence of *Ciz1*/CIZ1 in homozygous progeny was confirmed by qPCR, immunofluorescence and immunoblot⁵. Mouse PEFs were derived from individual embryos at days 13–14 of gestation. All primary and culture adapted MEFs were cultured in DMEM GlutaMAX (Gibco) supplemented with 10% FBS (PAA) and 1% Pen/Strep/Glutamine (Gibco). During rapid proliferation phase, cells were grown to 90% confluence then passaged 1:3 every 2–3 days. Populations of cells cultured up to passage (p) 5 were considered early passage (specified in text), and designated here as primary cells. Experiments were not performed after passage 5, unless to generate culture adapted lines (Supplementary Table 3). To achieve this, beyond passage 5, cells were supplemented with fresh media every 3 days and passaged at 90% confluence (typically once per week as proliferation rate slowed) without dilution (passage ratio 1:1). By passage 10–12 individual colonies of intermediate state cells emerged, that were dissociated and replated as a whole (1:1) until rapidly proliferating culture adapted populations emerged by passage 20 and designated as culture adapted lines. These were routinely grown to 90% confluence then passaged as for primary cells. For inducible cells harbouring transactivator and responder transgenes, the addition of doxycycline to media (5–10 µg/ml) was used to induce GFP-CIZ1 within 6 h. Female 3T3 cell line D001⁴⁵ (a kind gift from Stephen Downes) were grown in the same media.

Ethics. All work with animal models is compliant with UK ethical regulations. Breeding and genetic modification of mice were carried out under UK Home Office license and with the approval of the Animal Welfare and Ethical Review Body at the University of Leeds. Analysis on cells and tissues derived from these mice was carried out with the approval of the Animal Welfare and Ethical Review Body at the University of York.

Sub-nuclear fractionation. All buffers were made in RNase-free water. All reagents were from Sigma unless otherwise stated. All samples are named by the last treatment in the series that they received. Extraction for analysis by western blot⁴⁶ is detailed below, and extraction for analysis by immunofluorescence microscopy⁴⁷ was performed on cells grown on coverslips using the same buffers. Briefly, plates of adherent cells were washed twice in ice-cold PBS then twice in ice-cold cytoskeletal buffer (CSK: 10 mM PIPES/KOH pH 6.8, 100 mM NaCl, 300 mM Sucrose, 1 mM EGTA, 1 mM MgCl₂, 1 mM DTT, 1 cOmplete™ Protease Inhibitor Cocktail per 50 ml) and drained at a shallow angle on ice for 2 min. Cells were scrape harvested and supplemented with 2 mM PMSF before addition of Triton-X-100 to 0.1% and, to samples not subsequently receiving RNase, vanadyl ribonucleoside complex (VRC) to 2.5 mM (NEB). Samples were mixed by pipetting, and after 1 min on ice, centrifuged at 1000 × g. The soluble fraction from this stage was designated detergent (det.) supernatant (S, sample i Fig. 6b). The pellet fraction was either denatured for analysis by SDS-PAGE (det. P, ii) or resuspended in 0.1% Triton-X-100 and 0.5 M NaCl in CSK and incubated on ice for 1 min before second centrifugation at 1000 × g (salt S, iii, and P, iv). Pellets to be further extracted were washed with digestion buffer (40 mM Tris/HCl, 10 mM NaCl, 6 mM MgCl₂, 1 mM CaCl₂, pH 7.9, supplemented 1/500 with RNaseOUT for RNase-free samples), before resuspension in fresh digestion buffer. Samples were then incubated at 30 °C for 1 h with gentle agitation in the presence of DNase1 (Roche 04716728001) at 0.3 U/µl, RNase (Roche 11119915001) 0.5 U/µl, or both DNase1 and RNase, or no enzyme (mock). Before final centrifugation, reactions were supplemented to 0.5 M NaCl for 5 min, then separated to yield pellets (P, vi and viii) and supernatants (S, v and vii). Pellets were resuspended in 1× denaturing buffer (2% SDS, 15% glycerol, 1.7% betamercaptoethanol, 75 mM Tris pH 6.8 with bromophenol blue), and supernatants supplemented with 4× denaturing buffer and heated to 95 °C with repeated vortexing to shear remaining nucleic acid. For extraction for analysis by immunofluorescence, coverslips were incubated with CSK supplemented with 0.1% Triton-X-100 and 2.5 mM VRC (CSK-D) for 1 min. Following CSK-D removal coverslips were either fixed with 4% paraformaldehyde (detergent sample) or incubated with CSK-D with 0.5 M NaCl (CSK-DS) for 1 min. For coverslips to be treated with RNase, VRC was left out from this step onwards. Coverslips were washed twice for 1 min with digestion buffer (supplemented with 2.5 mM VRC for non-RNase samples) before incubation for 1 h at 37 °C in digestion buffer with 2.5 mM VRC (High Salt sample), digestion buffer supplemented with 2.5 mM VRC and DNase1 at 0.45 U/µl (DNase sample) or digestion buffer supplemented with RNase at 0.75 U/µl (RNase sample). Coverslips were incubated for 1 min with CSK-

DS (supplemented with 2.5 mM VRC for non-RNase samples) before fixation with 4% paraformaldehyde.

Antibodies and detection protocols. For western blots, adherent cells were either extracted as above or, after washing in ice-cold PBS, scrape harvested into denaturing buffer with fresh 2 mM PMSF, to generate whole cell lysates. Samples were heated to 95 °C and separated through 4–15% gradient gels (Bio-Rad) then transferred onto nitrocellulose membranes using the iBlot system (Invitrogen). Blots were blocked in either 5% low-fat milk in PBS with 0.1% Tween-20, or 5% BSA in TBS with 0.05% Tween-20, depending on antibody, then incubated with primary antibodies overnight at 4 °C with gentle agitation. Blots were washed and probed with HRP-conjugated anti-species secondary antibodies for 1 h at room temperature (Jackson Immunochemicals 115-035-174 and 211-032-171), and imaged using EZ-ECL (Biological Industries) with a Syngene PXi chemiluminescence imaging system. Band intensities were quantified using GeneSys 4.03.05.0 software. Uncropped blots for the images used in main figures are shown in Supplementary Fig. 5. For immunofluorescence microscopy, cells grown on glass coverslips were washed with PBS before fixation in 4% paraformaldehyde, or where indicated bathed in CSK with 0.1% Triton-X-100 for 1 min prior to fixation. After fixation cells were rinsed twice with PBS then incubated for 15 min in BSA Antibody buffer (0.02% SDS, 0.1% TX100, 10 mg/ml nuclease-free BSA in PBS), followed by 2 h at 37 °C with primary antibody (in BSA antibody buffer). After three washes in the same buffer, anti-species secondary antibodies (Alexa Fluor 488 or 568) were applied for 1 h, followed by three further washes, and mounting in VectorShield with DAPI (Vector Labs). All antibodies including dilutions are listed in Supplementary Table 4.

Proximity ligation assay. Primary WT, primary CIZ1 null and culture adapted WT murine embryonic fibroblast cells were grown on glass coverslips and washed with 0.1% Triton-X-100 containing PBS prior to fixation with 4% paraformaldehyde for 15 min. Coverslips were blocked and incubated with primary antibodies as usual before washing with Duolink® In Situ Wash Buffer A and the proximity ligation assay detailed in Duolink® PLA fluorescence protocol (Sigma), using species-specific secondary antibody probes and green detection reagents. In brief, following standard primary antibody incubation as detailed in antibodies and detection protocols, coverslips were washed twice for 5 min with Wash Buffer A and incubated with species-specific PLUS and MINUS PLA probes in the provided antibody diluent for 1 h at 37 °C. Coverslips were washed twice for 5 min with Wash Buffer A before incubation for 30 min at 37 °C with 1× Ligation Buffer supplemented with DNA Ligase. Coverslips were washed twice for 5 min with Wash Buffer A followed by incubation for 100 min at 37 °C in Amplification buffer supplemented with DNA polymerase. Coverslips were washed twice for 10 min with Wash Buffer B before a final 1 min wash in 0.01× Wash Buffer B and mounting in Vector Shield with DAPI (Vector Labs).

Imaging. Fluorescence images were captured using a Zeiss Axiovert 200M fitted with a 63×/1.40 Plan-Apochromat objective and Zeiss filter sets 2, 10, 15 (G365 FT395 LP420, BP450-490 FT510 BP515-565, BP546/12 FT580 LP590), using AxioCam 506 mono and Axiovision image acquisition software (SE64 release 4.9.1). Where changes in fluorescence intensity are quantified across an extraction series, coverslips were imaged as a set with all images for each filter set captured with the same exposure time. Images were saved at 1499 by 1205 pixels in tagged image file format for downstream analysis. For EZH2 at least 50 individual nuclei from unmodified images were quantified for each cell population and each extraction condition, using Fiji⁴⁸. Where indicated, EZH2 levels across an extraction series were normalized to the detergent-treated sample (given an arbitrary value of 100) to enable comparison between cell lines. For presentation, images were enhanced using Adobe Photoshop CS4 or Affinity Photo 1.5.2, maintaining identical manipulations across extraction series so that image intensities reflect actual relationships (unless specifically stated otherwise). All quantification of image intensity was carried out prior to manipulation. For the presentation of images illustrating positional, rather than intensity information, images were not necessarily modified identically.

S phase labelling. 5-Ethynyl-2'-deoxyuridine (EdU, 10 µM) was added to adherent cells on coverslips at ~70% confluence for a 30 min pulse period under standard growth conditions. For pulse/chase experiments, coverslips were transferred to warm PBS then into fresh media (without EdU) for 30–60 min. To visualize newly synthesized DNA, coverslips were washed briefly in CSK with 0.1% Triton-X-100, or subjected to extraction up to the RNase treatment step before fixation with 4% paraformaldehyde for 15 min. Coverslips were then washed in PBS and EdU detected using the Click-iT® EdU Alexa Fluor® 594 kit (ThermoFisher), as recommended. Briefly, coverslips were blocked with 3% BSA before incubation in a light-proof humidified chamber with Click-iT® reaction cocktail for 60 min. For dual staining (e.g. CIZ1 or fibrillarlin), coverslips were first incubated with primary antibody as described under antibodies and detection protocols before blocking with 3% BSA. Anti-species secondary antibody diluted in Click-iT® reaction buffer was included in the EdU detection step. Coverslips were then washed and mounted using VectorShield with DAPI (Vector Labs).

Positional analysis of Xi. Typically, 30–50% of cells in a rapidly cycling population incorporate EdU during a 30 min pulse labelling period. Of these, 6–16% have one or two compact discrete patches of incorporated EdU against an early or mid S-phase replication pattern, verified as Xi by co-staining with CIZ1. The location of EdU-labelled Xi was scored in two ways. (i) By proximity to the nucleolar marker fibrillarin and the nuclear envelope, generating four categories illustrated in Supplementary Fig. 1. Xis were classified as peripheral if any part of its territory or individual foci that are part of a cluster were not resolvable from the nuclear perimeter, confirmed by comparison with DAPI. Xis were scored as nucleolar if any part of the territory or foci were unresolvable from fibrillarin stain. Xi which meets both of these criteria were classified as peripheral/nucleolar, or neither if both of the criteria were not met. (ii) A binary classification, based on the criteria above (Supplementary Fig. 1), was used to gather time-resolved data points. Xis were scored as internal if all of the territory, including all foci, was resolvable from the nuclear periphery, and peripheral if these criteria were not met. For nuclei with two EdU-labelled Xis both were classified individually. Nuclei with more than two EdU-labelled Xis were excluded from the analysis. At least 100 labelled nuclei were imaged per coverslip, with at least 3 replicate coverslips, and three biological replicate cell populations (from independent embryos) of each genotype. Results in Fig. 4a show compromised of Xi relocation in CIZ1 null cells, and results in Fig. 1f show a gradual diminishing of Xi relocation in WT adapted lines (evidenced by lack of change in the proportion of peripheral Xis between pulse and chase). However, in some adapted cells we were not able to score location because of the absence of a clearly defined CIZ1 mark and also apparent incoherent replication of Xi. In these cells, replicating Xi was no longer detectable as a patch of dense EdU against a mid-S phase nucleus, and neither was it identifiable against early or late replicating nuclei because CIZ1 was dispersed or absent. This most likely reflects instability of Xi in transforming cells, illustrated previously in the context of breast cancer⁴⁹. These cell populations were typically late passage (p20 or more) and were excluded from the analysis.

Statistical analysis. Experiments were designed to use the minimum number of animals while achieving statistically valid data and include two types of analysis. For multivariate data (lists of genes or proteins that are changed in one genotype compared to another) three biological replicates (independent PEF lines) enables calculation of average log₂ fold change and *T*-test *P*-values, as well as FDR-adjusted *P*-value for each data point (gene), and is the minimum required for differential expression analysis using Cuffdiff. Randomization is not appropriate, and PEF lines are matched as closely as possible. For single parameter endpoints such as analysis the requirements of Xi relocation, the variable is sampling time, and the output one of two positions, with typical effect size of 3 and analysis using two sample *t*-test in Excel. Unless indicated, data is represented as means with SEM. Comparison of the differing Xi relocation behaviour of WT and CIZ1 null cells was evaluated in R using a two-way ANOVA. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Avoidance of bias is achieved by independent, blinded analysis of archived images, for tests with optimised parameters. For example, if 1 mg/ml etoposide in the 30 min EdU chase period influences return of Xi to the nuclear periphery after replication, data (images) from replicate experiments will be archived, coded and blind scored by independent workers.

Transcriptome analysis. Primary (before passage 5) and culture-adapted derivative cell lines of murine embryonic fibroblasts 13.1, 13.8 and 14.4 (female WT), and 13.15, 13.17, 14.2 and 14.19 (female CIZ1 null) were grown to 80% confluence. RNA was extracted with TRIzol (Ambion 15596-026) following manufacturer's instructions. Briefly, adherent cells grown to 80% confluence were washed twice with PBS, drained on a shallow angle for 2 min and excess PBS removed. 1 ml of TRIzol was added per 28 cm² and incubated for 3–5 min at room temperature with periodic agitation. Lysates were collected in clean Eppendorf tubes. Chloroform was added to a ratio of 1:5 and lysates were shaken vigorously for 15 s before incubation at room temperature for 3 min before centrifugation at 12,000 × *g* for 15 min. The aqueous phase was transferred into a clean Eppendorf and mixed with equal volume of isopropanol through gentle inversion and incubated at room temperature for 10 min, followed by a 10 min centrifugation at 12,000 × *g*. Supernatant was removed, and the RNA pellet washed with an equal volume of 75% ethanol to the volume of TRIzol used to extract the RNA. Sample was centrifuged for 5 min at 12,000 × *g* and the supernatant removed. RNA pellet was resuspended in nuclease-free water. Isolated RNA was then treated with DNase (Roche 04716728001), before quality analysis by agarose gel, NanoDrop spectrophotometer and Agilent 2100 Bioanalyzer. Libraries were prepared with NEBNext® Ultra™ RNA Library Prep Kit for Illumina®, and enriched for mRNA using NEBNext Poly(A) mRNA Magnetic Isolation Module, which is optimized for the production of libraries with 250–400 bp inserts. Enriched mRNA was fragmented by heating to 95 °C for 12 min, cDNA synthesised from random primers, followed by end repair, dA-tailing, adaptor ligation and PCR enrichment. Libraries were sequenced at the Leeds Institute for Molecular Medicine (LIMM) using Illumina 3000 system, using paired-end sequencing to generate ~50 million reads per sample. Sequence reads were trimmed to remove any adapter sequences using Cutadapt version 1.8.3⁵⁰ then aligned to version GRCm38 of the mouse genome using HISAT2⁵¹. Transcriptomes were assembled and gene expression quantified using the Tuxedo pipeline (version 2.2.1)⁵², from which predicted splice variant assemblies were extracted, including those that map to the EZH2 locus 6:47530040–47595351 (Supplementary Fig. 3,

Supplementary data 6) and other PRC subunits (Supplementary Fig. 4). Average expression of each splice variant was calculated from the 3 independent lines for WT and CIZ1 null primary and culture adapted cells. Variants arising from sequencing of the opposite strand were removed. Total expression of the gene was calculated for each cell type along with the percentage of each splice variant. Splice variants contributing less than 5% of the total in all cell types were removed from the analysis to simplify visual representation. Cufflinks was used to assemble transcriptomes for each sample using the GTF annotation file for the GRCm38 mouse genome (C57BL/6), followed by Cuffmerge to merge individual sample transcriptomes. Quantification, normalisation and differential expression were carried out using Cuffquant, Cuffnorm and Cuffdiff, respectively. GSEA^{22,53} was performed in Python 3.6 using one-sided Fisher's Exact tests as part of the SciPy library (v.0.19.0). FDR was controlled using the Benjamini–Hochberg method in the StatsModels library (v.0.8.0), to generate *q* values. As this correction is based on *P*-values and different comparisons have different numbers of *P*-significant expression changes, the threshold at which *q*-significance is reached will vary between comparisons, as reflected in the volcano plots in Fig. 2. Volcano plots were generated in Excel. Heat maps were generated using Spyder (v.3.1.4), accessed via Anaconda Distribution (v.1.6.2), using the pandas, seaborn and matplotlib modules. Transcription units which did not have a numerical value for log₂(fold change) due to mean expression of 0 in one condition were manually removed before generating the plots. Individual fragment counts per kilobase per million (FPKM) were extracted for replicate cell lines to calculate means and SEM.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The custom code used to perform Gene Set Enrichment Analysis is available from authors upon request. All versions of software used are described in transcriptome analysis methods.

