Role of nuclear matrix protein CIZ1 in X chromosome inactivation

Rebeca Ridings Figueroa

MSc (by research)

University of York

Biology

September 2016

ABSTRACT

X chromosome inactivation (XCI) is a gene dosage compensation mechanism that results in transcriptional silencing of one of the two X chromosomes in female cells. Polycomb repressive complexes PRC1 and PRC2 play essential roles in the process, through deposition of repressive epigenetic marks H2AK119ub and H3K27me3, respectively. Nuclear matrix protein CIZ1, which has established links with diseases such as cancer, plays a role in initiation of mammalian DNA replication through interaction with cyclin-dependent kinases. Recently it has also been observed to localize to the inactive X chromosome and loss of the protein results in disruption of the long non-coding RNA *Xist*. In mice, loss of CIZ1 leads to development of a lymphoma-like phenotype in female mice. My project has used RNAseq analysis, immune-detection techniques, and transfections to investigate the relationship between CIZ1 and X inactivation. I present evidence that implicates the C-terminus end of CIZ1 in binding at the Xi region and the N-terminus in recruitment of the protein to the Xi. I have also shown that loss of CIZ1 leads to disruption of the H3K27me3 at the Xi. At the gene expression level, absence of CIZ1 does not lead to any evident gene reactivation from the inactive X or changes in expression of genes at the X inactivation centre (Xic). However, whole-genome analysis showed that loss of CIZ1 does in fact lead to highly significant changes in the expression of PRC1 and PRC2 target genes, while leaving PRC1 and PRC2 subunits unaffected. This data suggests that loss of CIZ1 causes widespread changes in gene expression, which could underlie the lymphoma-like phenotype in female mice, and begins to suggest a mechanism of action for CIZ1 that involves recruitment of template to sites of polycomb action.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 5

CHAPTER 5

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

ACKNOWLEDGEMENTS

This research was carried out with support of the University of York, which provided funding for the RNA sequencing work, and the Radhika Scholarship Fund, who provided me with the necessary funds to complete the last few months of my Masters. I would also like to express my gratitude to Sally James, Dr. Richard Randle and Dr. Peter Ashton for their technical support for the RNA sequencing and bioinformatics work.

I would like to thank my supervisor Dr. Dawn Coverley for her constant support, enthusiasm and great advice. I would also like to thank Dr. Justin Ainscough for his technical assistance and insightful discussions. I have thoroughly enjoyed my time in the Coverley lab and would like to thank all of its members, present and past, for such an enjoyable experience.

Most importantly, I would like to thank my parents and sisters, for their unconditional love, support and patience.

AUTHOR'S DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. All the research in this thesis is my own, with the following exceptions:

All primary embryonic cell lines (PEFs) were generated by Justin Ainscough.

Chapter 3:
All GFP-CIZ1 mouse constructs were provided by Dr. Justin Ainscough (Coverley lab) Human GFP-CIZ1 constructs were provided by Dr. Dorian Swarts (Coverley lab) and Gillian Higgins (Coverley lab).

Chapter 4:
Reagents for nuclear matrix extraction protocol of cells on coverslips, previous to immunofluorescence, were kindly provided by Emma Stewart.

Chapter 5:
Preparation of cDNA libraries for sequencing was carried out by Sally James (Technology Facility, University of York)

Illumina Sequencing was done at the Leeds Institute for Molecular Medicine (LIMM) at St. James's University Hospital in Leeds.

Initial bioinformatics processing, including alignment of reads to the mouse genome, transcriptome assembly and differential gene analysis was carried out by Dr. Richard Randle Boggis with support of Dr. Peter Ashton (Technology Facility, University of York).

CHAPTER 1. INTRODUCTION

1.1 X chromosome inactivation

X chromosome inactivation (XCI) is the mechanism by which one of the two X chromosomes in female cells is inactivated. This process evolved in eutherian mammals in order to establish a gene dosage compensation mechanism between the sexes (Gribnau and Grootegoed, 2012); females cells have two X chromosomes (XX) while male cells have only one (XY). A key element in the process of XCI is the 17kb long non-coding RNA (lncRNA) molecule known as *Xist*, which is expressed from a locus on the X chromosome known as the X inactivation centre (Xic). This locus also contains other genetic elements that are necessary for XCI (Galupa and Heard, 2015).

Two types of XCI are observed in mice (Pollex and Heard, 2012). The first one is known as imprinted XCI and occurs early in female embryogenesis, transcriptionally silencing the paternally derived X chromosome (Xp). During the blastula stage, the Xp is reactivated and this is immediately followed by random XCI (rXCI), the second type of XCI. During this stage either the maternal (Xm) or paternal X (Xp) chromosome can be inactivated, and the choice is thought to be random. Following random XCI the same Xi is propagated in subsequent cell divisions, and is clonally maintained. *Xist* is crucial for XCI in humans as well as mouse, but in humans the choice of which X chromosome is inactivated is thought to occur at random also throughout development (Moreira de Mello et al, 2010).

Mechanistically, XCI is triggered following expression of *Xist* from the future inactive X chromosome (Xi), followed by accumulation in cis and recruitment of other factors, such as Polycomb Repressive Complex 2 (PRC2) and 1 (PRC1), and DNA-methyltransferases that contribute to silencing (Cerase et al, 2015). Evidence from several studies show that *Xist* contains different conserved repeat domains named A to F (Figure 1.1), which are suggested to have distinct roles (Sado and Brockdorff, 2012) (Galupa and Heard, 2015). Repeats B and F have an essential role in recruitment of PRC2 while A-repeat is essential for silencing but not PRC2 recruitment (da Rocha et al, 2014). Repeat C on the other hand is thought to be

important for association of *Xist* RNA to the X chromosome (Sado and Brockdorff, 2012). Function of the E-repeats remains unknown.

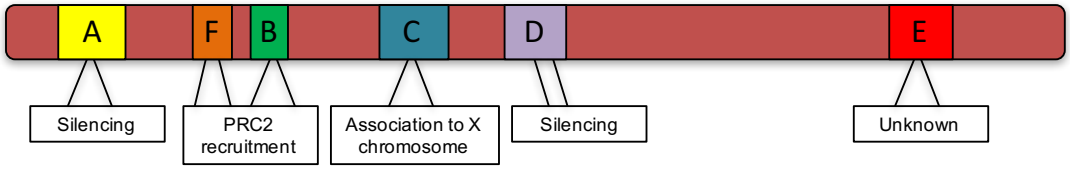


Figure 1.1. Schematic of *Xist* long non-coding RNA. Schematic shows the repeats present in *Xist* with an indication of their function.