Data availability

Transcriptome data is available at GEO repository under accession code [GSE122235](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122235). All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

Received: 21 May 2018 Accepted: 4 December 2018

Published online: 28 January 2019

References

- Brockdorff, N. & Turner, B. M. Dosage compensation in mammals. *Cold Spring Harb. Perspect. Biol.* **7**, a019406 (2015).
- da Rocha, S. T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.* **24**, 197–204 (2017).
- Pintacuda, G. et al. hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell* **68**, 955–969 (2017).
- Almeida, M. et al. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081–1084 (2017).
- Ridings-Figueroa, R. et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* **31**, 876–888 (2017).
- Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* **114**, 10654–10659 (2017).
- Xiao, J., Vemula, S. R. & LeDoux, M. S. Recent advances in the genetics of dystonia. *Curr. Neurol. Neurosci. Rep.* **14**, 462 (2014).
- Dahmcke, C. M., Buchmann-Moller, S., Jensen, N. A. & Mitchelmore, C. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol. Cell Neurosci.* **38**, 589–594 (2008).
- Rahman, F. A., Ainscough, J. F.-X., Copeland, N. & Coverley, D. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor Ciz1. *Hum. Mutat.* **28**, 993–1004 (2007).
- den Hollander, P., Rayala, S. K., Coverley, D. & Kumar, R. Ciz1, a novel DNA-binding coactivator of the estrogen receptor alpha, confers hypersensitivity to estrogen action. *Cancer Res.* **66**, 11021–11029 (2006).
- Higgins, G. et al. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc. Natl Acad. Sci. USA* **109**, E3128–E3135 (2012).
- Wu, J., Lei, L., Gu, D., Liu, H. & Wang, S. CIZ1 is upregulated in hepatocellular carcinoma and promotes the growth and migration of the cancer cells. *Tumour Biol.* **37**, 4735–4742 (2016).

13. Yin, J. et al. CIZ1 regulates the proliferation, cycle distribution and colony formation of RKO human colorectal cancer cells. *Mol. Med. Rep.* **8**, 1630–1634 (2013).
14. Nishibe, R. et al. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. *FEBS Lett.* **587**, 1529–1535 (2013).
15. Greaves, E. A., Copeland, N. A., Coverley, D. & Ainscough, J. F. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. *J. Cell Sci.* **125**, 2466–2477 (2012).
16. Ainscough, J. F. et al. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. *J. Cell Sci.* **120**, 115–124 (2007).
17. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
18. Casas-Delucchi, C. S. et al. Histone acetylation controls the inactive X chromosome replication dynamics. *Nat. Commun.* **2**, 222 (2011).
19. Koren, A. & McCarroll, S. A. Random replication of the inactive X chromosome. *Genome Res.* **24**, 64–69 (2014).
20. Chen, C. K. et al. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468–472 (2016).
21. Barr, M. L. & Bertram, E. G. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**, 676 (1949).
22. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
23. Nuytten, M. et al. The transcriptional repressor NIPPI1 is an essential player in EZH2-mediated gene silencing. *Oncogene* **27**, 1449–1460 (2008).
24. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell Biol.* **27**, 3769–3779 (2007).
25. Zhang, L. F., Huynh, K. D. & Lee, J. T. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* **129**, 693–706 (2007).
26. Chen, S., Birve, A. & Rasmuson-Lestander, A. In vivo analysis of Drosophila SU(Z)12 function. *Mol. Genet. Genomics* **279**, 159–170 (2008).
27. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
28. Grzenda, A. et al. Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. *Epigenetics Chromatin* **6**, 3 (2013).
29. Beltran, M. et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
30. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).
31. Munkley, J. et al. Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells. *Nucleic Acids Res.* **39**, 2671–2677 (2011).
32. Kim, K. H. & Roberts, C. W. Targeting EZH2 in cancer. *Nat. Med.* **22**, 128–134 (2016).
33. Chuang, C. H. et al. Long-range directional movement of an interphase chromosome site. *Curr. Biol.* **16**, 825–831 (2006).
34. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439–445 (2002).
35. Kulashreshtha, M., Mehta, I. S., Kumar, P. & Rao, B. J. Chromosome territory relocation during DNA repair requires nuclear myosin I recruitment to chromatin mediated by Upsilon-H2AX signaling. *Nucleic Acids Res.* **44**, 8272–8291 (2016).
36. Mehta, I. S., Amira, M., Harvey, A. J. & Bridger, J. M. Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol.* **11**, R5 (2010).
37. Pestic-Dragovich, L. et al. A myosin I isoform in the nucleus. *Science* **290**, 337–341 (2000).
38. Yang, F. et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16**, 52 (2015).
39. Mohammad, F. et al. Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell Biol.* **28**, 3713–3728 (2008).
40. Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
41. Terranova, R. et al. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* **15**, 668–679 (2008).
42. Copeland, N. A., Sercombe, H. E., Ainscough, J. F. & Coverley, D. Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. *J. Cell Sci.* **123**, 1108–1115 (2010).
43. Copeland, N. A., Sercombe, H. E., Wilson, R. H. & Coverley, D. Cyclin-A-CDK2-mediated phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. *J. Cell Sci.* **128**, 1518–1527 (2015).
44. Nozawa, R. S. et al. SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAs. *Cell* **169**, 1214–1227 (2017).
45. Coverley, D., Laman, H. & Laskey, R. A. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. *Nat. Cell Biol.* **4**, 523–528 (2002).
46. Wilson, R. H., Hesketh, E. L. & Coverley, D. Preparation of the nuclear matrix for parallel microscopy and biochemical analyses. *Cold Spring Harb. Protoc.* **2016**, prot083758 (2016).
47. Stewart, E. R. & Coverley, D. Visualization of hidden epitopes at the inactive X chromosome. *Methods Mol. Biol.* **1861**, 103–112 (2018).
48. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
49. Chaligne, R. et al. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Res.* **25**, 488–503 (2015).
50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Bioinformatics Action* **17**, 10–12 (2012).
51. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
52. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
53. Mootha, V. K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

Acknowledgements

The authors are grateful to Bob White and Neil Brockdorff for critical comments, and Sally James of the University of York Genomics facility for library preparation. This work was supported by BBSRC doctoral training funds to E.R.S. (BB/M011151/1), a University of York Facilities award, and MRC award MR/R008981/1.

Author contributions

E.R.S., R.M.L.T., V.S. and R.R.-F. carried out the investigations. J.F.X.A. designed and generated CIZ1 animal models and isolated primary cells. K.N. and P.D.A. supported transcriptome analysis. E.R.S., R.M.L.T., J.F.X.A. and D.C. designed the experiments. E.R.S. and D.C. wrote the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-08072-2>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2019

Appendix E

Example python scripts used to perform bioinformatic analysis. Author of each script is stated.

Gene Set Enrichment Analysis – written by Katherine Newling

```
import pandas as pd
from statsmodels.stats.multitest import multipletests
from scipy import stats
import argparse

parser=argparse.ArgumentParser(description=""
    -s sets_file
    -d differentially_expressed_genes_file
    -o output_file_name
    "")

parser.add_argument('-s', '--sets_file', type=str, required=True)
parser.add_argument('-d', '--differentially_expressed_genes_file', type=str, required=True)
parser.add_argument('-o', '--output_file_name', type=str, required=True)

args = parser.parse_args()

data = pd.read_table(args.sets_file)

d = pd.read_table(args.differentially_expressed_genes_file)

for i in d.index:
    d["gene"][i] = d["gene"][i].upper()
degs = d["gene"]
degs = set(degs)

total_genes = 45956

def check_if_float(item):
    return type(item) is not float

tmpfile = open("GSEA_sets_tmp.txt", "w")

tmpfile.write ('Set_name' + '\t' + 'Set_size' + '\t' + 'Overlap_size' + '\t' + 'p-value' + '\t' + 'q-value(BH 0.1)' + '\n')

for col in data:
    current_set = set(data[col])
    current_set = set(filter(check_if_float, current_set))
    current_set_size = len(current_set)
    overlap = degs.intersection(current_set)
    overlap_size = len(overlap)
    not_overlapping_size = len(degs) - overlap_size
    not_in_set_size = total_genes - current_set_size
    oddsratio, pvalue =
stats.fisher_exact([[overlap_size,not_overlapping_size],[current_set_size,not_in_set_size]], alternative='greater')
    print (col, overlap_size, not_overlapping_size, current_set_size, not_in_set_size, pvalue, oddsratio)
    tmpfile.write (col + '\t' + str(current_set_size) + '\t' + str(overlap_size) + '\t' + str(pvalue) + '\t' + '\n')

tmpfile.close()

sets = pd.read_table("GSEA_sets_tmp.txt")

sets["q-value(BH 0.1)"] = multipletests(sets["p-value"], 0.1, 'fdr_bh')[1]

sets.to_csv(args.output_file_name, sep = '\t')
```

Generation of heat maps – written by Emma Stewart

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.colors import LinearSegmentedColormap

x = open("Heatmap_data.txt", 'r')

df = pd.read_table("Heatmap_data.txt")

df['gene'] = df['gene'].str.upper()

#Drop columns not needed

df = df.drop('column_name', axis = 1)

#Add extra column 'log_foldchange' which will only copy across the log fold change in the data file if the significance
of the change is greater than a set parameter, otherwise will assume is 0 so will show as black.

df['log_foldchange'] = df['data_log2foldchange']
df['log_foldchange'][df['q_value'] > 0.05] = 0.00
df['log_foldchange'][df['q_value'] < 0.05] = df['data_log2foldchange']

#Drop columns not required in in heatmap

df = df.drop('data_log2foldchange', axis = 1)
df = df.drop('q_value', axis = 1)

#Index column want to plot heatmap by

df = df.sort_values(by=['log_foldchange'], ascending=False)
df = df.set_index('gene_id')

#Makes heatmap

plt.figure(figsize=(x,y))
cmap = LinearSegmentedColormap.from_list('mycmap', ['blue', 'black', 'red'])
sns.set(font_scale= 0.5)
b = sns.heatmap(df, vmin = -n , vmax = n , cmap = cmap, rasterized=True)
plt.yticks(rotation=0)
b.set_xlabel("x_label_name",font_size=10)
b.set_ylabel("y_label_name",font_size=10)
b.tick_params(labelsize= 0.3)
plt.savefig("/filepath/file_name.png", dpi=500)
```

#Sets size of the figure
#Sets blue to red scale bar
#Sets scale bar range
#Sets x label parameters
#Sets y label parameters
#Sets location of saved file and resolution

Comparing gene lists – written by Emma Stewart

This script reads in two txt files of differentially expressed genes from two comparisons and identifies which genes are common to both files and those which differ, with an output of 3 txt files.

```
import pandas as pd
import argparse

#Assign in configure panel. d1 and d2 are the differentially expressed genes files to be compared
#o1, o2 and oBoth are output files showing genes only significantly changed in d1 (o1), d2 (o2) or both (oBoth)

parser=argparse.ArgumentParser(description=""
    -d1 differentially_expressed_genes_file_1
    -d2 differentially_expressed_genes_file_2
    -o1 output_file_name_1
    -o2 output_file_name_2
    -oBoth output_file_name_Both
    "")

parser.add_argument('-d1', '--differentially_expressed_genes_file_1', type=str, required=True)
parser.add_argument('-d2', '--differentially_expressed_genes_file_2', type=str, required=True)
parser.add_argument('-o1', '--output_file_name_1', type=str, required=True)
parser.add_argument('-o2', '--output_file_name_2', type=str, required=True)
parser.add_argument('-oBoth', '--output_file_name_Both', type=str, required=True)

args = parser.parse_args()

#Configure output files with the desired heading

output_Both = open(args.output_file_name_Both, 'w')
output_1 = open(args.output_file_name_1, 'w')
output_2 = open(args.output_file_name_2, 'w')
output_Both.write("Heading" + "\n")
output_1.write("Heading " + "\n")
output_2.write("Heading " + "\n")

#Read in differentially expressed genes from d1

diff_1 = pd.read_table(args.differentially_expressed_genes_file_1)

for i in diff_1.index:
    diff_1["gene"][i] = diff_1["gene"][i].upper()
degs_diff_1 = diff_1["gene"]
degs_diff_1 = set(degs_diff_1)

#Read in differentially expressed genes from d2

diff_2 = pd.read_table(args.differentially_expressed_genes_file_2)

for i in diff_2.index:
    diff_2["gene"][i] = diff_2["gene"][i].upper()
degs_diff_2 = diff_2["gene"]
degs_diff_2 = set(degs_diff_2)

#Sorts genes into the output files

overlap = degs_diff_1.intersection(degs_diff_2)
for item in overlap:
    output_Both.write(item + "\n")

diff_1_only = list(set(degs_diff_1) - set(degs_diff_2))
for item in diff_1_only:
    output_1.write(item + "\n")

diff_2_only = list(set(degs_diff_2) - set(degs_diff_1))
for item in diff_2_only:
    output_2.write(item + "\n")

output_Both.close()
output_1.close()
output_2.close()
```

Extracting expression data – written by Emma Stewart

This script locates genes within a set, for example a GSEA set, in a differentially expressed genes file. This allows investigation of how genes within a set are behaving on an individual gene basis.

```
import pandas as pd
import argparse

parser=argparse.ArgumentParser(description=""
    -s sets_file_1
    -d differentially_expressed_genes_file
    -o output_file
    "")

#Assign in configure panel. 's' is the sets file, e.g. a file containing the genes within a GSEA gene set. 'd' is the
differentially expressed genes file. The output file 'o' will contain the expression of all the genes in 's' that can be
located in file 'd'.

parser.add_argument('-s', '--sets_file_1', type=str, required=True)
parser.add_argument('-d', '--differentially_expressed_genes_file', type=str, required=True)
parser.add_argument('-o', '--output_file', type=str, required=True)

args = parser.parse_args()

#Generate a list of genes within the set file

data = pd.read_table(args.sets_file_1)
for i in data.index:
    data["gene"][i] = data["gene"][i].upper()
degs_data = data["gene"]
degs_data = set(degs_data)
print(degs_data)

set_size = len(degs_data)
print(set_size)

#Configure output files with the headings

output = open(args.output_file, 'w')
output.write("Heading_1" + "\t" + "Heading_2" + "\t" + "Heading_3" + "\t" + "Heading_4" + "\t" + "Heading_5" + "\t" +
"Heading_6" + "\t" + "Heading_7" + "\n")

#Identifies genes in both data sets and writes to output

diff = pd.read_table(args.differentially_expressed_genes_file)

diff["gene"] = diff["gene"].str.upper()

for i in diff.index:
    gene_name = diff["gene"][i]
    gene_name_s = gene_name.split(",")
    for gene in gene_name_s:
        if gene in degs_data:
            output.write(str(diff["Column_1"][i]) + "\t" + str(diff["Column_2"][i]) + "\t" + str(diff["Column_3"][i]) + "\t" +
str(diff["Column_4"][i]) + "\t" + str(diff["Column_5"][i]) + "\t" + str(diff["Column_6"][i]) + "\t" + str(diff["Column_7"][i]) +
"\n")

output.close()
```

Appendix F

Sequences of EZH2 splice variants 'TCONS_00153598' and 'TCONS_00153599'.

'TCONS_00153598' matches UniProt Q61188 and is suspected to be EZH2 α , which is prevalent in culture adapted cells. 'TCONS_00153599' has a 5' extension on 'TCONS_00153598', part of which has been previously reported as a 5'UTR. This isoform is suspected to produce EZH2p which is prevalent in primary cells. The ATG which encodes the start codon of UniProt Q61188 is shown in bold and underlined in both sequences.

>TCONS_00153598

GGAATAATCATGGGCCAGACTGGGAAGAAATCTGAGAAGGGACCGGTTTGTGGCG
GAAGCGTGTAATAATCAGAGTACATGAGACTGAGACAGCTCAAGAGGTTTCAGAAGAG
CTGATGAAGTAAAGACTATGTTTAGTTCCAATCGTCAGAAAATTTTGGAAAGAACT
GAAACCTTAAACCAAGAGTGAAGCAGCGGAGGATACAGCCTGTGCACATCATGAC
TTCTGTGAGCTCATTGCGCGGGACTAGGGAGTGTTTCAGTCACCAGTGACTTGGATT
TTCCAGCACAAAGTCATCCCGTTAAAGACCCTGAATGCAGTCGCCTCGGTGCCTATA
ATGTACTCTTGGTCGCCCTTACAACAGAATTTTATGGTGGAAAGACGAAACTGTTTT
ACATAACATTCCCTTATATGGGGGATGAAGTTCTGGATCAGGATGGCACTTTCATTG
AAGAATAATAAAAAATTTATGATGGAAAAGTGCATGGTGACAGAGAATGTGGATTT
ATAAATGATGAAATTTTTGTGGAGTTGGTAAATGCTCTTGGTCAATATAATGATGA
TGATGATGACGATGATGGAGATGATCCAGATGAAAGAGAAGAAAAACAGAAAGATC
TAGAGGATAATCGAGATGATAAAGAACTTGCCCACCTCGGAAATTTCTGCTGAT
AAAATATTTGAAGCCATTTCTCAATGTTTCCAGATAAGGGCACCGCAGAAGAACT
GAAAGAAAAATATAAAGAACTCACGGAGCAGCAGCTCCCAGGTGCTCTGCCTCCTG
AATGTACTCCAAACATCGATGGACCAATGCCAAATCTGTTTCAGAGGGAGCAAAGC
TTGCATTCATTTTCATACGCTCTTCTGTGCGACGATGTTTTAAGTATGACTGCTTCT
ACATCCCTTCCATGCAACACCCAACACATATAAGAGGAAGAACACAGAAACAGCTT
TGGACAACAAGCCTTGTGGACCACAGTGTTACCAGCATCTGGAGGGAGCTAAGGAG
TTTGCTGCTGCTCTTACTGCTGAGCGTATAAAGACACCACCTAAACGCCCAGGGG
CCGCAGAAGAGGAAGACTTCCGAATAACAGTAGCAGACCCAGCACCCCCACCATCA
GTGTGCTGGAGTCAAAGGATACAGACAGTGACAGAGAAGCAGGGACTGAAACTGGG
GGAGAGAACAATGATAAAGAAGAAGAAGAGAAAAAAGATGAGACGTCCAGCTCCTC
TGAAGCAAATTTCTCGGTGTCAAACACCAATAAAGATGAAGCCAAATATTGAACCTC
CTGAGAAATGTGGAGTGGAGTGGTGCTGAAGCCTCCATGTTTAGAGTCCTCATTGGT
ACTTACTACGATAACTTTTTGTGCCATTGCTAGGCTAATTGGGACCAAAACATGTAG
ACAGGTGTATGAGTTTAGAGTCAAGGAGTCCAGTATCATAGCACCTGTTCCCACTG
AGGATGTAGACACTCCTCCAAGAAAGAAGAAAAGGAAACATCGGTTGTGGGCTGCA
CACTGCAGAAAGATACAACTGAAAAGGACGGCTCCTCTAACCATGTTTTACAACATA
TCAACCCTGTGACCATCCACGGCAGCCTTGTGACAGTTCGTGCCCTTGTGTGATAG
CACAAAATTTTTGTGAAAAGTTTTGTCAATGTAGTTCAGAGTGTCAAACCGCTTT
CCTGGATGTGGTGCAAAGCACAATGCAACACCAAACAGTGTCCATGCTACCTGGC
TGTCCGAGAGTGTGACCCTGACCTCTGTCTCACGTGTGGAGCTGCTGACCATTGGG
ACAGTAAAAATGTATCCTGTAAGAACTGTAGCATTTCAGCGGGGCTCTAAAAGCAC
TTACTGCTGGCACCGTCTGATGTGGCAGGCTGGGGCATCTTTATCAAAGATCCTGT
ACAGAAAAATGAATTCATCTCAGAATACTGTGGGGAGATTATTTCTCAGGATGAAG
CAGACAGAAGAGGAAAAGTGTATGACAAATACATGTGCAGCTTCTGTTCAACTTG
AACAAATGATTTTTGTGGTGGATGCAACCCGAAAGGGCAACAAAATTCGTTTTGCTAA
TCATTCAGTAAATCCAACTGCTATGCAAAAGTTATGATGGTTAATGGTGACCACA
GGATAGGCATCTTTGCTAAGAGGGCTATCCAGACTGGTGAAGAGTTGTTTTTTGAT
TACAGATACAGCCAGGCTGATGCCCTGAAGTATGTGGGCATCGAACGAGAAATGGA
AATCCCTTGACATCTACTACCTCTTCCCCTTCTCTGAAACAGCTGCCTTAGCTT
CAGGAACCTTGAGTACTGTGGGCAATTTAGAAAACGGAAATGCAGTTTGAAATTCT
GAATTTGCAAAGTACTGTAACAGTAATTTATAGTAATGTTTTAAAATCAACTTTTT
ATTGCCTTCTCACCAGCTGCAAAGTGTTTTGTACCAGTGAGTTTTTGCAATAATGC
AGTATGGTACATTTTTCAAATTTGAATAAAGAATACTTGAACCTGTTGTTGAATCA
CTTGTCTTAGAACATCACCGACTGATTCTAGAAAGACTATATCTAGTGGGAAAACA
CAACTTTGCTTTGTTTTTCCAGCTTGCTATGAATATTTAATAACCAATAAACATAC
TGGCATAGTGTTCCTAA

>TCONS_00153599

GCCTTTGGCGCTCGGTCCGGTCCGGTCCGACACCCAGTGGGACATCGAAGGCAGTG
GAGTCCCGGGCGGCGGCGGTGGCGGTGGCGGCGGGTTGGGGGCGACGCGCGGGAGAG
GCGCGGGCTGGCGCGCGGGACGAAGAATAATCATGGGCCAGACTGGGAAGAAATCT
GAGAAGGGACCGTTTTGTTGGCGGAAGCGTGTAAAATCAGAGTACATGAGACTGAG
ACAGCTCAAGAGGTTTCAGAAGAGCTGATGAAGTAAAGACTATGTTTAGTTCCAATC
GTCAGAAAATTTTGAAAAGAACTGAAACCTTAAACCAAGAGTGGAAGCAGCGGAGG
ATACAGCCTGTGCACATCATGACTTCTGTGAGCTCATTGCGCGGGACTAGGGAGTG
TTCAGTCACCAGTGACTTGGATTTTCCAGCACAAAGTCATCCCGTTAAAGACCCTGA
ATGCAGTCGCCTCGGTGCCTATAATGTACTCTTGGTCGCCCTTACAACAGAATTTT
ATGGTGGAAAGACGAAACTGTTTTACATAACATTCTTATATGGGGGATGAAGTTCT
GGATCAGGATGGCACTTTCATTGAAGAACTAATAAAAAATTTATGATGGAAAAGTGC
ATGGTGACAGAGAATGTGGATTTATAAATGATGAAATTTTTTGTGGAGTTGGTAAAT
GCTCTTGGTCAATATAATGATGATGATGATGACGATGATGGAGATGATCCAGATGA
AAGAGAAGAAAAACAGAAAGATCTAGAGGATAATCGAGATGATAAAGAACTTGCC
CACCTCGGAAATTTCTGCTGATAAAAATTTGAAGCCATTTCTCAATGTTTCCA
GATAAGGGCACC GCAGAAGAACTGAAAGAAAAATATAAAGAACTCACGGAGCAGCA
GCTCCCAGGTGCTCTGCCTCCTGAATGTACTCCAAACATCGATGGACCAAATGCCA
AATCTGTT CAGAGGGAGCAAAGCTTGCATTCATTT CATAACGCTCTTCTGTCGACGA
TGTTTTAAGTATGACTGCTTCCCTACATCCCTTCCATGCAACACCCAACACATATAA
GAGGAAGAACACAGAAACAGCTTTGGACAACAAGCCTTGTGGACCACAGTGTTACC
AGCATCTGGAGGGAGCTAAGGAGTTTGCTGCTGCTCTTACTGCTGAGCGTATAAAG
ACACCACCTAAACGCCAGGGGGCCGCAGAAGAGGAAGACTTCCGAATAACAGTAG
CAGACCCAGCACCCCCACCATCAGTGTGCTGGAGTCAAAGGATACAGACAGTGACA
GAGAAGCAGGGACTGAAACTGGGGGAGAGAACAATGATAAAGAAGAAGAAGAGAAA
AAAGATGAGACGTCCAGCTCCTCTGAAGCAAATTTCTCGGTGTCAAACACCAATAAA
GATGAAGCCAAATATTGAACCTCCTGAGAATGTGGAGTGGAGTGGTGCTGAAGCCT
CCATGTTTAGAGTCCCTCATTGGTACTTACTACGATAACTTTTGTGCCATTGCTAGG
CTAATTGGGACCAAACATGTAGACAGGTGTATGAGTTTAGAGTCAAGGAGTCCAG
TATCATAGCACCTGTTCCCACTGAGGATGTAGACACTCCTCCAAGAAAGAAGAAAA
GGAAACATCGGTGTGGGCTGCACACTGCAGAAAGATACAACTGAAAAAGGACGGC
TCTCTAACCATGTTTACAACCTATCAACCCTGTGACCATCCACGGCAGCCTTGTGA
CAGTTCGTGCCCTTGTGTGATAGCACAAAATTTTTGTGAAAAGTTTTGTCAATGTA
GTT CAGAGTGTCAAACCGCTTTCCTGGATGTCGGTGCAAAGCACAATGCAACACC
AAACAGTGTCCATGCTACCTGGCTGTCCGAGAGTGTGACCCTGACCTCTGTCTCAC
GTGTGGAGCTGCTGACCATTGGGACAGTAAAAATGTATCCTGTAAGAACTGTAGCA
TTCAGCGGGGCTCTAAAAGCACTTACTGCTGGCACCCTGTGATGTGGCAGGCTGG
GGCATCTTTATCAAAGATCCTGTACAGAAAAATGAATTCATCTCAGAATACTGTGG
GGAGATTATTTCTCAGGATGAAGCAGACAGAAGAGGAAAAGTGTATGACAAATACA
TGTGCAGCTTTCTGTTCAACTTGAACAATGATTTTGTGGTGGATGCAACCCGAAAG
GGCAACAAAATTCGTTTTGCTAATCATT CAGTAAATCCAACTGCTATGCAAAAGT
TATGATGGTTAATGGTGACCACAGGATAGGCATCTTTGCTAAGAGGGCTATCCAGA
CTGGTGAAGAGTTGTTTTTTGATTACAGATACAGCCAGGCTGATGCCCTGAAGTAT
GTGGGCATCGAACGAGAAATGGAATCCCTTGACATCTACTACCTCTTCCCCTTCT
CTCTGAAACAGCTGCCTTAGCTTCAGGAACCTTGAGTACTGTGGGCAATTTAGAAA
ACGGAATGCAGTTTTGAAATTTCTGAAATTTGCAAAGTACTGTAACAGTAATTTATAG
TAATGTTTTTAAAATCAACTTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTTGT
CCAGTGAGTTTTTTGCAATAATGCAGTATGGTACATTTTTTCAAATTTGAATAAAGAA
TACTTGAACCTGTTGTTGAATCACTTGTCTTAGAACATCACCGACTGATTCTAGAA
AGACTATATCTAGTGGGAAAACACAACCTTTGCTTTGTTTTTCCAGCTTGCTATGAA
TATTTAATAACCAATAAACATACTGGCATAAGTGTTCCTAAA

Appendix G

Table of X-linked genes significantly changed in CIZ1 null spleens compared to WT, relating to **(Chapter 6)**.

X-linked genes significantly changed in CIZ1 null spleens compared to WT ($q < 0.05$). $\text{Log}_2(\text{fold change})$ could not be calculated for genes not expressed in one genotype, and are instead represented as NE.

Locus	Gene	$\text{log}_2(\text{fold change})$	p value	q value
X:8271132-8280179	SLC38A5	2.35	5.00E-05	1.59E-03
X:9272624-9344332	GM14863,XK	1.89	5.00E-05	1.59E-03
X:36795650-36798807	SLC25A5	0.98	1.45E-03	0.02
X:42502452-43429126	SH2D1A	1.64	2.50E-04	5.99E-03
X:57212096-57224554	CD40LG	-2.88	1.90E-03	0.03
X:71555917-71560676	HMGB3	2.36	5.00E-05	1.59E-03
X:72825177-72842602	GABRQ	-4.24	1.10E-03	0.02
X:73774404-73778925	SRPK3	-4.30	5.00E-05	1.59E-03
X:82948869-85206141	DMD	-1.66	2.05E-03	0.03
X:93654862-93749951	PCYT1B	2.25	5.00E-05	1.59E-03
X:100619580-100727214	KIF4	2.48	5.00E-05	1.59E-03
X:106187099-106203699	PGK1	0.87	3.50E-03	0.04
X:107060419-107108627	P2RY10	-1.34	4.00E-04	8.59E-03
X:134308083-134362639	CENPI	2.20	5.00E-05	1.59E-03
X:136138995-136142054	BEX4	1.99	1.40E-03	0.02
X:150519275-150643878	ALAS2	1.81	4.00E-04	8.59E-03
X:152004581-152062694	SMC1A	1.07	2.00E-03	0.03
X:153237288-153396946	KLF8	-2.29	2.75E-03	0.04
X:161117192-161258214	SCML2	2.56	1.55E-03	0.02
X:164419592-164434184	PIGA	1.41	5.00E-04	0.01
X:166238910-166388988	GPM6B	-1.28	8.50E-04	0.01
X:7959259-7978071	GATA1	2.22	5.00E-05	1.59E-03
X:8042942-8074760	SUV39H1	1.14	1.70E-03	0.03
X:8136414-8147964	RBM3	0.94	3.10E-03	0.04
X:8238667-8252406	FTSJ1	1.49	3.00E-04	6.85E-03
X:20059539-20291883	SLC9A7	-2.35	2.00E-04	5.05E-03
X:51113494-51206532	MBNL3	1.19	5.50E-04	0.01
X:53344526-53370502	MOSPD1	1.33	1.30E-03	0.02
X:75095853-75131025	MPP1	1.48	3.00E-04	6.85E-03
X:94535473-94542143	MAGED1	-1.31	1.50E-03	0.02
X:96247202-96293438	VSIG4	4.44	1.00E-03	0.02
X:101261375-101269023	GM20489,GM614,IL2RG	-1.10	8.50E-04	0.01
X:101731534-101734269	CXCR3	-1.28	1.65E-03	0.02
X:102141715-102157091	ERCC6L	2.42	5.00E-05	1.59E-03
X:109004471-109013716	HMG5	1.28	5.00E-04	0.01
X:134059114-134221956	XKRX	-3.60	2.75E-03	0.04
X:134537255-134541865	TIMM8A1	1.36	3.10E-03	0.04
X:134971371-135009780	ZMAT1	-1.37	3.80E-03	0.04
X:140539527-140600659	TSC22D3	-1.90	5.00E-05	1.59E-03
X:142851145-142967363	AMMECR1	1.60	1.50E-04	4.01E-03
X:152294814-152327524	KANTR	-1.55	2.55E-03	0.03
X:164070494-164076493	SIAH1B	1.55	3.85E-03	0.05
X:166882411-167102090	GM15232,GM8817	-2.85	1.60E-03	0.02
X:167346226-167382802	PRPS2	1.39	9.00E-04	0.02
X:9639522-9640217	-	NE	7.00E-04	0.01
X:13492030-13492814	-	NE	5.50E-04	0.01
X:38180034-38180609	-	NE	5.00E-05	1.59E-03
X:42340508-42340995	-	NE	5.50E-04	0.01
X:121670841-121671415	-	NE	4.50E-04	0.01
X:150477096-150477349	-	NE	9.00E-04	0.02

Appendix H

A series of tables, labelled 1-8, containing top 20 outputs from gene set enrichment analysis (GSEA). Relates to **Chapter 6**.

Table H.1 The top 20 outputs following GSEA analysis using the MSig Curated Gene Set database for the annotated genes significantly ($q < 0.05$) upregulated upon B (upper, 51 genes) or T (lower, 290 genes) cell activation of CI21 null lymphocytes only (inappropriate upregulation set). Sets which appear in both analyses are highlighted.

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (51)	q-value
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	11	21.57	5.86E-10
ZHANG_TLX_TARGETS_60HR_DN	10	19.61	2.35E-09
FISCHER_DREAM_TARGETS	14	27.45	2.89E-09
ZHANG_TLX_TARGETS_DN	7	13.73	2.18E-08
FEVR_CTNNB1_TARGETS_DN	11	21.57	3.16E-08
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP	11	21.57	3.61E-08
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	7	13.73	2.48E-07
ZHANG_TLX_TARGETS_36HR_DN	7	13.73	1.38E-06
PUJANA_BRCA1_PCC_NETWORK	14	27.45	1.48E-06
BENPORATH_CYCLING_GENES	10	19.61	1.48E-06
FISCHER_G2_M_CELL_CYCLE	7	13.73	3.66E-06
MARSON_BOUND_BY_E2F4_UNSTIMULATED	10	19.61	3.66E-06
NUYTEN_EZH2_TARGETS_DN	11	21.57	6.78E-06
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	6	11.76	9.02E-06
GAVIN_FOXP3_TARGETS_CLUSTER_P6	5	9.80	2.85E-05
KONG_E2F3_TARGETS	5	9.80	3.61E-05
KINSEY_TARGETS_OF_EWSR1_FLI1_FUSION_UP	11	21.57	4.74E-05
PUJANA_CHEK2_PCC_NETWORK	9	17.65	6.14E-05
MARKEY_RB1_ACUTE_LOF_UP	6	11.76	8.25E-05
KOBAYASHI_EGFR_SIGNALING_24HR_DN	6	11.76	0.000114

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (290)	q-value
FISCHER_DREAM_TARGETS	52	17.93	1.35E-28
MARSON_BOUND_BY_E2F4_UNSTIMULATED	38	13.10	2.71E-19
KINSEY_TARGETS_OF_EWSR1_FLI1_FUSION_UP	43	14.83	2.06E-15
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP	25	8.62	1.49E-10
HORIUCHI_WTAP_TARGETS_DN	19	6.55	4.72E-10
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_UP	25	8.62	9.07E-10
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	21	7.24	3.47E-09
LEE_BMP2_TARGETS_DN	28	9.66	4.31E-09
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	18	6.21	5.10E-09
PUJANA_CHEK2_PCC_NETWORK	26	8.97	7.01E-09
BLUM_RESPONSE_TO_SALIRASIB_DN	18	6.21	9.70E-09
FISCHER_G2_M_CELL_CYCLE	15	5.17	1.95E-08
MARKEY_RB1_ACUTE_LOF_UP	15	5.17	3.11E-08
JOHNSTONE_PARVB_TARGETS_3_DN	27	9.31	3.11E-08
KONG_E2F3_TARGETS	11	3.79	3.56E-08
KOBAYASHI_EGFR_SIGNALING_24HR_DN	15	5.17	6.14E-08
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	12	4.14	8.58E-08
PUJANA_BRCA1_PCC_NETWORK	35	12.07	2.09E-07
PATIL_LIVER_CANCER	23	7.93	2.59E-07
DODD_NASOPHARYNGEAL_CARCINOMA_DN	31	10.69	4.58E-07

Table H.2 The top 20 outputs following GSEA analysis using the MSig Gene Ontology database for the annotated genes significantly ($q < 0.05$) upregulated upon B (upper, 51 genes) or T (lower, 290 genes) cell activation of CIZ1 null lymphocytes only (inappropriate upregulation set). Sets which appear in both analyses are highlighted.

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (51)	q-value
GO_REGULATION_OF_CELL_CYCLE	12	23.53	3.65E-06
GO_CELL_CYCLE	13	25.49	6.07E-06
GO_CHROMOSOME	11	21.57	7.81E-06
GO_CHROMOSOME_ORGANIZATION	11	21.57	2.35E-05
GO_CHROMOSOME_SEGREGATION	7	13.73	3.13E-05
GO_CELL_CYCLE_PROCESS	11	21.57	3.13E-05
GO_SISTER_CHROMATID_SEGREGATION	6	11.76	5.22E-05
GO_ORGANELLE_FISSION	8	15.69	7.16E-05
GO_CONDENSED_CHROMOSOME	6	11.76	7.30E-05
GO_MITOTIC_NUCLEAR_DIVISION	7	13.73	0.00010804
GO_MITOTIC_CELL_CYCLE	9	17.65	0.00010804
GO_REGULATION_OF_CELL_CYCLE_PROCESS	8	15.69	0.00011518
GO_NUCLEAR_CHROMOSOME_SEGREGATION	6	11.76	0.0001237
GO_STRAND_DISPLACEMENT	3	5.88	0.00205965
GO_SISTER_CHROMATID_COHESION	4	7.84	0.00360532
GO_KINETOCHORE	4	7.84	0.00455229
GO_CELL_DIVISION	6	11.76	0.00499457
GO_REGULATION_OF_CELL_DIVISION	5	9.80	0.00517202
GO_CELLULAR_RESPONSE_TO_DNA_DAMAGE_STIMULUS	7	13.73	0.00537062