Despite XCI, a small percentage of genes from the inactive X chromosome, known as 'escape genes' remain transcriptionally active. In the mouse, this set of genes represents approximately 3.3% (Yang et al, 2010), while in humans they represent a much larger percentage of 15% (Carrel and Willard, 2005). These genes are thought to have female-specific functions, working as factors that contribute to sexual dimorphism (Berletch et al., 2011).

1.1.1 Initiation

Several key elements are necessary for both the initiation and maintenance of X chromosome inactivation, and likely more remain undiscovered. Initiation of random XCI begins with the expression of 17kb *Xist* lncRNA, which occurs following downregulation of pluripotency factors such as octamer-binding transcription factor 4 (*Oct4*) and Kruppel Like Factor 4 (*Klf4*). Evidence shows that these pluripotency factors can bind to and repress *Xist* expression while positively regulating *Tsix* (Jiang et al, 2014; Navarro et al, 2008). *Tsix* is a non-coding gene, which is also located in the Xic and is transcribed antisense across Xist, suppressing its transcription (Lee et al, 1999). Another two transcribed loci from the X inactivation centre, *Jpx* and *Ftx*, as well as *Xpr* (Barakat et al, 2014), are thought to positively regulate *Xist*, although the mechanism is unknown.

Inactivation of the X chromosome is also dependent on a set of trans-acting factors. One of the most crucial factors is LIM Domain-Interacting RING Finger Protein (*Rnf12/RLIM*), which acts as an activator of XCI and works in a dosage-dependent manner (Jonkers et al, 2009). Evidence also points at a crucial role for this factor in imprinted XCI in mice, where maternal germ line deletion of *Rnf12* leads to lethality of female embryos due to defective XCI (Shin et al., 2010). *Jpx*, *Ftx* and *Xpr* cooperate

with and lower the threshold in order for *Rnf12* to activate *Xist* and overcome repression by *Tsix* (Barakat et al, 2014).

The mechanism by which the future inactive X chromosome is chosen during random XCI remains unclear despite intense research. Early evidence suggests the existence of an initial 'counting' mechanism by which the cell is able to determine the number of X chromosomes relative to autosomes (Lyon, 1972). This is followed by the 'choice' of which X chromosome will undergo inactivation and which one will remain active. Several different hypotheses that attempt to explain this phenomenon have been proposed (Starmer and Magnuson, 2009), with some of the most key ones being described below. The 'blocking factor' model suggests that an unknown molecule named the blocking factor, and produced by autosomes, binds to one of the X chromosomes and prevents it from being inactivated, while the active state of the remaining degrades, rendering them inactive (Lyon, 1971). The 'two factor model' builds upon the previous model, but proposes the existence of a competence factor, which is produced by both X chromosomes (Lee, 2005). The blocking factor binds one of the two competence factors, neutralizing it and rendering that X chromosome active, while the other chromosome is inactivated. In the 'sensing and choice model' the process of counting and choice is regulated by elements of the X inactivation center (Augui et al, 2007). A mechanism of 'pairing' that occurs between the two X chromosomes, through the Xpr region, is suggested to play a role in the choice step by allowing persistent *Tsix* expression on the future active X chromosome and thus repressing *Xist* expression. On the future inactive X chromosome on the other hand, *Tsix* expression is downregulated, which leads to expression of *Xist* (Stavropoulos et al, 2005). Overall the process of X chromosome inactivation has been described as a complex mechanism where both cis- and trans- regulators control *Xist* expression and activation (Galupa and Heard, 2015).

1.1.2 Spreading and recruitment of polycomb repressive complexes

Following choice and counting, *Xist* is expressed from the future inactive X chromosome. *Xist* transcripts spread in cis- and coat the X chromosome, leading to the formation of transcriptionally inactive heterochromatin. Several different chromatin modifications are observed at the inactive X. Polycomb repressive complexes (PRC) (Figure 1.2) are responsible for some of the characteristic

chromatin modifications: PRC1 is responsible for ubiquitination of lysine 119 at histone H2A (H2A-K119ub) and PRC2 is responsible for trimethylation of lysine 27 in histone 3 (H3K27me3) (Chen et al, 2004). PRC1 and PRC2 can be composed by several different subunits, which give rise to PRC1 and PRC2 variants. However, the main subunits, which form what is known as canonical PRC1 and PRC2, are indicated in Figure 1.2. Despite intense research into this area, the mechanisms underlying target specificity and action of these complexes are still incompletely understood. Recruitment of PRC2 by *Xist* is one of the earliest events occurring during X inactivation. Alpha Thalassemia/Mental Retardation Syndrome X-Linked protein (ATRX), an RNA-binding protein, has been recently identified as a necessary factor for the binding of PRC2 to *Xist* as well as other polycomb targets (Lee et al, 2014). Other proteins such as the Jumonji And AT-Rich Interaction Domain Containing 2 (JARID2), from the Jumonji family of proteins has also been described as cofactors involved in the recruitment of PRC2 (da Rocha et al, 2014). Initial observations that the PRC1 is able to recognize H3K27me3 via the chromodomain of its chromobox homolog (CBX) subunit, led to the idea that PRC2 is recruited first and that, following trymethylation, PRC1 is subsequently recruited (Schwartz and Pirrotta, 2014). However, recent studies analysing the general mechanism of polycomb recruitment in the cell suggest that it is variant PRC1, composed by Really Interesting New Gene 1 Protein A and B, (RING1A/B), polycomb group ring finger proteins 1 to 6 (PCGF1-6) and RING1 and YY1 Binding Protein (RYBP), which drives recruitment of PRC2 in a H2AK119ub1 dependent manner and that this is followed by recruitment of canonical PRC1 (Blackledge et al, 2014). These studies also showed that disruption of PRC1 mediated H2AK119ub1 also leads to disruption of PRC2 binding genome-wide (Schwartz and Pirrotta, 2014).

The study of these polycomb repressive complexes and their functions becomes of even greater importance when considering the fact that they have been implicated in regulatory pathways that are known to be disrupted in cancer (Jones and Baylin, 2011). Overexpression of polycomb repressive complexes has been linked with several types of cancer, including lung cancer (Sato et al, 2013), breast cancer (Yoo et al, 2012) and lymphoma (Raaphorst, 2005).