GO_DNA_REPAIR	6	11.76	0.00537713

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (290)	q-value
GO_RNA_BINDING	34	11.72	3.18E-06
GO_MITOTIC_CELL_CYCLE	23	7.93	3.18E-06
GO_RNA_PROCESSING	24	8.28	3.18E-06
GO_MICROTUBULE_CYTOSKELETON	27	9.31	3.22E-06
GO_CELL_CYCLE	30	10.34	3.27E-06
GO_CELL_CYCLE_PROCESS	26	8.97	1.16E-05
GO_CHROMOSOME	23	7.93	1.59E-05
GO_MITOTIC_NUCLEAR_DIVISION	15	5.17	1.59E-05
GO_POLY_A_RNA_BINDING	25	8.62	0.00013372
GO_CHROMOSOME_SEGREGATION	12	4.14	0.00018086
GO_SISTER_CHROMATID_SEGREGATION	10	3.45	0.00018088
GO_NUCLEAR_CHROMOSOME_SEGREGATION	11	3.79	0.00020285
GO_SPINDLE	12	4.14	0.00025775
GO_CHROMOSOME_ORGANIZATION	22	7.59	0.00032021
GO_NUCLEOLUS	20	6.90	0.00032021
GO_SISTER_CHROMATID_COHESION	8	2.76	0.00035383
GO_ORGANELLE_FISSION	15	5.17	0.00037797
GO_REGULATION_OF_CELL_CYCLE	21	7.24	0.00038027
GO_CELLULAR_RESPONSE_TO_DNA_DAMAGE_STIMULUS	18	6.21	0.00040962
GO_CYTOSKELETON	32	11.03	0.00042198

Table H.3 The top 20 outputs following GSEA analysis using the MSig Curated Gene Set database for the annotated genes significantly ($q < 0.05$) upregulated upon B (upper, 335 genes) or T (lower, 720 genes) cell activation of WT lymphocytes only (failure to upregulate set). Sets which appear in both analyses are highlighted.

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (335)	q-value
PUJANA_BRCA1_PCC_NETWORK	96	28.66	1.59E-52
LEE_BMP2_TARGETS_DN	67	20.00	4.89E-42
PUJANA_CHEK2_PCC_NETWORK	62	18.51	7.14E-40
WONG_EMBRYONIC_STEM_CELL_CORE	45	13.43	2.92E-37
DODD_NASOPHARYNGEAL_CARCINOMA_DN	71	21.19	8.41E-35
KINSEY_TARGETS_OF_EWSR1_FLII_FUSION_UP	68	20.30	5.28E-34
CAIRO_HEPATOBLASTOMA_CLASSES_UP	51	15.22	1.10E-33
FISCHER_DREAM_TARGETS	59	17.61	4.21E-33
PENG_GlutAMINE_DEPRIVATION_DN	41	12.24	1.67E-32
BERENJENO_TRANSFORMED_BY_RHOA_UP	46	13.73	1.34E-30
WEI_MYCN_TARGETS_WITH_E_BOX	53	15.82	1.49E-30
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	42	12.54	5.73E-29
GRADE_COLON_CANCER_UP	51	15.22	7.03E-27
KRIGE_RESPONSE_TO_TOSEDOSTAT_24HR_DN	54	16.12	8.88E-27
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP	42	12.54	1.78E-25
DIAZ_CHRONIC_MEYLOGENOUS_LEUKEMIA_UP	59	17.61	1.10E-24
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_UP	41	12.24	6.70E-23
MANALO_HYPOXIA_DN	31	9.25	6.70E-23
KRIGE_RESPONSE_TO_TOSEDOSTAT_6HR_DN	47	14.03	1.52E-22
RHEIN_ALL_GLUcOCORTICOID_THERAPY_DN	33	9.85	1.83E-22

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (720)	q-value
LEE_BMP2_TARGETS_DN	85	11.81	6.74E-35
PUJANA_BRCA1_PCC_NETWORK	109	15.14	5.17E-32
PILON_KLF1_TARGETS_DN	113	15.69	2.55E-28
KRIGE_RESPONSE_TO_TOSEDOSTAT_24HR_DN	77	10.69	1.24E-25
DIAZ_CHRONIC_MEYLOGENOUS_LEUKEMIA_UP	86	11.94	2.36E-23
HSIAO_HOUSEKEEPING_GENES	48	6.67	2.36E-23
KRIGE_RESPONSE_TO_TOSEDOSTAT_6HR_DN	65	9.03	4.41E-20
PUJANA_ATM_PCC_NETWORK	81	11.25	2.49E-19
REACTOME_METABOLISM_OF_RNA	40	5.56	3.88E-19
KEGG_RIBOSOME	25	3.47	4.09E-19
REACTOME_METABOLISM_OF_PROTEINS	48	6.67	8.07E-19
REACTOME_METABOLISM_OF_MRNA	37	5.14	1.14E-18
GRADE_COLON_CANCER_UP	61	8.47	1.28E-18
CAIRO_HEPATOBLASTOMA_CLASSES_UP	51	7.08	1.57E-18
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_DN	89	12.36	1.58E-18
KIM_BIPOLAR_DISORDER_OLIGODENDROCYTE_DENSITY_CORR_UP	52	7.22	3.29E-17
TIEN_INTESTINE_PROBIOTICS_24HR_UP	47	6.53	4.56E-17
DANG_BOUND_BY_MYC	66	9.17	4.80E-17
BENPORATH_MYC_MAX_TARGETS	55	7.64	4.87E-17
PUJANA_CHEK2_PCC_NETWORK	55	7.64	5.74E-17

Table H.4 The top 20 outputs following GSEA analysis using the MSig Gene Ontology database for the annotated genes significantly ($q < 0.05$) upregulated upon B (upper, 335 genes) or T (lower, 720 genes) cell activation of WT lymphocytes only (failure to upregulate set). Sets which appear in both analyses are highlighted.

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (335)	q-value
GO_CELL_CYCLE	66	19.70	1.40E-30
GO_MITOTIC_CELL_CYCLE	50	14.93	1.80E-27
GO_CELL_CYCLE_PROCESS	54	16.12	8.18E-25
GO_RNA_BINDING	64	19.10	8.18E-25
GO_ORGANONITROGEN_COMPOUND_METABOLIC_PROCESS	66	19.70	8.49E-24
GO_POLY_A_RNA_BINDING	53	15.82	1.44E-22
GO_ORGANONITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS	45	13.43	2.40E-18
GO_RIBONUCLEOPROTEIN_COMPLEX	38	11.34	9.57E-18
GO_ENVELOPE	45	13.43	1.93E-17
GO_SMALL_MOLECULE_METABOLIC_PROCESS	54	16.12	7.87E-16
GO_MITOCHONDRION	52	15.52	7.87E-16
GO_MITOCHONDRIAL_ENVELOPE	35	10.45	7.87E-16
GO_PROTEIN_COMPLEX_SUBUNIT_ORGANIZATION	50	14.93	1.03E-15
GO_ORGANELLE_INNER_MEMBRANE	31	9.25	1.03E-15
GO_MITOCHONDRIAL_PART	40	11.94	1.03E-15
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	28	8.36	7.27E-15
GO_REGULATION_OF_CELL_CYCLE	38	11.34	2.83E-14
GO_REGULATION_OF_CELL_CYCLE_PROCESS	30	8.96	3.43E-14
GO_ANAPHASE_PROMOTING_COMPLEX_DEPENDENT_CATABOLIC_PROCESS	15	4.48	7.37E-14
GO_MACROMOLECULAR_COMPLEX_ASSEMBLY	45	13.43	7.41E-14

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (720)	q-value
GO_ORGANONITROGEN_COMPOUND_METABOLIC_PROCESS	124	17.22	1.58E-38
GO_RNA_BINDING	114	15.83	1.40E-36
GO_RIBOSOME_BIOGENESIS	54	7.50	3.21E-33
GO_MITOCHONDRION	110	15.28	3.21E-33
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	61	8.47	1.03E-32
GO_RRNA_METABOLIC_PROCESS	49	6.81	9.34E-32
GO_RNA_PROCESSING	78	10.83	1.15E-31
GO_RIBONUCLEOPROTEIN_COMPLEX	73	10.14	1.32E-31
GO_RIBOSOME	46	6.39	7.80E-31
GO_ORGANONITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS	83	11.53	6.25E-30
GO_PEPTIDE_METABOLIC_PROCESS	64	8.89	6.97E-30
GO_MITOCHONDRIAL_PART	80	11.11	8.37E-30
GO_POLY_A_RNA_BINDING	88	12.22	9.33E-30
GO_CELLULAR_AMIDE_METABOLIC_PROCESS	70	9.72	3.25E-29
GO_AMIDE_BIOSYNTHETIC_PROCESS	60	8.33	3.40E-29
GO_ENVELOPE	84	11.67	4.47E-29
GO_NCRNA_PROCESSING	52	7.22	1.46E-27
GO_RIBOSOMAL_SUBUNIT	38	5.28	2.31E-27
GO_NUCLEOLUS	72	10.00	4.38E-27
GO_ORGANELLE_INNER_MEMBRANE	58	8.06	7.27E-27

Table H.5 The top 20 outputs following GSEA analysis using the MSig Curated Gene Set database for the annotated genes significantly ($q < 0.05$) downregulated upon B (upper, 121 genes) or T (lower, 354 genes) cell activation of CIZ1 null lymphocytes only (inappropriate downregulation set). Sets which appear in both analyses are highlighted.

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (121)	q-value
RAMALHO_STEMNESS_DN	8	6.61	2.23E-07
CHEN_METABOLIC_SYNDROM_NETWORK	20	16.53	2.23E-07
HUANG_GATA2_TARGETS_UP	9	7.44	7.62E-07
HESS_TARGETS_OF_HOXA9_AND_MEIS1_DN	7	5.79	3.78E-06
RUTELLA_RESPONSE_TO_HGF_UP	11	9.09	2.06E-05
JOHNSTONE_PARVB_TARGETS_3_UP	11	9.09	2.06E-05
VERHAAK_AML_WITH_NPM1_MUTATED_DN	9	7.44	2.06E-05
HADDAD_T_LYMPHOCYTE_AND_NK_PROGENITOR_DN	6	4.96	2.06E-05
LENAOUR_DENDRITIC_CELL_MATURATION_UP	7	5.79	2.10E-05
QI_PLASMACYTOMA_UP	9	7.44	2.24E-05
ALCALAY_AML_BY_NPM1_LOCALIZATION_DN	8	6.61	2.26E-05
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP	16	13.22	3.11E-05
LI_INDUCED_T_TO_NATURAL_KILLER_UP	9	7.44	6.94E-05
LIU_CMYB_TARGETS_UP	7	5.79	0.000137002
BROWN_MYELOID_CELL_DEVELOPMENT_UP	7	5.79	0.000137002
LABBE_TARGETS_OF_TGFB1_AND_WNT3A_UP	6	4.96	0.00022033
SANSOM_APC_TARGETS_DN	9	7.44	0.00022033
GRAESSMANN_RESPONSE_TO_MC_AND_DOXORUBICIN_UP	11	9.09	0.000236824
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	7	5.79	0.000236824
SWEET_LUNG_CANCER_KRAS_UP	10	8.26	0.000236824

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (354)	q-value
CHEN_METABOLIC_SYNDROM_NETWORK	40	11.30	1.58E-10
REN_ALVEOLAR_RHABDOMYOSARCOMA_DN	23	6.50	1.25E-09
WONG_ADULT_TISSUE_STEM_MODULE	29	8.19	2.28E-09
MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	35	9.89	2.28E-09
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_DN	34	9.60	9.56E-09
BROWN_MYELOID_CELL_DEVELOPMENT_UP	15	4.24	9.56E-09
OSWALD_HEMATOPOIETIC_STEM_CELL_IN_COLLAGEN_GEL_UP	17	4.80	9.56E-09
JOHNSTONE_PARVB_TARGETS_3_UP	21	5.93	4.12E-08
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_UP	28	7.91	5.54E-08
DELYS_THYROID_CANCER_UP	21	5.93	5.54E-08
RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_DN	18	5.08	6.55E-08
DUTERTRE ESTRADIOL_RESPONSE_24HR_DN	22	6.21	7.12E-08
FOSTER_TOLERANT_MACROPHAGE_DN	20	5.65	7.12E-08
FORTSCHEGGER_PHF8_TARGETS_DN	27	7.63	8.54E-08
RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY_DN	19	5.37	1.35E-07
BILD_HRAS_ONCOGENIC_SIGNATURE	16	4.52	1.84E-07
DURCHDEWALD_SKIN_CARCINOGENESIS_DN	16	4.52	2.03E-07
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP	32	9.04	2.06E-07
MILI_PSEUDOPODIA_CHEMOTAXIS_DN	20	5.65	3.00E-07
PETROVA_ENDOTHELIUM_LYMPHATIC_VS_BLOOD_DN	13	3.67	3.00E-07

Table H.6 The top 20 outputs following GSEA analysis using the MSig Gene Ontology database for the annotated genes significantly ($q < 0.05$) downregulated upon B (upper, 121 genes) or T (lower, 354 genes) cell activation of *CI21* null lymphocytes only (inappropriate downregulation set). Sets which appear in both analyses are highlighted.

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (121)	q-value
GO_IMMUNE_SYSTEM_PROCESS	29	23.97	3.37E-10
GO_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	21	17.36	5.24E-07
GO_REGULATION_OF_CELL_DIFFERENTIATION	20	16.53	6.38E-06
GO_REGULATION_OF_IMMUNE_RESPONSE	15	12.40	1.80E-05
GO_TISSUE_DEVELOPMENT	19	15.70	2.90E-05
GO_IMMUNE_RESPONSE	16	13.22	4.64E-05
GO_REGULATION_OF_CELL_DEATH	18	14.88	6.43E-05
GO_POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	14	11.57	6.43E-05
GO_CELLULAR_RESPONSE_TO_ORGANIC_SUBSTANCE	20	16.53	6.43E-05
GO_REGULATION_OF_MULTICELLULAR_ORGANISMAL_DEVELOPMENT	19	15.70	6.43E-05
GO_CELL_SURFACE	13	10.74	7.97E-05
GO_POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS	17	14.05	0.000102745
GO_REGULATION_OF_PHOSPHORUS_METABOLIC_PROCESS	18	14.88	0.000150558
GO_POSITIVE_REGULATION_OF_CELL_DIFFERENTIATION	13	10.74	0.000155416
GO_RESPONSE_TO_EXTERNAL_STIMULUS	19	15.70	0.000155416
GO_RECEPTOR_ACTIVITY	18	14.88	0.000160532
GO_RESPONSE_TO_CYTOKINE	12	9.92	0.000199295
GO_NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS	16	13.22	0.000254339
GO_IMMUNE_EFFECTOR_PROCESS	10	8.26	0.000260524
GO_IDENTICAL_PROTEIN_BINDING	15	12.40	0.000260524

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (354)	q-value
GO_VACUOLE	41	11.58	1.80E-11
GO_LOCOMOTION	38	10.73	1.73E-10
GO_REGULATION_OF_CELL_DIFFERENTIATION	43	12.15	5.88E-10
GO_ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	29	8.19	7.07E-10
GO_POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS	41	11.58	7.07E-10
GO_POSITIVE_REGULATION_OF_BIOSYNTHETIC_PROCESS	47	13.28	7.07E-10
GO_REGULATION_OF_MULTICELLULAR_ORGANISMAL_DEVELOPMENT	44	12.43	2.52E-09
GO_IMMUNE_SYSTEM_PROCESS	48	13.56	3.71E-09
GO_CELLULAR_RESPONSE_TO_ORGANIC_SUBSTANCE	46	12.99	3.79E-09
GO_RESPONSE_TO_WOUNDING	25	7.06	4.16E-09
GO_POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	35	9.89	5.96E-09
GO_REGULATION_OF_RESPONSE_TO_STRESS	39	11.02	2.14E-08
GO_WOUND_HEALING	22	6.21	2.18E-08
GO_REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS	32	9.04	2.18E-08
GO_ENDOSOME	28	7.91	2.57E-08
GO_TISSUE_DEVELOPMENT	39	11.02	4.14E-08
GO_HEMOSTASIS	18	5.08	4.15E-08
GO_REGULATION_OF_BODY_FLUID_LEVELS	22	6.21	6.34E-08
GO_VESICLE_MEDIATED_TRANSPORT	34	9.60	1.15E-07
GO_POSITIVE_REGULATION_OF_GENE_EXPRESSION	41	11.58	1.17E-07

Table H.7 The top 20 outputs following GSEA analysis using the MSig Curated Gene Set database for the annotated genes significantly ($q < 0.05$) downregulated upon B (upper, 185 genes) or T (lower, 558 genes) cell activation of WT lymphocytes only (failure to downregulate set). Sets which appear in both analyses are highlighted.

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (185)	q-value
ZHENG_BOUND_BY_FOXP3	25	13.51	4.30E-16
CUI_TCF21_TARGETS_2_DN	23	12.43	1.72E-09
ZHENG_FOXP3_TARGETS_IN_THYMUS_UP	13	7.03	5.69E-09
VERHAAK_AML_WITH_NPM1_MUTATED_DN	13	7.03	6.21E-08
LI_INDUCED_T_TO_NATURAL_KILLER_UP	12	6.49	5.17E-06
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_UP	10	5.41	5.17E-06
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_UP	18	9.73	5.17E-06
WONG_ADULT_TISSUE_STEM_MODULE	17	9.19	5.17E-06
SMID_BREAST_CANCER_NORMAL_LIKE_UP	14	7.57	7.10E-06
RUIZ_TNC_TARGETS_UP	9	4.86	1.05E-05
FEVR_CTNNB1_TARGETS_UP	16	8.65	1.12E-05
MARSON_BOUND_BY_FOXP3_UNSTIMULATED	21	11.35	1.38E-05
BOYLAN_MULTIPLE_MYELOMA_C_D_DN	10	5.41	4.46E-05
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_DN	8	4.32	4.64E-05
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP	19	10.27	7.70E-05
REACTOME_IMMUNE_SYSTEM	17	9.19	9.21E-05
BAKKER_FOXP3_TARGETS_UP	6	3.24	9.21E-05
MOSERLE_IFNA_RESPONSE	5	2.70	9.65E-05
WOO_LIVER_CANCER_RECURRENCE_UP	7	3.78	9.65E-05
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_DN	18	9.73	0.000118869

Name of Curated Gene Set	Genes in Overlap	Percentage of genes in set (558)	q-value
CUI_TCF21_TARGETS_2_DN	49	8.78	2.75E-15
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_UP	48	8.60	2.91E-15
ZHENG_BOUND_BY_FOXP3	34	6.09	3.16E-12
PILON_KLF1_TARGETS_DN	70	12.54	3.16E-12
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_DN	48	8.60	4.08E-11
DUTERTRE ESTRADIOL_RESPONSE_24HR_DN	32	5.73	1.15E-10
GABRIELY_MIR21_TARGETS	24	4.30	5.97E-10
DACOSTA_UV_RESPONSE_VIA_ERCC3_DN	39	6.99	3.23E-09
FOSTER_TOLERANT_MACROPHAGE_DN	26	4.66	1.30E-08
BLALOCK_ALZHEIMERS_DISEASE_UP	56	10.04	1.30E-08
ZWANG_TRANSIENTLY_UP_BY_1ST_EGF_PULSE_ONLY	58	10.39	3.17E-08
NAKAMURA_TUMOR_ZONE_PERIPHERAL_VS_CENTRAL_DN	31	5.56	6.91E-08
HADDAD_B_LYMPHOCYTE_PROGENITOR	21	3.76	1.08E-07
RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_DN	28	5.02	1.11E-07
NUYTEN_EZH2_TARGETS_UP	40	7.17	1.11E-07
MARSON_BOUND_BY_FOXP3_UNSTIMULATED	44	7.89	1.18E-07
KOINUMA_TARGETS_OF_SMAD2_OR_SMAD3	35	6.27	1.18E-07
WONG_ADULT_TISSUE_STEM_MODULE	32	5.73	2.30E-07
MANALO_HYPOXIA_UP	17	3.05	5.60E-07
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_DN	33	5.91	7.41E-07

Table H.8 The top 20 outputs following GSEA analysis using the MSig Gene Ontology database for the annotated genes significantly ($q < 0.05$) downregulated upon B (upper, 185 genes) or T (lower, 558 genes) cell activation of WT lymphocytes only (failure to downregulate set). Sets which appear in both analyses are highlighted.

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (185)	q-value
GO_CELL_ADHESION_MOLECULE_BINDING	11	5.95	3.05E-06
GO_IMMUNE_SYSTEM_PROCESS	29	15.68	5.33E-06
GO_MOVEMENT_OF_CELL_OR_SUBCELLULAR_COMPONENT	23	12.43	5.33E-06
GO_BIOLOGICAL_ADHESION	20	10.81	1.39E-05
GO_CELL_MOTILITY	18	9.73	1.39E-05
GO_INTEGRIN_BINDING	8	4.32	1.94E-05
GO_LOCOMOTION	20	10.81	2.88E-05
GO_IMMUNE_RESPONSE	19	10.27	9.91E-05
GO_LEUKOCYTE_MIGRATION	10	5.41	9.91E-05
GO_REGULATION_OF_CELL_PROLIFERATION	22	11.89	0.000121603
GO_RECEPTOR_ACTIVITY	23	12.43	0.000139745
GO_REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS	17	9.19	0.000139745
GO_RESPONSE_TO_EXTERNAL_STIMULUS	24	12.97	0.000177724
GO_CELL_CHEMOTAXIS	8	4.32	0.000195926
GO_INTRINSIC_COMPONENT_OF_PLASMA_MEMBRANE	22	11.89	0.000402128
GO_MACROMOLECULAR_COMPLEX_BINDING	20	10.81	0.000433033
GO_REGULATION_OF_RESPONSE_TO_WOUNDING	11	5.95	0.000434865
GO_REGULATION_OF_BODY_FLUID_LEVELS	12	6.49	0.000434865
GO_POSITIVE_REGULATION_OF_CELL_PROLIFERATION	15	8.11	0.000434865
GO_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	19	10.27	0.001388147

Name of Gene Ontology Set	Genes in Overlap	Percentage of genes in set (558)	q-value
GO_POSITIVE_REGULATION_OF_MOLECULAR_FUNCTION	62	11.11	2.01E-09
GO_POSITIVE_REGULATION_OF_CATALYTIC_ACTIVITY	52	9.32	8.97E-08
GO_REGULATION_OF_HYDROLASE_ACTIVITY	48	8.60	8.97E-08
GO_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	57	10.22	9.71E-08
GO_MACROMOLECULAR_COMPLEX_BINDING	49	8.78	9.71E-08
GO_TUBE_DEVELOPMENT	29	5.20	1.34E-07
GO_INTRACELLULAR_VESICLE	45	8.06	2.30E-07
GO_REGULATION_OF_GTPASE_ACTIVITY	31	5.56	4.95E-07
GO_TISSUE_DEVELOPMENT	49	8.78	7.31E-07
GO_POSITIVE_REGULATION_OF_HYDROLASE_ACTIVITY	36	6.45	7.31E-07
GO_CIRCULATORY_SYSTEM_DEVELOPMENT	33	5.91	9.88E-07
GO_VACUOLE	41	7.35	1.88E-06
GO_POSITIVE_REGULATION_OF_GENE_EXPRESSION	51	9.14	4.55E-06
GO_EPITHELIUM_DEVELOPMENT	35	6.27	5.06E-06
GO_GOLGI_APPARATUS	45	8.06	6.23E-06
GO_PROTEIN_C_TERMINUS_BINDING	15	2.69	8.58E-06
GO_ENDOSOME	31	5.56	9.04E-06
GO_REGULATION_OF_CELL_ADHESION	27	4.84	1.13E-05
GO_NEGATIVE_REGULATION_OF_GENE_EXPRESSION	45	8.06	1.24E-05
GO_REGULATION_OF_INTRACELLULAR_SIGNAL_TRANSDUCTION	48	8.60	1.28E-05

List of Abbreviations

°C	degrees centigrade
ARF	Alternative reading frame
bp	Base Pair
BRCA1	Breast Cancer Type 1 Susceptibility Protein
BSA	Bovine Serum Albumin
CBX	Chromobox protein
CDK	Cyclin Dependent Kinase
CIZ1	Cip1-interacting zinc finger protein 1
CSK	Cytoskeletal Buffer
CST	Cell Signalling Technologies
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DREAM	Dimerization partner, RB-like, E2F and multi-vulval class B
DTSP	dithiobis succinimidyl propionate
DTT	Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
EED	Embryonic ectoderm development
EGTA	ethylene glycol tetraacetic acid
ES	Embryonic Stem Cells
EZH2	Enhancer Of Zeste Homolog 2
FBS	Foetal bovine serum
<i>Firre</i>	Functional Intergenic Repeating RNA Element
FOXP3	Forkhead Box P3
G0	Quiescence
G1	Growth phase 1
G2	Growth phase 2
GFP	Green fluorescent protein
GSEA	Gene set enrichment analysis
H2AK119Ub1	Mono-ubiquitination of lysine 119 on histone 2A

H3	Histone 3
H3K27me3	Trimethylation of lysine 27 on histone 3
hnRNP U	Heterogenous Nuclear Ribonucleoprotein U
HRP	Horseradish peroxidase
IF	Immunofluorescence
iPS	Induced pluripotent stem cells
iRS	Intermediately reprogrammed stem cells
KB	Kilo base pairs
KDa	Kilodalton
LINES	Long interspersed repetitive sequence elements
lncRNA	Long Non-Coding RNA
LPS	Lipo poly-saccharides
M	Mitosis
MEF	Culture adapted mouse embryonic fibroblast
NEB	New England Biolabs
NM	Nuclear Matrix
nm	Nanometer
NM1	Nuclear Myosin 1
P	Phosphorylation site
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCGF	Polycomb group ring fingers
PEF	Primary mouse embryonic fibroblast
PFA	Paraformaldehyde
PHC	Polyhomeotic-like protein
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
PLA	Proximity ligation assay
PMSF	Phenylmethanesulfonylfluoride
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PTBP	Polypyrimidine tract binding protein
q value	False detection rate corrected p value
RB1	Retinoblastoma protein 1

RBAP	Retinoblastoma-Binding Protein
RING	Really interesting new gene
RNA	Ribonucleic acid
RXL	Cyclin binding motif
S	DNA synthesis phase
S/MAR	Scaffold/Matrix attachment region
SAF-A	Scaffold Attachment Factor A
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SUZ12	Suppressor Of Zeste 12 Protein Homolog
TBS	Tris Buffered Saline
TCF21	Transcription Factor 21
VRC	Vanadyl ribonucleoside complex
WT	Wild type
Xa	Active X Chromosome
XCI	X Chromosome Inactivation
xg	Multiplied by the force of gravity
Xi	Inactive X Chromosome
XIC	X Inactivation Centre
<i>Xist</i>	X-inactive specific transcript
Xm	Maternal X Chromosome
Xp	Paternal X Chromosome
YY1	Yin and Yang 1
ZF	Zinc finger

References

- ADAMS, R. R., ECKLEY, D. M., VAGNARELLI, P., WHEATLEY, S. P., GERLOFF, D. L., MACKAY, A. M., SVINGEN, P. A., KAUFMANN, S. H. & EARNSHAW, W. C. 2001. Human INCENP colocalizes with the Aurora-B/AIRK2 kinase on chromosomes and is overexpressed in tumour cells. *Chromosoma*, 110, 65-74.
- ADAMS, S. R., MAEZAWA, S., ALAVATTAM, K. G., ABE, H., SAKASHITA, A., SHRODER, M., BROERING, T. J., SROGA RIOS, J., THOMAS, M. A., LIN, X., PRICE, C. M., BARSKI, A., ANDREASSEN, P. R. & NAMEKAWA, S. H. 2018. RNF8 and SCML2 cooperate to regulate ubiquitination and H3K27 acetylation for escape gene activation on the sex chromosomes. *PLoS Genet*, 14, e1007233.
- AGRELO, R., SOUABNI, A., NOVATCHKOVA, M., HASLINGER, C., LEEB, M., KOMNENOVIC, V., KISHIMOTO, H., GRESH, L., KOHWI-SHIGEMATSU, T., KENNER, L. & WUTZ, A. 2009. SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Dev Cell*, 16, 507-16.
- AINSCOUGH, J. F., RAHMAN, F. A., SERCOMBE, H., SEDO, A., GERLACH, B. & COVERLEY, D. 2007. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. *J Cell Sci*, 120, 115-24.
- AN, J., DU, Y., FAN, X., WANG, Y., IVAN, C., ZHANG, X. G., SOOD, A. K., AN, Z. & ZHANG, N. 2019. EGFL6 promotes breast cancer by simultaneously enhancing cancer cell metastasis and stimulating tumor angiogenesis. *Oncogene*, 38, 2123-2134.
- AUGUI, S., FILION, G. J., HUART, S., NORA, E., GUGGIARI, M., MARESCA, M., STEWART, A. F. & HEARD, E. 2007. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science*, 318, 1632-6.
- AUGUI, S., NORA, E. P. & HEARD, E. 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet*, 12, 429-42.
- AYAD, N. G., RANKIN, S., MURAKAMI, M., JEBANATHIRAJAH, J., GYGI, S. & KIRSCHNER, M. W. 2003. Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell*, 113, 101-13.
- BARR, M. L. & BERTRAM, E. G. 1949. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature*, 163, 676.
- BELTRAN, M., YATES, C. M., SKALSKA, L., DAWSON, M., REIS, F. P., VIIRI, K., FISHER, C. L., SIBLEY, C. R., FOSTER, B. M., BARTKE, T., ULE, J. & JENNER, R. G. 2016. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res*, 26, 896-907.
- BENNETT, C. L., CHRISTIE, J., RAMSDELL, F., BRUNKOW, M. E., FERGUSON, P. J., WHITESELL, L., KELLY, T. E., SAULSBURY, F. T., CHANCE, P. F. & OCHS, H. D. 2001. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet*, 27, 20-1.
- BEREZNEY, R. & COFFEY, D. S. 1974. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun*, 60, 1410-7.
- BERLETCH, J. B., MA, W., YANG, F., SHENDURE, J., NOBLE, W. S., DISTECHE, C. M. & DENG, X. 2015. Escape from X inactivation varies in mouse tissues. *PLoS Genet*, 11, e1005079.

- BERLETCH, J. B., YANG, F., XU, J., CARREL, L. & DISTECHE, C. M. 2011. Genes that escape from X inactivation. *Hum Genet*, 130, 237-45.
- BHARADWAJ, R., QI, W. & YU, H. 2004. Identification of two novel components of the human NDC80 kinetochore complex. *J Biol Chem*, 279, 13076-85.
- BLACKLEDGE, N. P., FARCAS, A. M., KONDO, T., KING, H. W., MCGOURAN, J. F., HANSEN, L. L., ITO, S., COOPER, S., KONDO, K., KOSEKI, Y., ISHIKURA, T., LONG, H. K., SHEAHAN, T. W., BROCKDORFF, N., KESSLER, B. M., KOSEKI, H. & KLOSE, R. J. 2014. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, 157, 1445-59.
- BODNAR, A. G., OUELLETTE, M., FROLKIS, M., HOLT, S. E., CHIU, C. P., MORIN, G. B., HARLEY, C. B., SHAY, J. W., LICHTSTEINER, S. & WRIGHT, W. E. 1998. Extension of life-span by introduction of telomerase into normal human cells. *Science*, 279, 349-52.
- BROCKDORFF, N. 2017. Polycomb complexes in X chromosome inactivation. *Philos Trans R Soc Lond B Biol Sci*, 372.
- BROOKS, R. F. 1976. Regulation of fibroblast cell cycle by serum. *Nature*, 260, 248-50.
- BROWN, C. J., HENDRICH, B. D., RUPERT, J. L., LAFRENIERE, R. G., XING, Y., LAWRENCE, J. & WILLARD, H. F. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, 71, 527-42.
- BULT, C. J., BLAKE, J. A., SMITH, C. L., KADIN, J. A., RICHARDSON, J. E. & MOUSE GENOME DATABASE, G. 2019. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*, 47, D801-D806.
- BUONGIORNO-NARDELLI, M., MICHELI, G., CARRI, M. T. & MARILLEY, M. 1982. A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature*, 298, 100-2.
- BUSHMEYER, S. M. & ATCHISON, M. L. 1998. Identification of YY1 sequences necessary for association with the nuclear matrix and for transcriptional repression functions. *J Cell Biochem*, 68, 484-99.
- CAO, R., WANG, L., WANG, H., XIA, L., ERDJUMENT-BROMAGE, H., TEMPST, P., JONES, R. S. & ZHANG, Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298, 1039-43.
- CAPCO, D. G., WAN, K. M. & PENMAN, S. 1982. The nuclear matrix: three-dimensional architecture and protein composition. *Cell*, 29, 847-58.
- CASAS-DELUCCHI, C. S., BRERO, A., RAHN, H. P., SOLOVEI, I., WUTZ, A., CREMER, T., LEONHARDT, H. & CARDOSO, M. C. 2011. Histone acetylation controls the inactive X chromosome replication dynamics. *Nat Commun*, 2, 222.
- CERASE, A., ARMAOS, A., NEUMAYER, C., AVNER, P., GUTTMAN, M. & TARTAGLIA, G. G. 2019. Phase separation drives X-chromosome inactivation: a hypothesis. *Nat Struct Mol Biol*, 26, 331-334.
- CHEN, C. K., BLANCO, M., JACKSON, C., AZNAURYAN, E., OLLIKAINEN, N., SURKA, C., CHOW, A., CERASE, A., MCDONEL, P. & GUTTMAN, M. 2016. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science*, 354, 468-472.

- CHEN, G. & DENG, X. 2018. Cell Synchronization by Double Thymidine Block. *Bio Protoc*, 8.
- CHEN, S., BIRVE, A. & RASMUSON-LESTANDER, A. 2008. In vivo analysis of *Drosophila* SU(Z)12 function. *Mol Genet Genomics*, 279, 159-70.
- CHEN, Y., HENNESSY, K. M., BOTSTEIN, D. & TYE, B. K. 1992. CDC46/MCM5, a yeast protein whose subcellular localization is cell cycle-regulated, is involved in DNA replication at autonomously replicating sequences. *Proc Natl Acad Sci U S A*, 89, 10459-63.
- CHU, C., ZHANG, Q. C., DA ROCHA, S. T., FLYNN, R. A., BHARADWAJ, M., CALABRESE, J. M., MAGNUSON, T., HEARD, E. & CHANG, H. Y. 2015. Systematic discovery of Xist RNA binding proteins. *Cell*, 161, 404-16.
- CHUANG, C. H., CARPENTER, A. E., FUCHSOVA, B., JOHNSON, T., DE LANEROLLE, P. & BELMONT, A. S. 2006. Long-range directional movement of an interphase chromosome site. *Curr Biol*, 16, 825-31.
- COOPER, S., DIENSTBIER, M., HASSAN, R., SCHERMELLEH, L., SHARIF, J., BLACKLEDGE, N. P., DE MARCO, V., ELDERKIN, S., KOSEKI, H., KLOSE, R., HEGER, A. & BROCKDORFF, N. 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep*, 7, 1456-1470.
- COOPER, S., GRIJZENHOUT, A., UNDERWOOD, E., ANCELIN, K., ZHANG, T., NESTEROVA, T. B., ANIL-KIRMIZITAS, B., BASSETT, A., KOOISTRA, S. M., AGGER, K., HELIN, K., HEARD, E. & BROCKDORFF, N. 2016. Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat Commun*, 7, 13661.
- COPELAND, N. A., SERCOMBE, H. E., AINSCOUGH, J. F. & COVERLEY, D. 2010. Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. *J Cell Sci*, 123, 1108-15.