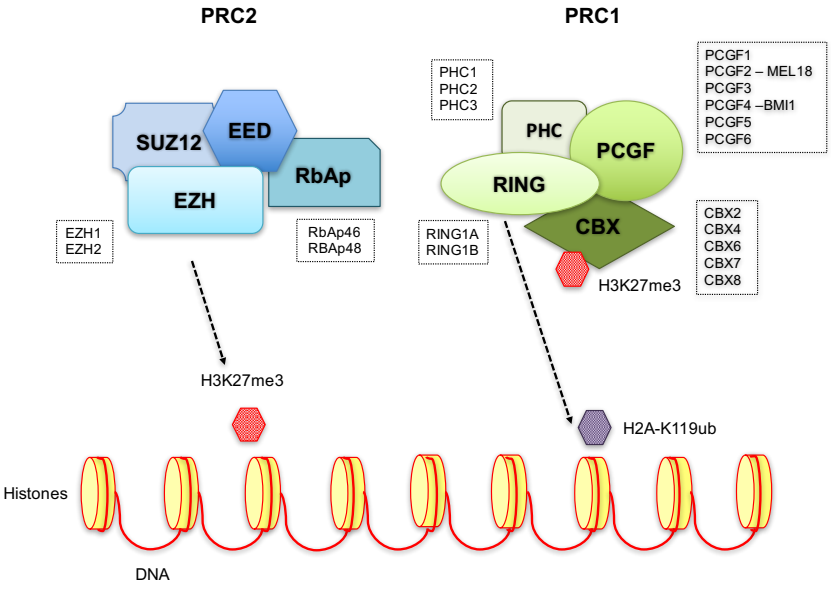


Figure 1.2. Schematic of the composition and function of PRC1 and PRC2. Canonical PRC2 contains the embryonic ectoderm development (EED) and suppressor of zeste 12 homolog (SUZ12) subunits, the histone-binding protein RbAP46/48 and the enhancer of zeste 2 (EZH) subunit, which catalyses the methylation at lysine 27 of histone H3 (H3K27me3). Canonical PRC1 includes RING finger proteins RING1A/B and the PCGF proteins MEL18 (PCGF2) or BMI1 (PCGF4). They also contain chromobox-containing protein (CBX) subunits, which recognize H3K27me3, and polyhomeotic (PHC) subunits.

It is not yet clear how the XCI signal spreads along the chromosome, however it has been proposed that long interspersed repetitive sequence elements (LINEs) could have a role in this process (Lyon, 2000). It has been observed that ectopic expression of *Xist* in autosomes leads to less efficient spreading of *Xist* and less efficient silencing. This suggested the existence of elements in the X chromosome, referred to as 'boosters', that somehow increase the efficiency of the spreading of the inactivation signal during XCI (Lyon, 2003). Although enriched in the X chromosome, LINES are also present in autosomes, consistent with the fact that autosomes are also capable of localized inactivation. In line with this hypothesis the density of LINEs appears to be highest in regions near the Xic and lowest in regions where genes are known to escape inactivation (Lyon, 2003).

1.1.3 Maintenance

Following random XCI, the inactive X chromosome enters a maintenance phase, during which different mechanisms maintain the silent state of the Xi. This phase occurs beyond mouse embryonic days 5.5 (Yildirim et al, 2013).

Despite the continued expression of *Xist*, it appears to be dispensable for maintenance of the Xi (Brown and Willard, 1994) (Csankovszki et al, 1999), as loss does not cause reactivation of gene expression. However, more recent work suggests a novel role for both *Xist* lncRNA and the perinucleolar ring in maintenance of the Xi (Zhang et al, 2007). This study provided evidence that following X-X pairing, the Xi migrates to the perinucleolar region by an unknown mechanism, and that this direct association is also necessary during mid-to-late S phase to ensure that genes from the Xi are continuously and faithfully repressed. In addition, results showed that *Xist* is in fact essential for targeting of the Xi to the perinucleolar compartment and that loss of *Xist*, and with it loss of perinucleolar localization, leads to loss of the H3K27 methylation mark and gene reactivation at the single gene level (Zhang et al, 2007). Another player that has been linked to the movement of the Xi to the perinucleolar ring is the long non-coding RNA Functional Intergenic Repeating RNA Element (*Firre)*, an X-linked gene. Knockdown of *Firre* leads to disruption of the targeting of the Xi to the nucleolus as well as loss of the H3K27 methylation mark (Yang et al, 2015).

Recent work has uncovered yet another new role for *Xist* RNA in *in vivo* suppression of cancer (Yildirim et al, 2013). In this study, loss of *Xist* in the blood compartment of mice was observed to cause changes in gene expression and subsequently, development of hematological cancer in female mice. This finding, together with other observations such as the loss of the inactive X in breast cancer cells (Sirchia et al, 2005), point towards a possible link between the X chromosome inactivation and cancer.

1.2 *Xist* and the nuclear matrix

One of the most interesting questions about the *Xist* lncRNA is the mechanism that allows it to remain in the nucleus, avoiding nuclear export and specifically coating the inactive X chromosome. Despite intensive effort this question remains largely unanswered.

A number of different players have been reported to localize to the inactive X chromosome, both during initiation and maintenance of XCI. Recent studies have uncovered a possible role of the nuclear matrix protein Scaffold Attachment Factor A

(SAF-A) in localization of the *Xist* RNA to the inactive X chromosome (Hasegawa et al, 2010). The nuclear matrix, also referred to as nuclear scaffold (Mirkovitch et al, 1984) or nucleoskeleton (Jackson and Cook, 1988) was first described in the 1960s as the non-chromatin structures that can be observed in the nucleus (Fawcett, 1966), following observation of extracted cells under an electron microscope. It is currently described as an insoluble proteinaceous subnuclear structure that is somewhat analogous to the cytoskeleton. This structure has been suggested to act as an anchor for several important cellular processes such as DNA replication (Anachkova et al, 2005), DNA transcription (Robinson et al, 1983) and DNA repair (Koehler and Hanawalt, 1996) as well as pre-mRNA splicing (Zeitlin et al, 1987).

An example of the interaction between a nuclear matrix protein and the process of XCI comes from the previously mentioned heterogeneous nuclear ribonucleoprotein U (hnRNPU), also known as SAF-A, which has been identified as a 120kDa nuclear matrix component with high affinity for scaffold attachment regions (SARs) (Romig et al, 1992). Studies have also shown enrichment of this protein in the Xi area, which showed to be dependent on the RGG box, the RNA binding domain of SAF-A (Helbig and Fackelmayer, 2003). Consistent with a role in XCI, knockdown of SAF-A resulted in lack of localization of the *Xist* RNA to the inactive X chromosome as well as absence of the H3K27m3, suggesting that it is involved in tethering *Xist* to the inactive X chromosome (Hasegawa et al, 2010).

In addition to SAF-A, our recent evidence, outlined in section 1.4 of this chapter, suggests that the nuclear matrix protein Cip1 interacting zinc finger protein 1 (CIZ1) may have a role in XCI (Ridings-Figueroa, unpublished). Investigation of the functional significance of this observation is the subject of my thesis.

1.3 Cip1 interacting zinc finger protein (CIZ1)

Precise and faithful replication of DNA during S-phase (Figure 1.3) is crucial in order to prevent genetic abnormalities that could give rise to disease, death and other abnormalities. To ensure this, the cell cycle is subject to a high degree of control and safety systems, known as checkpoints, are in place. Key control elements are cyclin-dependent kinases (CDKs), made up of a kinase subunit and a regulator cyclin subunit which activates and targets the kinase, to support phosphorylation of appropriate substrates. During G1 phase cyclin Ds associate with CDK4 and CDK6

and help transfer information that promotes or inhibits proliferation, leading to expression of cyclin E, passage through the restriction point (R) and assembly of pre-replication complexes. To allow S phase to progress cyclin A substitutes cyclin E as the partner of CDK2. For transition into the mitosis phase cyclin B associates with CDK1 (Alberts et al, 2008).

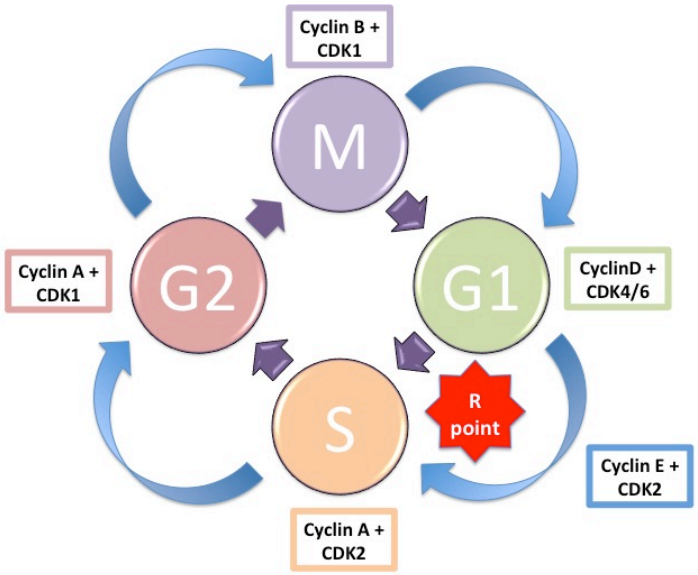


Figure 1.3. Cell cycle control. Progression through the cell cycle is governed by the interaction of cyclin-dependent kinases (CDKs) with their cyclin partners.

CIZ1 has been identified as a nuclear matrix protein and plays an important role in the initiation of mammalian DNA replication (Coverley et al., 2005). It was originally identified by its ability to bind to the cyclin-dependent kinase inhibitor p21 (Mitsui et al., 1999), which inhibits DNA replication causing cell cycle arrest. It has subsequently been reported to interact with CDK2, cyclin E, cyclin A and the pre-replication complex factor CDC6 (Copeland et al, 2010). CIZ1 has been reported to function as a molecular scaffold during DNA replication, allowing coupling of the activities of cyclin A and cyclin E. During mid-G1 phase cyclin E interacts with CDK2, which allows pre-replication complex (pre-RC) assembly. As the cell cycle progresses into late G1 the levels of cyclin A rise and cyclin A displaces cyclin E as the molecular partner of CDK2, which leads to activation of the DNA replication machinery and progression through S phase (Copeland et al, 2010). In vitro this cyclin exchange occurs in association with the N-terminal end (Figure 1.4) of the CIZ1 protein (Coverley et al, 2005). The C-terminal part of the protein, on the other hand, tethers the CIZ1 protein to the nuclear matrix (Figure 1.4), in a cell cycle dependent manner,

allowing the formation of CIZ1 foci that colocalise with replication factories (Ainscough et al, 2007). CIZ1 has also been linked with post-replicative functions in male germ cell differentiation (Greaves et al, 2012). Together these data suggest that CIZ1 may specify the locations in the nucleus where cyclin exchange takes place, or alternatively that CIZ1 may be involved in delivering first cyclin E then cyclin A to sites of DNA replication.

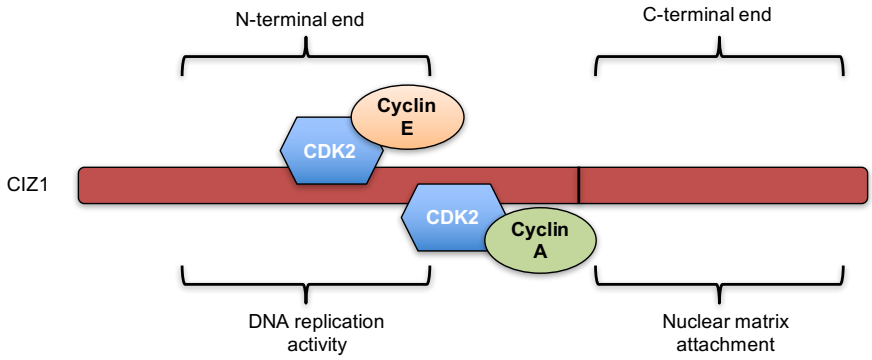


Figure 1.4. Schematic of the functional domains of CIZ1 protein. The N-terminal end of the protein is responsible for the DNA replication activity of the protein, acting as a platform for cyclin exchange. The C-terminal end of the protein is responsible for its attachment to the nuclear matrix.