- COPELAND, N. A., SERCOMBE, H. E., WILSON, R. H. & COVERLEY, D. 2015. Cyclin-A-CDK2-mediated phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. *J Cell Sci*, 128, 1518-27.
- COVERLEY, D., HIGGINS, G., WEST, D., JACKSON, O. T., DOWLE, A., HASLAM, A., AINSCOUGH, E., CHALKLEY, R. & WHITE, J. 2016. A quantitative immunoassay for lung cancer biomarker CIZ1b in patient plasma. *Clin Biochem*.
- COVERLEY, D., LAMAN, H. & LASKEY, R. A. 2002. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. *Nat Cell Biol*, 4, 523-8.
- COVERLEY, D., MARR, J. & AINSCOUGH, J. 2005. Ciz1 promotes mammalian DNA replication. *J Cell Sci*, 118, 101-12.
- CSANKOVSKI, G., PANNING, B., BATES, B., PEHRSON, J. R. & JAENISCH, R. 1999. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet*, 22, 323-4.
- DA ROCHA, S. T., BOEVA, V., ESCAMILLA-DEL-ARENAL, M., ANCELIN, K., GRANIER, C., MATIAS, N. R., SANULLI, S., CHOW, J., SCHULZ, E., PICARD, C., KANEKO, S., HELIN, K., REINBERG, D., STEWART, A. F., WUTZ, A., MARGUERON, R. & HEARD, E. 2014. Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell*, 53, 301-16.

- DAHMCCKE, C. M., BÜCHMANN-MØLLER, S., JENSEN, N. A. & MITCHELMORE, C. 2008. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol Cell Neurosci*, 38, 589-94.
- DALTON, W. B. & YANG, V. W. 2009. Role of prolonged mitotic checkpoint activation in the formation and treatment of cancer. *Future Oncol*, 5, 1363-70.
- DAVIDO, T. & GETZENBERG, R. H. 2000. Nuclear matrix proteins as cancer markers. *J Cell Biochem Suppl*, Suppl 35, 136-41.
- DELGADO, M. D. & LEON, J. 2006. Gene expression regulation and cancer. *Clin Transl Oncol*, 8, 780-7.
- DERE, E., WINKLER, D., RITTER, C., RONNENBERG, A., POGGI, G., PATZIG, J., GERNERT, M., MULLER, C., NAVE, K. A., EHRENREICH, H. & WERNER, H. B. 2015. Gpm6b deficiency impairs sensorimotor gating and modulates the behavioral response to a 5-HT2A/C receptor agonist. *Behav Brain Res*, 277, 254-63.
- DIAZ, E., MACHUTTA, C. A., CHEN, S., JIANG, Y., NIXON, C., HOFMANN, G., KEY, D., SWEITZER, S., PATEL, M., WU, Z., CREASY, C. L., KRUGER, R. G., LAFRANCE, L., VERMA, S. K., PAPPALARDI, M. B., LE, B., VAN ALLER, G. S., MCCABE, M. T., TUMMINO, P. J., POPE, A. J., THRALL, S. H., SCHWARTZ, B. & BRANDT, M. 2012. Development and validation of reagents and assays for EZH2 peptide and nucleosome high-throughput screens. *J Biomol Screen*, 17, 1279-92.
- DIJKWEL, P. A., VAUGHN, J. P. & HAMLIN, J. L. 1991. Mapping of replication initiation sites in mammalian genomes by two-dimensional gel analysis: stabilization and enrichment of replication intermediates by isolation on the nuclear matrix. *Mol Cell Biol*, 11, 3850-9.
- DONOHUE, M. E., SILVA, S. S., PINTER, S. F., XU, N. & LEE, J. T. 2009. The pluripotency factor Oct4 interacts with Ctf and also controls X-chromosome pairing and counting. *Nature*, 460, 128-32.
- DOWLING, M. R., KAN, A., HEINZEL, S., ZHOU, J. H., MARCHINGO, J. M., WELLARD, C. J., MARKHAM, J. F. & HODGKIN, P. D. 2014. Stretched cell cycle model for proliferating lymphocytes. *Proc Natl Acad Sci U S A*, 111, 6377-82.
- DRABEK, K., VAN DE PEPPEL, J., EIJKEN, M. & VAN LEEUWEN, J. P. 2011. GPM6B regulates osteoblast function and induction of mineralization by controlling cytoskeleton and matrix vesicle release. *J Bone Miner Res*, 26, 2045-51.
- EATON, W. W., ROSE, N. R., KALAYDJIAN, A., PEDERSEN, M. G. & MORTENSEN, P. B. 2007. Epidemiology of autoimmune diseases in Denmark. *J Autoimmun*, 29, 1-9.
- ENGELAND, K. 2018. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ*, 25, 114-132.
- FABREGAT, A., JUPE, S., MATTHEWS, L., SIDIROPOULOS, K., GILLESPIE, M., GARAPATI, P., HAW, R., JASSAL, B., KORNINGER, F., MAY, B., MILACIC, M., ROCA, C. D., ROTHFELS, K., SEVILLA, C., SHAMOVSKY, V., SHORSER, S., VARUSAI, T., VITERI, G., WEISER, J., WU, G., STEIN, L., HERMJAKOB, H. & D'EUSTACHIO, P. 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*, 46, D649-D655.
- FABREGAT, A., SIDIROPOULOS, K., VITERI, G., FORNER, O., MARIN-GARCIA, P., ARNAU, V., D'EUSTACHIO, P., STEIN, L. & HERMJAKOB, H. 2017. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics*, 18, 142.

- FAUST, C., LAWSON, K. A., SCHORK, N. J., THIEL, B. & MAGNUSON, T. 1998. The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development*, 125, 4495-506.
- FAY, M. M. & ANDERSON, P. J. 2018. The Role of RNA in Biological Phase Separations. *J Mol Biol*, 430, 4685-4701.
- FERNANDES, J. C. R., ACUNA, S. M., AOKI, J. I., FLOETER-WINTER, L. M. & MUXEL, S. M. 2019. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA*, 5.
- FIORE, A., RIBEIRO, P. F. & BRUNI-CARDOSO, A. 2018. Sleeping Beauty and the Microenvironment Enchantment: Microenvironmental Regulation of the Proliferation-Quiescence Decision in Normal Tissues and in Cancer Development. *Front Cell Dev Biol*, 6, 59.
- FISCHER, M., GROSSMANN, P., PADI, M. & DECAPRIO, J. A. 2016. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res*, 44, 6070-86.
- GARITANO-TROJAOLA, A., AGIRRE, X., PROSPER, F. & FORTES, P. 2013. Long non-coding RNAs in haematological malignancies. *Int J Mol Sci*, 14, 15386-422.
- GARTLER, S. M. & RIGGS, A. D. 1983. Mammalian X-chromosome inactivation. *Annu Rev Genet*, 17, 155-90.
- GASSMANN, R., CARVALHO, A., HENZING, A. J., RUCHAUD, S., HUDSON, D. F., HONDA, R., NIGG, E. A., GERLOFF, D. L. & EARNSHAW, W. C. 2004. Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. *J Cell Biol*, 166, 179-91.
- GOLUBOVSKAYA, V. & WU, L. 2016. Different Subsets of T Cells, Memory, Effector Functions, and CAR-T Immunotherapy. *Cancers (Basel)*, 8.
- GOS, M., MILOSZEWSKA, J., SWOBODA, P., TREMBACZ, H., SKIERSKI, J. & JANIK, P. 2005. Cellular quiescence induced by contact inhibition or serum withdrawal in C3H10T1/2 cells. *Cell Prolif*, 38, 107-16.
- GRZENDA, A., LOMBERK, G., SVINGEN, P., MATHISON, A., CALVO, E., IOVANNA, J., XIONG, Y., FAUBION, W. & URRUTIA, R. 2013. Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. *Epigenetics Chromatin*, 6, 3.
- GUENEAU, L., BERTRAND, A. T., JAIS, J. P., SALIH, M. A., STOJKOVIC, T., WEHNERT, M., HOELTZENBEIN, M., SPULER, S., SAITOH, S., VERSCHUEREN, A., TRANCHANT, C., BEUVIN, M., LACENE, E., ROMERO, N. B., HEATH, S., ZELENKA, D., VOIT, T., EYMARD, B., BEN YAOU, R. & BONNE, G. 2009. Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. *Am J Hum Genet*, 85, 338-53.
- GUO, B., ODGREN, P. R., VAN WIJNEN, A. J., LAST, T. J., NICKERSON, J., PENMAN, S., LIAN, J. B., STEIN, J. L. & STEIN, G. S. 1995. The nuclear matrix protein NMP-1 is the transcription factor YY1. *Proc Natl Acad Sci U S A*, 92, 10526-30.
- GUO, Y., KIM, C., AHMAD, S., ZHANG, J. & MAO, Y. 2012. CENP-E--dependent BubR1 autophosphorylation enhances chromosome alignment and the mitotic checkpoint. *J Cell Biol*, 198, 205-17.
- HAN LI, C. & CHEN, Y. 2015. Targeting EZH2 for cancer therapy: progress and perspective. *Curr Protein Pept Sci*, 16, 559-70.

- HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- HASEGAWA, Y., BROCKDORFF, N., KAWANO, S., TSUTUI, K. & NAKAGAWA, S. 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell*, 19, 469-76.
- HAYFLICK, L. & MOORHEAD, P. S. 1961. The serial cultivation of human diploid cell strains. *Exp Cell Res*, 25, 585-621.
- HELBIG, R. & FACKELMAYER, F. O. 2003. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma*, 112, 173-82.
- HESKETH, E. L., KNIGHT, J. R., WILSON, R. H., CHONG, J. P. & COVERLEY, D. 2015. Transient association of MCM complex proteins with the nuclear matrix during initiation of mammalian DNA replication. *Cell Cycle*, 14, 333-41.
- HIGGINS, G., ROPER, K. M., WATSON, I. J., BLACKHALL, F. H., ROM, W. N., PASS, H. I., AINSCOUGH, J. F. & COVERLEY, D. 2012. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc Natl Acad Sci U S A*, 109, E3128-35.
- HOGE, F. E., LYON, P. E., SWIFT, R. N., YUNGEL, J. K., ABBOTT, M. R., LETELIER, R. M. & ESAIAS, W. E. 2003. Validation of Terra-MODIS phytoplankton chlorophyll fluorescence line height. I. Initial airborne lidar results. *Appl Opt*, 42, 2767-71.
- HORI, S., NOMURA, T. & SAKAGUCHI, S. 2003. Control of regulatory T cell development by the transcription factor Foxp3. *Science*, 299, 1057-61.
- HUNG, T. & CHANG, H. Y. 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol*, 7, 582-5.
- ISHAK, C. A., MARSHALL, A. E., PASSOS, D. T., WHITE, C. R., KIM, S. J., CECCHINI, M. J., FERWATI, S., MACDONALD, W. A., HOWLETT, C. J., WELCH, I. D., RUBIN, S. M., MANN, M. R. W. & DICK, F. A. 2016. An RB-EZH2 Complex Mediates Silencing of Repetitive DNA Sequences. *Mol Cell*, 64, 1074-1087.
- JACKSON, D. A. & COOK, P. R. 1985. Transcription occurs at a nucleoskeleton. *EMBO J*, 4, 919-25.
- JACKSON, D. A. & COOK, P. R. 1986. Replication occurs at a nucleoskeleton. *EMBO J*, 5, 1403-10.
- JEON, Y. & LEE, J. T. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*, 146, 119-33.
- JONKERS, I., BARAKAT, T. S., ACHAME, E. M., MONKHORST, K., KENTER, A., RENTMEESTER, E., GROSVELD, F., GROOTEGOED, J. A. & GRIBNAU, J. 2009. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell*, 139, 999-1011.
- KALB, R., LATWIEL, S., BAYMAZ, H. I., JANSEN, P. W., MULLER, C. W., VERMEULEN, M. & MULLER, J. 2014. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol*, 21, 569-71.
- KANEKO, S., BONASIO, R., SALDANA-MEYER, R., YOSHIDA, T., SON, J., NISHINO, K., UMEZAWA, A. & REINBERG, D. 2014. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell*, 53, 290-300.

- KANEKO, S., LI, G., SON, J., XU, C. F., MARGUERON, R., NEUBERT, T. A. & REINBERG, D. 2010. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev*, 24, 2615-20.
- KHANUJA, P. S., LEHR, J. E., SOULE, H. D., GEHANI, S. K., NOTO, A. C., CHOUDHURY, S., CHEN, R. & PIENTA, K. J. 1993. Nuclear matrix proteins in normal and breast cancer cells. *Cancer Res*, 53, 3394-8.
- KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12, 357-60.
- KIMURA, K., CUVIER, O. & HIRANO, T. 2001. Chromosome condensation by a human condensin complex in *Xenopus* egg extracts. *J Biol Chem*, 276, 5417-20.
- KOHLMAIER, A., SAVARESE, F., LACHNER, M., MARTENS, J., JENUWEIN, T. & WUTZ, A. 2004. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol*, 2, E171.
- KOPP, F. & MENDELL, J. T. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, 172, 393-407.
- KOREN, A. & MCCARROLL, S. A. 2014. Random replication of the inactive X chromosome. *Genome Res*, 24, 64-9.
- KOUPRINA, N., PAVLICEK, A., COLLINS, N. K., NAKANO, M., NOSKOV, V. N., OHZEKI, J., MOCHIDA, G. H., RISINGER, J. I., GOLDSMITH, P., GUNSIOR, M., SOLOMON, G., GERSCH, W., KIM, J. H., BARRETT, J. C., WALSH, C. A., JURKA, J., MASUMOTO, H. & LARIONOV, V. 2005. The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. *Hum Mol Genet*, 14, 2155-65.
- KULASHRESHTHA, M., MEHTA, I. S., KUMAR, P. & RAO, B. J. 2016. Chromosome territory relocation during DNA repair requires nuclear myosin 1 recruitment to chromatin mediated by Upsilon-H2AX signaling. *Nucleic Acids Res*, 44, 8272-91.
- LAGARKOVA, M. A., SVETLOVA, E., GIACCA, M., FALASCHI, A. & RAZIN, S. V. 1998. DNA loop anchorage region colocalizes with the replication origin located downstream to the human gene encoding lamin B2. *J Cell Biochem*, 69, 13-8.
- LAPORTE, D., COURTOU, F., SALIN, B., CESCHIN, J. & SAGOT, I. 2013. An array of nuclear microtubules reorganizes the budding yeast nucleus during quiescence. *J Cell Biol*, 203, 585-94.
- LEBIEN, T. W. & TEDDER, T. F. 2008. B lymphocytes: how they develop and function. *Blood*, 112, 1570-80.
- LEE, J. T., DAVIDOW, L. S. & WARSHAWSKY, D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, 21, 400-4.
- LEE, K. M., CHOI, K. H. & OUELLETTE, M. M. 2004. Use of exogenous hTERT to immortalize primary human cells. *Cytotechnology*, 45, 33-8.
- LEE, S., LIU, B., LEE, S., HUANG, S. X., SHEN, B. & QIAN, S. B. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A*, 109, E2424-32.
- LEHTONEN, J. Y., DAVIET, L., NAHMIAS, C., HORIUCHI, M. & DZAU, V. J. 1999. Analysis of functional domains of angiotensin II type 2 receptor involved in apoptosis. *Mol Endocrinol*, 13, 1051-60.

- LEI, L., WU, J., GU, D., LIU, H. & WANG, S. 2016. CIZ1 interacts with YAP and activates its transcriptional activity in hepatocellular carcinoma cells. *Tumour Biol*, 37, 11073-9.
- LEMAITRE, J. M., DANIS, E., PASERO, P., VASSETZKY, Y. & MÉCHALI, M. 2005. Mitotic remodeling of the replicon and chromosome structure. *Cell*, 123, 787-801.
- LIBERT, C., DEJAGER, L. & PINHEIRO, I. 2010. The X chromosome in immune functions: when a chromosome makes the difference. *Nat Rev Immunol*, 10, 594-604.
- LIU, T., REN, X., LI, L., YIN, L., LIANG, K., YU, H., REN, H., ZHOU, W., JING, H. & KONG, C. 2015. Ciz1 promotes tumorigenicity of prostate carcinoma cells. *Front Biosci (Landmark Ed)*, 20, 705-15.