The CIZ1 protein is encoded by 17 exons, which are alternatively spliced. In fact, at least twenty-two variants of CIZ1 have been identified, including the isoform known as b-variant, which is being developed as a biomarker for early stage lung cancer (Higgins et al, 2012). This variant has a shorter alternative to exon 14, which has been named 14b (Higgins et al, 2012). CIZ1 has also been linked with other types of cancer, which include both paediatric cancers such as Ewing sarcoma (Rahman et al, 2007) and adult cancers such as breast cancer (den Hollander et al, 2006). More recently CIZ1 has also been linked with hepatocellular carcinoma in promoting both tumour growth and metastasis (Wu et al, 2015). It has also been linked with non-cancer chronic disorders such as Alzheimer's disease (Dahmcke et al, 2008) and cervical dystonia (Xiao et al, 2012).

Interestingly, work done by previous PhD students in the Coverley and Ainscough labs, showed that genetic deletion of *Ciz1* (*Ciz1*-KO) leads to development of non-Hodgkin follicular type lymphoma in 100% of tested female mice**,** suggesting a role as a tumour suppressor (unpublished).

1.4 CIZ1 and *Xist*

Initial experiments carried out in the Coverley lab have shown that when using an anti-CIZ1 antibody, such as 1794 (Coverley et al, 2005), high density patches that resemble the Barr body can be observed by immunofluorescence and that these high-density patches colocalise with the H3K27me3 marker (unpublished), characteristic of the inactive X chromosome. It has also been shown using fluorescence in situ hybridization (FISH) that CIZ1 is required for accurate *Xist* localization and H3K27me3 marker, which disappears in *Ciz1*-KO derived cells (Ridings-Figueroa, unpublished). Re-expression of *Ciz1* in a *Ciz1* KO cell line can restore both Xist localization and the H3K27me3 methylation mark. These results point towards a previously unknown role for CIZ1 in the process of X chromosome inactivation. This is supported by the finding that CIZ1 interacts with *Xist* lncRNA, though it is not known whether this is a direct or indirect interaction (Chu et al, 2015).

The aim of this project is to investigate the role of CIZ1 in XCI. I hypothesize that CIZ1, alike nuclear matrix protein SAF-A may have a function in tethering of *Xist* to the inactive X. However, absence of CIZ1 is unlikely to have major effects on gene expression from the Xi since other nuclear matrix proteins like SAF-A have been reported to have similar and likely redundant functions. My project is divided into three approaches, which address related questions:

1. Investigation of which domains of CIZ1 protein are required for recruitment and binding at the inactive X.
2. Confirmation of our evidence in support of loss of and reinstatement of the H3K27me3 mark at the Xi, in response to loss and re-induction of *Ciz1* expression.
3. Analysis, by mean of an RNA sequencing experiment, of the effect that loss of *Ciz1* has on genes from the X chromosome, and specifically whether *Ciz1* knockout leads to reactivation of genes from the inactive X chromosome and whether this could account for the lymphoma-like phenotype observed in *Ciz1* knockout female mice. In addition, data derived from whole-genome analysis will be used to obtain information on other functions or biological pathways that CIZ1 might be involved in.

CHAPTER 2. MATERIALS AND METHODS

2.1 Cell lines

2.1.1 Cell line information

CELL LINE		DERIVATION	♀/♂	GENOTYPE
E14.4		Primary mouse embryonic fibroblasts	♀	Wild-type
E13.1		Primary mouse embryonic fibroblasts	♀	Wild-type
E13.8		Primary mouse embryonic fibroblasts	♀	Wild-type
E14.2		Primary mouse embryonic fibroblasts	♀	*Ciz1* KO
E13.15	- dox	Primary mouse embryonic fibroblasts	♀	*Ciz1* KO
	+ dox			
E13.17	- dox	Primary mouse embryonic fibroblasts	♀	*Ciz1* KO (contains *Ciz1* inducible transgene)
	+ dox			
E13.2		Primary mouse embryonic fibroblasts	♂	Wild-type
E13.6		Primary mouse embryonic fibroblasts	♂	*Ciz1* KO

Table 2.1. Cell lines used for RNA sequencing experiment. Table represents the list of primary mouse embryonic fibroblasts cell lines that were used for RNA extraction and sequencing. Doxycycline was added to the e13.17 cell line which contains a doxycycline inducible GFP tagged Ciz1 transgene, and to the e13.15 cell line which does not contain the transgene but was used to test for direct effects of doxycycline on gene expression. All cell lines were generated by Justin Ainscough.

CELL LINE	DERIVATION	♀/♂	GENOTYPE
3T3 (Coverley et al, 2002)	Fibroblasts	♀	Wild-type

Table 2.2. Use of other cell lines.

2.1.1.1 Design and generation of Ciz1 KO cell lines (Justin Ainscough)

CIZ1 null mice were generated from C57BL/6 ES clone IST13830B6 (TIGM) harboring a neomycin resistance gene trap inserted downstream of exon 1. Absence of CIZ1 in homozygous progeny was confirmed by qPCR, immunofluorescence and immunoblot with CIZ1 N-term antibody 1793.

2.1.1.2 Design and generation of e13.17 Ciz1 KO cell line (Justin Ainscough)

Primary embryonic fibroblasts (e13.17 *Ciz1* KO PEFs) were derived from double transgenic mice harbouring responder and activator transgenes on a CIZ1 -/- background (Figure 2.1) (Ainscough, unpublished). The responder transgene contains regulatable GFP tagged full-length mouse CIZ1 and a LacZ reporter under control of a bidirectional minimal CMV promoter. The activator transgene encodes reverse-tetracycline regulatable trans-activator (rtTA) under control of the ubiquitous ROSA26 promoter. Addition of doxycycline to media induces rtTA-mediated activation of GFP-CIZ1 via the responder tetracycline response element (TRE) within 6 hours (Ainscough, unpublished). The construct and mouse from which cells were derived were made by Justin Ainscough.

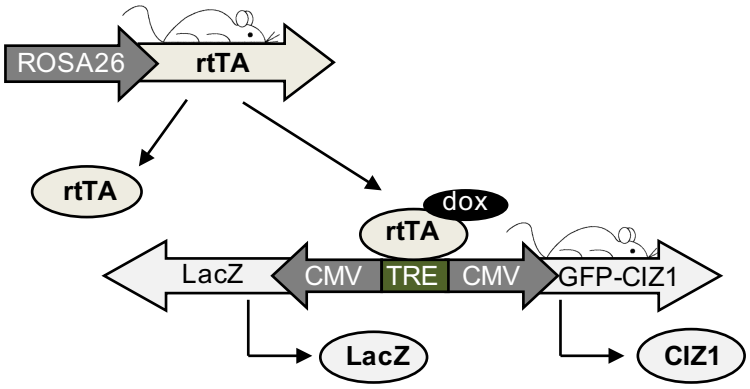


Figure 2.1. Schematic of CIZ1 induction mouse model (KO-I) showing regulated expression of full length CIZ1.

2.1.2 Media

Mouse primary embryonic fibroblasts (mPEFs) as well as 3T3 cells (Coverley et al, 2002) were grown in Dulbecco's Modified Eagle Medium (DMEM) with 10% foetal bovine serum (Cat. No. A15-151) and 1% penicillin/streptomycin.

2.1.3 Maintenance

Cells were grown at 37°C in a humidified atmosphere in 5% CO_2 air. Cells were usually kept at approximately 80% and passaged when necessary. Passaging was done by removing media, washing with warm Dulbeco's phosphate buffered saline

(D-PBS) and then lifting them with 1mL 1x 0.5 % Trypsin-EDTA (10X) (Life Technologies, Cat. 15400054).

2.2 Tissue samples

Tissue samples were obtained from the spleen of a WT (No.181) and a *Ciz1* KO (No.97) female mouse, both at 15 months of age (Ainscough, unpublished) (Table 2.3). *Ciz1* KO mice develop normally but 100% of adult females show enlarged spleen, liver and lungs with observable gross lymphoid abnormalities (Albadrani and Ainscough, unpublished). These abnormalities were described as consistent with B-cell lymphoma with T cell infiltration (Roulson, unpublished).

TISSUE SAMPLE	DERIVATION	♀/♂	Body weight	Spleen weight
Spleen tissue WT	15 month WT mouse	♀	31 grams	110 mg
Spleen tissue KO	15 month Ciz1 KO mouse	♀	31 grams	1.21 grams

Table 2.3. Information on tissue samples used for extraction of RNA for RNA sequencing work. RNA samples from these tissues were generated by Justin Ainscough. Although these samples were sent for sequencing and data is available no analysis was carried out due to time constraints.

2.3 RNA extraction

RNA homogenization was achieved using 2mL of TRIzol Reagent (Ambion Life Technologies, Cat. No.15596-026) per 9cm plate of cells. Phase separation step involved addition of chloroform and isolation of aqueous phase followed by RNA precipitation with 100% isopropanol and a final RNA wash step with 75% ethanol. RNA was then resuspended in DEPC water. All RNA samples for the RNA sequencing were then subjected to DNase treatment (Roche, Cat. No. 04716728001).

2.4 Quantitative PCR (qPCR)

Expression levels of *Ciz1* and *Xist* were analysed using real-time, quantitative PCR. Total RNA was extracted from e14.4 WT PEFs and e13.17 *Ciz1* KO PEFs. The latter were exposed to 10 μg/mL of doxycycline for periods ranging from 6 to 72h. All

samples were treated with DNase to remove any DNA contamination. RNA was reverse transcribed using SuperScript® III Reverse Transcriptase (ThermoFisher Scientific). Taqman qPCR was carried out using Taqman (Taqman® Fast Universal PCR Master Mix, AB, TFS Inc) reagents and already existing gene-specific primers for *Ciz1* and *Xist* (Table 2.4), using the AB StepOnePlus™ Real-Time PCR Systems available at the Technology Facility at the University of York. 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 analysed using the StepOne Software v. 2.3 (AB, TFS Inc).

The relative quantification of gene expression was done using the 2^{-ddCt} formula (Schmittgen et al, 2008). The fold changes of gene expression were normalised to an internal control gene, in this case, *GapdH* (not shown). The e14.4 WT was used as a calibrator.

PRIMER	EXON	PRODUCT SIZE (bp)	ASSAY
Ciz1	10-11	77	Mm00503766_m1
Xist	*1-2*	65	Mm01232884_m1
GapdH	2-3	107	Mm99999915_g1

Table 2.4. Primer-probe sets used for quantitative analysis.

2.5 Whole transcriptome RNA sequencing

2.5.1 Analysis of RNA integrity

Following RNA extraction and DNase treatment all samples were sent to the Technology Facility at the University of York. Quality of the RNA samples was analysed by Sally James, using the Agilent 2100 Bioanalyzer. RNA integrity numbers are determined by the electrophoretic trace of the RNA and indicate the quality of the submitted RNA samples with the highest value,10, indicating intact RNA and the lowest number, 1, indicating degraded RNA. For this specific experiment a RIN value of 7 or higher was recommended (James, personal communication) which puts all but one sample, spleen KO, within a good quality range.

SAMPLE	RNA integrity number (RIN)
E14.4 WT	9.60
E13.1 WT	9.80
E13.8 WT	9.90
E14.2 Ciz1 KO	9.60
E13.15 Ciz1 KO	9.90
E13.15 Ciz1 KO (+dox)	10
E13.17 Ciz1 KO	9.20
E13.17 Ciz1 KO (+dox)	9.60
E13.2 WT (male)	9.70
E13.6 Ciz1 KO (male)	8.70
Spleen WT	7.50
Spleen KO	6.10

Table 2.5. Summary results from the Bioanalyzer analysis for RNA quality.

2.5.2 cDNA library preparations

1μg of RNA was provided to Sally James, member of the technology facility at the University of York, who carried out the library preparations following the NEBNext® Ultra™ RNA library Prep Kit for Illumina® manual. The protocol followed is that for use with NEBNext Poly(A) mRNA Magnetic Isolation Module. This protocol uses the feature that mRNA have poly(A) tails, while rRNA, which comprises approximately 90% of RNA within a cell doesn't have this poly(A) tail. The protocol involves the use of specific magnetic beads that use an oligodT tag to bind to mRNA, pulling this out and into a magnet, and washing out the remaining rRNA species. The enriched mRNA is then subjected to random fragmentation by sonication and priming, producing libraries with 250-400bp insert sizes. This is followed by synthesis of cDNA, adaptor ligation and PCR enrichment.

2.5.3 Sequencing

The combined libraries were sent to the Leeds Institute for Molecular Medicine (LIMM) at St. James's University Hospital in Leeds, where they were run on a single lane of the Illumina 3000 system, using paired-end sequencing (Table 2.6), which allows a maximum read length of 150bp from each end.