- LOHELA, M., BRY, M., TAMMELA, T. & ALITALO, K. 2009. VEGFs and receptors involved in angiogenesis versus lymphangiogenesis. *Curr Opin Cell Biol*, 21, 154-65.
- LORSCH, J. R., COLLINS, F. S. & LIPPINCOTT-SCHWARTZ, J. 2014. Cell Biology. Fixing problems with cell lines. *Science*, 346, 1452-3.
- LU, J., CHANG, P., RICHARDSON, J. A., GAN, L., WEILER, H. & OLSON, E. N. 2000. The basic helix-loop-helix transcription factor capsulin controls spleen organogenesis. *Proc Natl Acad Sci U S A*, 97, 9525-30.
- LUDERUS, M. E., DE GRAAF, A., MATTIA, E., DEN BLAAUWEN, J. L., GRANDE, M. A., DE JONG, L. & VAN DRIEL, R. 1992. Binding of matrix attachment regions to lamin B1. *Cell*, 70, 949-59.
- LV, Q., YUAN, L., SONG, Y., SUI, T., LI, Z. & LAI, L. 2016. D-repeat in the XIST gene is required for X chromosome inactivation. *RNA Biol*, 13, 172-6.
- LYON, M. F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*, 190, 372-3.
- LYON, M. F. 1971. Possible mechanisms of X chromosome inactivation. *Nat New Biol*, 232, 229-32.
- LYON, M. F. 2000. LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? *Proc Natl Acad Sci U S A*, 97, 6248-9.
- LYU, G., TAN, T., GUAN, Y., SUN, L., LIANG, Q. & TAO, W. 2018. Changes in the position and volume of inactive X chromosomes during the G0/G1 transition. *Chromosome Res*, 26, 179-189.
- MA, N., MATSUNAGA, S., MORIMOTO, A., SAKASHITA, G., URANO, T., UCHIYAMA, S. & FUKUI, K. 2011. The nuclear scaffold protein SAF-A is required for kinetochore-microtubule attachment and contributes to the targeting of Aurora-A to mitotic spindles. *J Cell Sci*, 124, 394-404.
- MAEZAWA, S., HASEGAWA, K., ALAVATTAM, K. G., FUNAKOSHI, M., SATO, T., BARSKI, A. & NAMEKAWA, S. H. 2018. SCML2 promotes heterochromatin organization in late spermatogenesis. *J Cell Sci*, 131.
- MAHERALI, N., SRIDHARAN, R., XIE, W., UTIKAL, J., EMINLI, S., ARNOLD, K., STADTFELD, M., YACHECHKO, R., TCHIEU, J., JAENISCH, R., PLATH, K. & HOCHEDLINGER, K. 2007. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell*, 1, 55-70.
- MAKHLOUF, M., OUIMETTE, J. F., OLDFIELD, A., NAVARRO, P., NEUILLET, D. & ROUGEULLE, C. 2014. A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat Commun*, 5, 4878.

- MARAHRENS, Y., PANNING, B., DAUSMAN, J., STRAUSS, W. & JAENISCH, R. 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev*, 11, 156-66.
- MARGUERON, R. & REINBERG, D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature*, 469, 343-9.
- MARKEY, M. P., BERGSEID, J., BOSCO, E. E., STENGEL, K., XU, H., MAYHEW, C. N., SCHWEMBERGER, S. J., BRADEN, W. A., JIANG, Y., BABCOCK, G. F., JEGGA, A. G., ARONOW, B. J., REED, M. F., WANG, J. Y. & KNUDSEN, E. S. 2007. Loss of the retinoblastoma tumor suppressor: differential action on transcriptional programs related to cell cycle control and immune function. *Oncogene*, 26, 6307-18.
- MARKLUND, U., LARSSON, N., GRADIN, H. M., BRATTSAND, G. & GULLBERG, M. 1996. Oncoprotein 18 is a phosphorylation-responsive regulator of microtubule dynamics. *EMBO J*, 15, 5290-8.
- MARTIN, M. 2012. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Bioinformatics in Action*, 17, 10-12.
- MARTIN, R. G. & STEIN, S. 1976. Resting state in normal and simian virus 40 transformed Chinese hamster lung cells. *Proc Natl Acad Sci U S A*, 73, 1655-9.
- MATSUOKA, S., ROTMAN, G., OGAWA, A., SHILOH, Y., TAMAI, K. & ELLEDGE, S. J. 2000. Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc Natl Acad Sci U S A*, 97, 10389-94.
- MAXWELL, C. A., BENITEZ, J., GOMEZ-BALDO, L., OSORIO, A., BONIFACI, N., FERNANDEZ-RAMIREZ, R., COSTES, S. V., GUINO, E., CHEN, H., EVANS, G. J., MOHAN, P., CATALA, I., PETIT, A., AGUILAR, H., VILLANUEVA, A., AYLES, A., SERRA-MUSACH, J., RENNERT, G., LEJBKOWICZ, F., PETERLONGO, P., MANOUKIAN, S., PEISSEL, B., RIPAMONTI, C. B., BONANNI, B., VIEL, A., ALLAVENA, A., BERNARD, L., RADICE, P., FRIEDMAN, E., KAUFMAN, B., LAITMAN, Y., DUBROVSKY, M., MILGROM, R., JAKUBOWSKA, A., CYBULSKI, C., GORSKI, B., JAWORSKA, K., DURDA, K., SUKIENNICKI, G., LUBINSKI, J., SHUGART, Y. Y., DOMCHEK, S. M., LETRERO, R., WEBER, B. L., HOGERVORST, F. B., ROOKUS, M. A., COLLEE, J. M., DEVILEE, P., LIGTENBERG, M. J., LUIJT, R. B., AALFS, C. M., WAISFISZ, Q., WIJNEN, J., ROOZENDAAL, C. E., HEBON, EMBRACE, EASTON, D. F., PEOCK, S., COOK, M., OLIVER, C., FROST, D., HARRINGTON, P., EVANS, D. G., LALLOO, F., EELES, R., IZATT, L., CHU, C., ECCLES, D., DOUGLAS, F., BREWER, C., NEVANLINNA, H., HEIKKINEN, T., COUCH, F. J., LINDOR, N. M., WANG, X., GODWIN, A. K., CALIGO, M. A., LOMBARDI, G., LOMAN, N., KARLSSON, P., EHRENCRONA, H., WACHENFELDT, A., SWE, B., BARKARDOTTIR, R. B., HAMANN, U., RASHID, M. U., LASA, A., CALDES, T., ANDRES, R., SCHMITT, M., ASSMANN, V., STEVENS, K., OFFIT, K., CURADO, J., TILGNER, H., GUIGO, R., AIZA, G., BRUNET, J., CASTELLSAGUE, J., MARTRAT, G., et al. 2011. Interplay between BRCA1 and RHMAM regulates epithelial apicobasal polarization and may influence risk of breast cancer. *PLoS Biol*, 9, e1001199.
- MCHUGH, C. A., CHEN, C. K., CHOW, A., SURKA, C. F., TRAN, C., MCDONEL, P., PANDYA-JONES, A., BLANCO, M., BURGHARD, C., MORADIAN, A., SWEREDOSKI, M. J., SHISHKIN, A. A., SU, J., LANDER, E. S., HESS, S., PLATH, K. & GUTTMAN, M. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, 521, 232-6.
- MCNEIL, S., GUO, B., STEIN, J. L., LIAN, J. B., BUSHMEYER, S., SETO, E., ATCHISON, M. L., PENMAN, S., VAN WIJNEN, A. J. & STEIN, G. S. 1998. Targeting of the YY1 transcription factor to the nucleolus and the nuclear matrix in situ: the C-terminus is a principal determinant for nuclear trafficking. *J Cell Biochem*, 68, 500-10.

- MEHTA, I. S., AMIRA, M., HARVEY, A. J. & BRIDGER, J. M. 2010. Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol*, 11, R5.
- MIGNONE, F., GISSI, C., LIUNI, S. & PESOLE, G. 2002. Untranslated regions of mRNAs. *Genome Biol*, 3, REVIEWS0004.
- MITSUI, K., MATSUMOTO, A., OHTSUKA, S., OHTSUBO, M. & YOSHIMURA, A. 1999. Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. *Biochem Biophys Res Commun*, 264, 457-64.
- MIYAMOTO, T., FURUSAWA, C. & KANEKO, K. 2015. Pluripotency, Differentiation, and Reprogramming: A Gene Expression Dynamics Model with Epigenetic Feedback Regulation. *PLoS Comput Biol*, 11, e1004476.
- MOINDROT, B., CERASE, A., COKER, H., MASUI, O., GRIJZENHOUT, A., PINTACUDA, G., SCHERMELLEH, L., NESTEROVA, T. B. & BROCKDORFF, N. 2015. A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep*, 12, 562-72.
- MONFORT, A., DI MININ, G., POSTLMAYR, A., FREIMANN, R., ARIETI, F., THORE, S. & WUTZ, A. 2015. Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Rep*, 12, 554-61.
- MONKHORST, K., DE HOON, B., JONKERS, I., MULUGETA ACHAME, E., MONKHORST, W., HOOGERBRUGGE, J., RENTMEESTER, E., WESTERHOFF, H. V., GROSVELD, F., GROOTEGOED, J. A. & GRIBNAU, J. 2009. The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. *PLoS One*, 4, e5616.
- MONKHORST, K., JONKERS, I., RENTMEESTER, E., GROSVELD, F. & GRIBNAU, J. 2008. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. *Cell*, 132, 410-21.
- MOOTHA, V. K., LINDGREN, C. M., ERIKSSON, K. F., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRALE, M., LAURILA, E., HOUSTIS, N., DALY, M. J., PATTERSON, N., MESIROV, J. P., GOLUB, T. R., TAMAYO, P., SPIEGELMAN, B., LANDER, E. S., HIRSCHHORN, J. N., ALTSHULER, D. & GROOP, L. C. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 34, 267-73.
- MUMBACH, M. R., GRANJA, J. M., FLYNN, R. A., ROAKE, C. M., SATPATHY, A. T., RUBIN, A. J., QI, Y., JIANG, Z., SHAMS, S., LOUIE, B. H., GUO, J. K., GENNERT, D. G., CORCES, M. R., KHAVARI, P. A., ATIANAND, M. K., ARTANDI, S. E., FITZGERALD, K. A., GREENLEAF, W. J. & CHANG, H. Y. 2019. HiChIRP reveals RNA-associated chromosome conformation. *Nat Methods*, 16, 489-492.
- MUNKLEY, J., COPELAND, N. A., MOIGNARD, V., KNIGHT, J. R., GREAVES, E., RAMSBOTTOM, S. A., POWNALL, M. E., SOUTHGATE, J., AINSCOUGH, J. F. & COVERLEY, D. 2011. Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells. *Nucleic Acids Res*, 39, 2671-7.
- NABATIYAN, A. & KRUDE, T. 2004. Silencing of chromatin assembly factor 1 in human cells leads to cell death and loss of chromatin assembly during DNA synthesis. *Mol Cell Biol*, 24, 2853-62.
- NABETANI, A., KOUJIN, T., TSUTSUMI, C., HARAGUCHI, T. & HIRAOKA, Y. 2001. A conserved protein, Nuf2, is implicated in connecting the centromere to the spindle

during chromosome segregation: a link between the kinetochore function and the spindle checkpoint. *Chromosoma*, 110, 322-34.

NAKAGAWA, S. & PRASANTH, K. V. 2011. eXIST with matrix-associated proteins. *Trends Cell Biol*, 21, 321-7.

NISHIBE, R., WATANABE, W., UEDA, T., YAMASAKI, N., KOLLER, R., WOLFF, L., HONDA, Z., OHTSUBO, M. & HONDA, H. 2013. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. *FEBS Lett*, 587, 1529-35.

NOJIMA, H., ADACHI, M., MATSUI, T., OKAWA, K., TSUKITA, S. & TSUKITA, S. 2008. IQGAP3 regulates cell proliferation through the Ras/ERK signalling cascade. *Nat Cell Biol*, 10, 971-8.

NORBURY, C. & NURSE, P. 1992. Animal cell cycles and their control. *Annu Rev Biochem*, 61, 441-70.

O'CARROLL, D., ERHARDT, S., PAGANI, M., BARTON, S. C., SURANI, M. A. & JENUWEIN, T. 2001. The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol*, 21, 4330-6.

O'KEEFE, R. T., HENDERSON, S. C. & SPECTOR, D. L. 1992. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J Cell Biol*, 116, 1095-110.

OBERAUER, R., RIST, W., LENTER, M. C., HAMILTON, B. S. & NEUBAUER, H. 2010. EGFL6 is increasingly expressed in human obesity and promotes proliferation of adipose tissue-derived stromal vascular cells. *Mol Cell Biochem*, 343, 257-69.

OKAMOTO, I., OTTE, A. P., ALLIS, C. D., REINBERG, D. & HEARD, E. 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science*, 303, 644-9.

PAN, C., KUMAR, C., BOHL, S., KLINGMUELLER, U. & MANN, M. 2009. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics*, 8, 443-50.

PARDEE, A. B. 1974. A restriction point for control of normal animal cell proliferation. *Proc Natl Acad Sci U S A*, 71, 1286-90.

PARK, H. D., KIM, Y. K., PARK, K. U., KIM, J. Q., SONG, Y. H. & SONG, J. 2009. A novel c.-22T>C mutation in GALK1 promoter is associated with elevated galactokinase phenotype. *BMC Med Genet*, 10, 29.

PASINI, D., BRACKEN, A. P., JENSEN, M. R., LAZZERINI DENCHI, E. & HELIN, K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J*, 23, 4061-71.

PAULEY, R. J., PAINE, T. J. & SOULE, H. D. 1993. *Immortal Human Mammary Epithelial Cell Sublines*.

PAYER, B. & LEE, J. T. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet*, 42, 733-72.

PINTACUDA, G., WEI, G., ROUSTAN, C., KIRMIZITAS, B. A., SOLCAN, N., CERASE, A., CASTELLO, A., MOHAMMED, S., MOINDROT, B., NESTEROVA, T. B. & BROCKDORFF, N. 2017. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell*, 68, 955-969 e10.

- PRESCOTT, J. E., OSTHUS, R. C., LEE, L. A., LEWIS, B. C., SHIM, H., BARRETT, J. F., GUO, Q., HAWKINS, A. L., GRIFFIN, C. A. & DANG, C. V. 2001. A novel c-Myc-responsive gene, JPO1, participates in neoplastic transformation. *J Biol Chem*, 276, 48276-84.
- PUJANA, M. A., HAN, J. D., STARITA, L. M., STEVENS, K. N., TEWARI, M., AHN, J. S., RENNERT, G., MORENO, V., KIRCHHOFF, T., GOLD, B., ASSMANN, V., ELSHAMY, W. M., RUAL, J. F., LEVINE, D., ROZEK, L. S., GELMAN, R. S., GUNSALUS, K. C., GREENBERG, R. A., SOBHIAN, B., BERTIN, N., VENKATESAN, K., AYIVI-GUEDEHOUSOU, N., SOLE, X., HERNANDEZ, P., LAZARO, C., NATHANSON, K. L., WEBER, B. L., CUSICK, M. E., HILL, D. E., OFFIT, K., LIVINGSTON, D. M., GRUBER, S. B., PARVIN, J. D. & VIDAL, M. 2007. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet*, 39, 1338-49.
- QIAO, F., MOSS, A. & KUPFER, G. M. 2001. Fanconi anemia proteins localize to chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *J Biol Chem*, 276, 23391-6.
- RAAPHORST, F. M. 2005. Deregulated expression of Polycomb-group oncogenes in human malignant lymphomas and epithelial tumors. *Hum Mol Genet*, 14 Spec No 1, R93-R100.
- RAHMAN, F., AINSCOUGH, J. F., COPELAND, N. & COVERLEY, D. 2007. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. *Hum Mutat*, 28, 993-1004.
- RAHMAN, F. A., AZIZ, N. & COVERLEY, D. 2010. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. *BMC Cancer*, 10, 482.
- RICHLY, H., ALOIA, L. & DI CROCE, L. 2011. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis*, 2, e204.
- RIDINGS-FIGUEROA, R., STEWART, E. R., NESTEROVA, T. B., COKER, H., PINTACUDA, G., GODWIN, J., WILSON, R., HASLAM, A., LILLEY, F., RUIGROK, R., BAGEGHNI, S. A., ALBADRANI, G., MANSFIELD, W., ROULSON, J. A., BROCKDORFF, N., AINSCOUGH, J. F. X. & COVERLEY, D. 2017. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev*.