Sample	Forward / reverse reads
S9 - E14.4 WT ♀	25666486/25666486
S10 - E13.1 WT ♀	38425408/38425408
S11 - E13.8 WT ♀	34507392/34507392
S12 - E14.2 KO ♀	32424390/32424390
S13 - E13.15 KO ♀	30167428/30167428
S14 - E13.15 KO (+dox) ♀	36654457/36654457
S15 - E13.17KO-I (no dox) ♀	32331645/32331645
S16 - E13.17KO-I (+dox) ♀	29412405/29412405
S17 - E13.2 WT ♂	31185113/31185113
S18 - E13.6 KO ♂	33393640/33393640
S19 - SPLEEN WT ♀	29086647/29086647
S20 – SPLEEN KO ♀	34223665/34223665

Table 2.6. Summary of number of reads per library. Table shows the number of reads obtained for each sample from the Illumina sequencing. Paired-end sequencing was used which results in both forward and reverse reads, which correspond with reads from both ends of the fragment.

2.5.4 Bioinformatics analysis

Initial processing of RNA sequencing data by members of the Technology Facility at the University of York involved use of STAR software (Dobin et al, 2013) for alignment of reads to the X chromosome, using the reference genome of the C57BL/6 mouse (Mus_musculus GRCm38 primary assembly, downloaded from Ensembl.org [Access date: 4/05/2016]). Cufflinks software was used for transcriptome assembly of the mapped reads (Trapnell et al, 2010), (Roberts et al, 2011), followed by differential gene expression analysis with Cuffdiff program, which provided quantified levels of expression (Trapnell et al, 2013).

The output from Cuffdiff shows a series of values for which an example can be found in Table 2.7. In addition, for the purpose of referring to each of the individual entities for which Cuffdiff returns an expression value I used the term ‘transcription unit’, rather than ‘gene’, since the latter is not an accurate term for non-annotated transcripts.

Column name	Example	Definition
Gene_ID	XLOC_000004	A unique identifier describing the transcript, gene, primary transcript, or CDS being tested
Gene	Gm14360	The gene_name(s) or gene_id(s) being tested
Locus	X:3461642-3708291	Genomic coordinates for easy browsing to the genes or transcripts being tested.
Sample_1	WT	Label (or number if no labels provided) of the first sample being tested
Sample_2	KO	Label (or number if no labels provided) of the second sample being tested
Test status	OK	Can be one of OK (test successful), NOTEST (not enough alignments for testing), LOWDATA (too complex or shallowly sequenced), HIDATA (too many fragments in locus), or FAIL, when an ill-conditioned covariance matrix or other numerical exception prevents testing.
Value_1	11.151	FPKM of the gene in sample x
Value_2	2.0901	FPKM of the gene in sample y
Log_2(fold change)	-2.41554	The (base 2) log of the fold change y/x
Test stat	-1.59114	The value of the test statistic used to compute significance of the observed change in FPKM
p_value	0.0191	The uncorrected p-value of the test statistic
q_value	0.603433	The FDR-adjusted p-value of the test statistic
Significant	No	Can be either "yes" or "no", depending on whether p is greater then the FDR after Benjamini-Hochberg correction for multiple-testing

Table 2.7. Example of Cuffdiff analysis output file. The definitions of the terms that appear on a typical cuffdiff file have been extracted from the Trapnell Lab website (http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/) [Accessed 17.08.2016]. An example gene obtained from one of my Cuffdiff files has also been included.

2.5.5 Data interpretation

For the X chromosome analysis, initial filtering involved exclusion of transcriptional units with "no test" or "fail" test status, which indicate that no statistical test could be performed and therefore no p-value can be calculated either. This was followed by selection of transcriptional units of interest, that is, those with a p-value equal to or less than 0.05. Further selection included manual filtering of any transcriptional units where no gene expression was recorded for all but one of the samples. This rendered a total of 62 transcriptional units from the X chromosome with a p-value equal or less than 0.05. A total of 27 transcriptional units within the 62, correspond to non-annotated genes and therefore possible novel-transcripts, which have been indicated with their locus ID. In this particular case, no attention was paid to the significance of the q-value since the main aim of this analysis is to look specifically for changes to gene expression from the X and possible candidate genes of X reactivation. If further analysis was to be carried out on specific genes from the 62 significant p-value list validation by qPCR would be necessary.

For the whole genome analysis, the main aim was to look for groups of related genes with significant changes as well as candidate genes for the *Ciz1* KO phenotype. For this reason, the q-value was again not considered for the selection of candidate genes. In agreement with the opinion of bioinformatics experts another layer of analysis was carried out, using the gene set enrichment analysis software (GSEA), described below.

All RNA sequencing output data files are safely stored in the Coverley lab and are available upon request.

2.5.6 Gene set enrichment analysis (GSEA)

Gene enrichment analysis was done using the Gene Set Enrichment Analysis (GSEA) computational method (Subramanian et al, 2005) (Mootha et al, 2003). This software allows to compare a set of predefined genes from your own analysis, in this case the list of genes with a significant p-value ($p<0.05$) from the whole-genome analysis, with another set of genes which have been grouped together due to different criteria such as being involved in the same biological pathway, disease or chromosomal location for example. These annotated gene sets are collected in the Molecular Signatures Database (MSigDB). An example of a GSEA output together with a definition of each of the terms is shown in Table 2.8.

Column name	Example	Definition
Gene set name	BMI1_DN.V1_UP	Name given to a specific gene collection from the MSigDB
Description	Genes up-regulated in DAOY cells (medulloblastoma) upon knockdown of BMI1 [Gene ID=648] gene by RNAi.	Description of the criteria used to define a particular gene set
Genes in gene set	147	Indicates the total number of genes that appear in a specific gene collection as defined by the MSigDB
Genes in overlap	37	Indicates the number of genes from the submitted list that overlap with the annotated gene set from the MSigDB.
p-value	p = 2.67E-30	Statistical significance of the enrichment score for a single gene set
q-value	q = 5.05E-28	The FDR corrected p-value used for multiple hypothesis testing. Indicates likelihood that the reported p-value is a false positive (when q>0.05)

Table 2.8. Example of GSEA analysis result. Each of the terms included in a typical GSEA output file has been defined. An example obtained from one of my GSEA analysis files has also been included.