- ROSHON, M. J. & RULEY, H. E. 2005. Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res*, 14, 179-92.
- ROSS, M. T., GRAFHAM, D. V., COFFEY, A. J., SCHERER, S., MCLAY, K., MUZNY, D., PLATZER, M., HOWELL, G. R., BURROWS, C., BIRD, C. P., FRANKISH, A., LOVELL, F. L., HOWE, K. L., ASHURST, J. L., FULTON, R. S., SUDBRAK, R., WEN, G., JONES, M. C., HURLES, M. E., ANDREWS, T. D., SCOTT, C. E., SEARLE, S., RAMSER, J., WHITTAKER, A., DEADMAN, R., CARTER, N. P., HUNT, S. E., CHEN, R., CREE, A., GUNARATNE, P., HAVLAK, P., HODGSON, A., METZKER, M. L., RICHARDS, S., SCOTT, G., STEFFEN, D., SODERGREN, E., WHEELER, D. A., WORLEY, K. C., AINSCOUGH, R., AMBROSE, K. D., ANSARI-LARI, M. A., ARADHYA, S., ASHWELL, R. I., BABBAGE, A. K., BAGGULEY, C. L., BALLABIO, A., BANERJEE, R., BARKER, G. E., BARLOW, K. F., BARRETT, I. P., BATES, K. N., BEARE, D. M., BEASLEY, H., BEASLEY, O., BECK, A., BETHEL, G., BLECHSCHMIDT, K., BRADY, N., BRAY-ALLEN, S., BRIDGEMAN, A. M., BROWN, A. J., BROWN, M. J., BONNIN, D., BRUFORD, E. A., BUHAY, C., BURCH, P., BURFORD, D., BURGESS, J., BURRILL, W., BURTON, J., BYE, J. M., CARDER, C., CARREL, L., CHAKO, J., CHAPMAN, J. C., CHAVEZ, D., CHEN, E., CHEN, G., CHEN, Y., CHEN, Z., CHINAULT, C., CICCOCICOLA, A., CLARK, S. Y., CLARKE,

- G., CLEE, C. M., CLEGG, S., CLERC-BLANKENBURG, K., CLIFFORD, K., COBLEY, V., COLE, C. G., CONQUER, J. S., CORBY, N., CONNOR, R. E., DAVID, R., DAVIES, J., DAVIS, C., DAVIS, J., DELGADO, O., DESHAZO, D., et al. 2005. The DNA sequence of the human X chromosome. *Nature*, 434, 325-37.
- ROUX, K. J., KIM, D. I. & BURKE, B. 2013. BioID: a screen for protein-protein interactions. *Curr Protoc Protein Sci*, 74, Unit 19 23.
- SADASIVAM, S. & DECAPRIO, J. A. 2013. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer*, 13, 585-95.
- SADO, T. & BROCKDORFF, N. 2013. Advances in understanding chromosome silencing by the long non-coding RNA Xist. *Philos Trans R Soc Lond B Biol Sci*, 368, 20110325.
- SATO, T., KANEDA, A., TSUJI, S., ISAGAWA, T., YAMAMOTO, S., FUJITA, T., YAMANAKA, R., TANAKA, Y., NUKIWA, T., MARQUEZ, V. E., ISHIKAWA, Y., ICHINOSE, M. & ABURATANI, H. 2013. PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. *Sci Rep*, 3, 1911.
- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9, 676-82.
- SCHOEFTNER, S., SENGUPTA, A. K., KUBICEK, S., MECHTLER, K., SPAHN, L., KOSEKI, H., JENUWEIN, T. & WUTZ, A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J*, 25, 3110-22.
- SCHWARTZ, Y. B. & PIRROTTA, V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet*, 14, 853-64.
- SELLERS, W. R. & KAELIN, W. G., JR. 1997. Role of the retinoblastoma protein in the pathogenesis of human cancer. *J Clin Oncol*, 15, 3301-12.
- SENNER, C. E., NESTEROVA, T. B., NORTON, S., DEWCHAND, H., GODWIN, J., MAK, W. & BROCKDORFF, N. 2011. Disruption of a conserved region of Xist exon 1 impairs Xist RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation. *Development*, 138, 1541-50.
- SHARMAN, G. B. 1971. Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature*, 230, 231-2.
- SHIN, J., BOSSENZ, M., CHUNG, Y., MA, H., BYRON, M., TANIGUCHI-ISHIGAKI, N., ZHU, X., JIAO, B., HALL, L. L., GREEN, M. R., JONES, S. N., HERMANS-BORGMEYER, I., LAWRENCE, J. B. & BACH, I. 2010. Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature*, 467, 977-81.
- SHIRATORI, A., OKUMURA, K., NOGAMI, M., TAGUCHI, H., ONOZAKI, T., INOUE, T., ANDO, T., SHIBATA, T., IZUMI, M., MIYAZAWA, H. & ET AL. 1995. Assignment of the 49-kDa (PRIM1) and 58-kDa (PRIM2A and PRIM2B) subunit genes of the human DNA primase to chromosome bands 1q44 and 6p11.1-p12. *Genomics*, 28, 350-3.
- SHUKUNAMI, C., YOSHIMOTO, Y., TAKIMOTO, A., YAMASHITA, H. & HIRAKI, Y. 2016. Molecular characterization and function of tenomodulin, a marker of tendons and ligaments that integrate musculoskeletal components. *Jpn Dent Sci Rev*, 52, 84-92.
- SIMON, J. A. 2009. Transcription. Sweet silencing. *Science*, 325, 45-6.

- SMEETS, D., MARKAKI, Y., SCHMID, V. J., KRAUS, F., TATTERMUSCH, A., CERASE, A., STERR, M., FIEDLER, S., DEMMERLE, J., POPKEN, J., LEONHARDT, H., BROCKDORFF, N., CREMER, T., SCHERMELLEH, L. & CREMER, M. 2014. Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics Chromatin*, 7, 8.
- SMITH, J. A. & MARTIN, L. 1973. Do cells cycle? *Proc Natl Acad Sci U S A*, 70, 1263-7.
- STADTFELD, M. & HOCHEDLINGER, K. 2010. Induced pluripotency: history, mechanisms, and applications. *Genes Dev*, 24, 2239-63.
- STARMER, J. & MAGNUSON, T. 2009. A new model for random X chromosome inactivation. *Development*, 136, 1-10.
- STEWART, E. R. & COVERLEY, D. 2018. Visualization of Hidden Epitopes at the Inactive X Chromosome. *Methods Mol Biol*, 1861, 103-112.
- STEWART, E. R., TURNER, R. M. L., NEWLING, K., RIDINGS-FIGUEROA, R., SCOTT, V., ASHTON, P. D., AINSCOUGH, J. F. X. & COVERLEY, D. 2019. Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture. *Nat Commun*, 10, 460.
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102, 15545-50.
- SUN, Y., KUCEJ, M., FAN, H. Y., YU, H., SUN, Q. Y. & ZOU, H. 2009. Separase is recruited to mitotic chromosomes to dissolve sister chromatid cohesion in a DNA-dependent manner. *Cell*, 137, 123-32.
- SUNWOO, H., COLOGNORI, D., FROBERG, J. E., JEON, Y. & LEE, J. T. 2017. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc Natl Acad Sci U S A*, 114, 10654-10659.
- SURYADINATA, R., SADOWSKI, M. & SARCEVIC, B. 2010. Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates. *Biosci Rep*, 30, 243-55.
- SWARTS, D. R. A., STEWART, E. R., HIGGINS, G. S. & COVERLEY, D. 2018. CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours. *Cell Cycle*, 17, 2268-2283.
- SYRETT, C. M., PANERU, B., SANDOVAL-HEGLUND, D., WANG, J., BANERJEE, S., SINDHAVA, V., BEHRENS, E. M., ATCHISON, M. & ANGUERA, M. C. 2019. Altered X-chromosome inactivation in T cells may promote sex-biased autoimmune diseases. *JCI Insight*, 4.
- TADA, T., NAGATA, S. & HIRANO, K. 2014. *Reprogrammed Stem Cell*. PCT/JP2013/08535.
- TAKAGI, N. & SASAKI, M. 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, 256, 640-2.
- TAVARES, L., DIMITROVA, E., OXLEY, D., WEBSTER, J., POOT, R., DEMMERS, J., BEZSTAROSTI, K., TAYLOR, S., URA, H., KOIDE, H., WUTZ, A., VIDAL, M., ELDERKIN, S. & BROCKDORFF, N. 2012. RYBP-PRC1 complexes mediate H2A

- ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell*, 148, 664-78.
- TOWNSLEY, F. M., ARISTARKHOV, A., BECK, S., HERSHKO, A. & RUDERMAN, J. V. 1997. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. *Proc Natl Acad Sci U S A*, 94, 2362-7.
- TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7, 562-78.
- VAN DER LUGT, N. M., DOMEN, J., LINDERS, K., VAN ROON, M., ROBANUS-MAANDAG, E., TE RIELE, H., VAN DER VALK, M., DESCHAMPS, J., SOFRONIEW, M., VAN LOHUIZEN, M. & ET AL. 1994. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev*, 8, 757-69.
- VARMA, P. & MISHRA, R. K. 2011. Dynamics of nuclear matrix proteome during embryonic development in *Drosophila melanogaster*. *J Biosci*, 36, 439-59.
- VERHEIJEN, R., VAN VENROOIJ, W. & RAMAEKERS, F. 1988. The nuclear matrix: structure and composition. *J Cell Sci*, 90 (Pt 1), 11-36.
- VERVOORT, V. S., BEACHEM, M. A., EDWARDS, P. S., LADD, S., MILLER, K. E., DE MOLLERAT, X., CLARKSON, K., DUPONT, B., SCHWARTZ, C. E., STEVENSON, R. E., BOYD, E. & SRIVASTAVA, A. K. 2002. AGTR2 mutations in X-linked mental retardation. *Science*, 296, 2401-3.
- VONCKEN, J. W., ROELEN, B. A., ROEFS, M., DE VRIES, S., VERHOEVEN, E., MARINO, S., DESCHAMPS, J. & VAN LOHUIZEN, M. 2003. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A*, 100, 2468-73.
- WAN, K. M., NICKERSON, J. A., KROCKMALNIC, G. & PENMAN, S. 1999. The nuclear matrix prepared by amine modification. *Proc Natl Acad Sci U S A*, 96, 933-8.
- WANG, D. Q., WANG, K., YAN, D. W., LIU, J., WANG, B., LI, M. X., WANG, X. W., LIU, J., PENG, Z. H., LI, G. X. & YU, Z. H. 2014. Ciz1 is a novel predictor of survival in human colon cancer. *Exp Biol Med (Maywood)*, 239, 862-870.
- WANG, J., SYRETT, C. M., KRAMER, M. C., BASU, A., ATCHISON, M. L. & ANGUERA, M. C. 2016. Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for increased expression from the inactive X. *Proc Natl Acad Sci U S A*, 113, E2029-38.
- WANG, R., DILLON, C. P., SHI, L. Z., MILASTA, S., CARTER, R., FINKELSTEIN, D., MCCORMICK, L. L., FITZGERALD, P., CHI, H., MUNGER, J. & GREEN, D. R. 2011. The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity*, 35, 871-82.
- WATERS, L. R., AHSAN, F. M., WOLF, D. M., SHIRIHAI, O. & TEITELL, M. A. 2018. Initial B Cell Activation Induces Metabolic Reprogramming and Mitochondrial Remodeling. *iScience*, 5, 99-109.
- WILLIAMS, B. C., LI, Z., LIU, S., WILLIAMS, E. V., LEUNG, G., YEN, T. J. & GOLDBERG, M. L. 2003. Zwilch, a new component of the ZW10/ROD complex required for kinetochore functions. *Mol Biol Cell*, 14, 1379-91.
- WILSON, R. H. & COVERLEY, D. 2013. Relationship between DNA replication and the nuclear matrix. *Genes Cells*, 18, 17-31.

- WILSON, R. H., HESKETH, E. L. & COVERLEY, D. 2016a. The Nuclear Matrix: Fractionation Techniques and Analysis. *Cold Spring Harb Protoc*, 2016, pdb top074518.
- WILSON, R. H., HESKETH, E. L. & COVERLEY, D. 2016b. Preparation of the Nuclear Matrix for Parallel Microscopy and Biochemical Analyses. *Cold Spring Harb Protoc*, 2016, pdb prot083758.
- WILSON, R. H. C. & COVERLEY, D. 2017. Transformation-induced changes in the DNA-nuclear matrix interface, revealed by high-throughput analysis of DNA halos. *Sci Rep*, 7, 6475.
- WITTMANN, T., WILM, M., KARSENTI, E. & VERNOS, I. 2000. TPX2, A novel xenopus MAP involved in spindle pole organization. *J Cell Biol*, 149, 1405-18.
- WU, J., LEI, L., GU, D., LIU, H. & WANG, S. 2016. CIZ1 is upregulated in hepatocellular carcinoma and promotes the growth and migration of the cancer cells. *Tumour Biol*, 37, 4735-42.
- WUTZ, A., RASMUSSEN, T. P. & JAENISCH, R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*, 30, 167-74.
- XIAO, J., UITTI, R. J., ZHAO, Y., VEMULA, S. R., PERLMUTTER, J. S., WSZOLEK, Z. K., MARAGANORE, D. M., AUBURGER, G., LEUBE, B., LEHNHOFF, K. & LEDOUX, M. S. 2012. Mutations in CIZ1 cause adult onset primary cervical dystonia. *Ann Neurol*, 71, 458-69.
- YAMADA, N., HASEGAWA, Y., YUE, M., HAMADA, T., NAKAGAWA, S. & OGAWA, Y. 2015. Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome. *PLoS Genet*, 11, e1005430.
- YAMAGISHI, M. & UCHIMARU, K. 2017. Targeting EZH2 in cancer therapy. *Curr Opin Oncol*, 29, 375-381.
- YAMAGUCHI, H. & HUNG, M. C. 2014. Regulation and Role of EZH2 in Cancer. *Cancer Res Treat*, 46, 209-22.
- YANG, F., BABAK, T., SHENDURE, J. & DISTECHE, C. M. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res*, 20, 614-22.
- YANG, F., DENG, X., MA, W., BERLETCH, J. B., RABAIA, N., WEI, G., MOORE, J. M., FILIPPOVA, G. N., XU, J., LIU, Y., NOBLE, W. S., SHENDURE, J. & DISTECHE, C. M. 2015. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol*, 16, 52.
- YAO, Y. & DAI, W. 2012. Mitotic checkpoint control and chromatin remodeling. *Front Biosci (Landmark Ed)*, 17, 976-83.
- YILDIRIM, E., KIRBY, J. E., BROWN, D. E., MERCIER, F. E., SADREYEV, R. I., SCADDEN, D. T. & LEE, J. T. 2013. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell*, 152, 727-42.
- YIN, J., WANG, C., TANG, X., SUN, H., SHAO, Q., YANG, X. & QU, X. 2013. CIZ1 regulates the proliferation, cycle distribution and colony formation of RKO human colorectal cancer cells. *Mol Med Rep*, 8, 1630-4.
- YOO, K. H. & HENNIGHAUSEN, L. 2012. EZH2 methyltransferase and H3K27 methylation in breast cancer. *Int J Biol Sci*, 8, 59-65.

- YU, Y., WANG, X. Y., SUN, L., WANG, Y. L., WAN, Y. F., LI, X. Q. & FENG, Y. M. 2014. Inhibition of KIF22 suppresses cancer cell proliferation by delaying mitotic exit through upregulating CDC25C expression. *Carcinogenesis*, 35, 1416-25.
- YUSUF, I. & FRUMAN, D. A. 2003. Regulation of quiescence in lymphocytes. *Trends Immunol*, 24, 380-6.
- ZEITLIN, S., PARENT, A., SILVERSTEIN, S. & EFSTRATIADIS, A. 1987. Pre-mRNA splicing and the nuclear matrix. *Mol Cell Biol*, 7, 111-20.
- ZETTERBERG, A. & LARSSON, O. 1985. Kinetic analysis of regulatory events in G1 leading to proliferation or quiescence of Swiss 3T3 cells. *Proc Natl Acad Sci U S A*, 82, 5365-9.
- ZHANG, D., WANG, Y., DAI, Y., WANG, J., SUO, T., PAN, H., LIU, H., SHEN, S. & LIU, H. 2015. CIZ1 promoted the growth and migration of gallbladder cancer cells. *Tumour Biol*, 36, 2583-91.
- ZHANG, F., WEN, Y. & GUO, X. 2014. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet*, 23, R40-6.
- ZHANG, L. F., HUYNH, K. D. & LEE, J. T. 2007. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, 129, 693-706.
- ZHAO, J., SUN, B. K., ERWIN, J. A., SONG, J. J. & LEE, J. T. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, 322, 750-6.