2.5.7 Heatmaps

Heatmaps were generated using R software (R Core Team, 2016) using the RStudio platform (RStudio Team, 2015) and the library gplots package (Warnes et al, 2016) with the hetmap.2 function. For heatmaps showing FPKM values, RColorBrewer (Neuwirth, 2014) was used to generate the orange-red colour palette.

2.6 Immunofluorescence

Cells were grown on glass coverslips and then briefly exposed to 0.1% Triton X-100 in PBS, fixed with 4% paraformaldehyde (PFA), rinsed twice with PBS, and then blocked using an antibody buffer (10mg/ml nuclease and protease free bovine serum albumin (BSA, Jackson, Cat:001-000-162), 0.02% w/v SDS, 0.1% v/v Triton X-100 in PBS). Primary antibody was added to the coverslip at the indicated concentration (Table 2.9) and diluted in antibody buffer, followed by incubation in a humidified chamber at 37°C for a period of 2 hours. Following wash with antibody buffer to get rid of excess primary antibody, the secondary antibody was added to the coverslip at the indicated concentration (Table 2.9) and incubated for 1 hour at 37°C.

2.6.1 Antibodies

CIZ1 antibodies				
Antibody	**Origin**	**Concentration**	**Secondary antibody**	**Concentration**
1794	Rabbit polyclonal	1:2000	Alexa Fluor 488 green goat α-rabbit /Alexa Fluor 568 red goat α-rabbit	1:1000 1:2000
Nov4 (ABIN261701, Novus Biologicals)	Rabbit polyclonal	1:500	Alexa Fluor 568 red goat α-rabbit	1:1000 1:5000
Other antibodies				
H3K27Me3 (ab6002, Abcam)	Mouse monoclonal	1:2000	Alexa Fluor 568 red goat α-mouse	1:2000

Table 2.9. Summary of antibodies used for immunofluorescence.

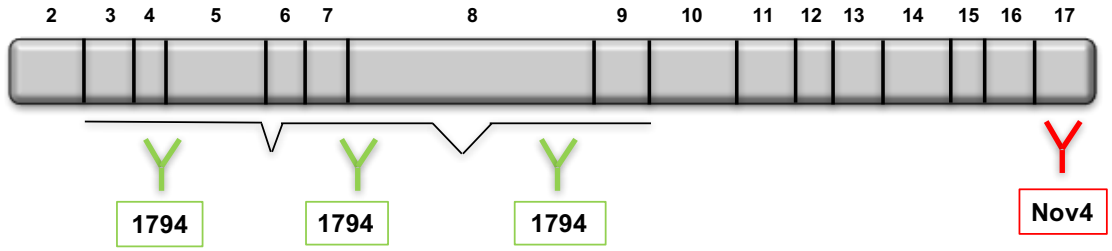


Figure 2.2. Schematic of binding epitopes of CIZ1 antibodies 1794 and Nov4.

2.6.2 Imaging

Imaging was done using a Zeiss Axiovert microscope with a x63/1,40 oil immersion objective. Data was collected using an AxioCam camera and Openlab software. Images were manipulated using Adobe Photoshop CS4 to enhance fluorescence for printing purposes. In cases where levels of fluorescence are relevant for interpretation of the results all images were equally manipulated (as indicated in figures).

2.7 Nuclear matrix extraction

Before immunofluorescence detection with H3K27me3 antibody coverslips were subjected to nuclear matrix extraction to maximize access of the antibody to the methylation mark. The reagents required for this extraction, kindly provided by PhD student Emma Stewart, are indicated below:

- Cytoskeletal buffer (CSK) with 0.1% Triton-X-100 (Table 2.10) in DEPC water.
- CSK DS (CSK with Triton ad indicated above and 0.5M NaCl)
- 1x DNase buffer (40mM Tris/HCl, 10mM NaCl, 6mM MgCl2, 1mM CaCl2, pH 7.9)
- PBS in DEPC water
- 8% paraformaldehyde (PFA)
- 5M NaCl in DEPC water
- 10% TX100 in DEPC water
- DNase (Roche, Cat. No. 04716728001) used at a concentration of 1:50 during a 15-minute incubation time.

Reagent	Quantity (for 50mL)	Final concentration
PIPES/KOH (1M, pH 6.8)	0.5 mL	10mM
NaCl (5M)	1 mL	100mM
Sucrose	5.135 g	300mM
EGTA (250mM)	0.2 mL	1mM
$MgCl_2$ (1M)	50 µl	1mM
DTT (1M)	50 µl	1mM
Protease inhibitor cocktail	1 tablet	-

Table 2.10. List of components of CSK buffer

Cells grown on coverslips were then subjected to nuclear matrix protocol, as indicated in Figure 2.3. Cells on coverslips were immersed in CSK buffer, followed by CSK buffer with detergent and then after immersion in DNase buffer treated with DNase at 37°C for 15 minutes. Cells were then washed in CSK-DS buffer and then fixed in 8% PFA for a period of 20 minutes. After washing in PBS cells were then subjected to standard immunofluorescence protocol.

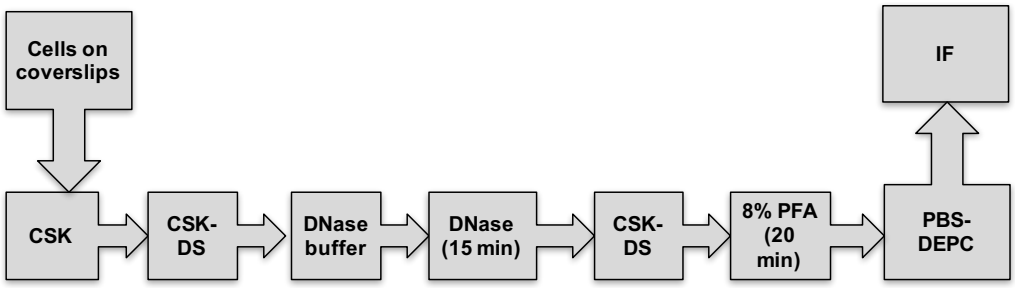


Figure 2.3. Schematic of nuclear matrix extraction protocol.

2.8 Plasmid DNA transfection

Cells on coverslips were grown to 80% confluence prior to transfection on 24-well plates. Cells were then transfected with approximately 0.5µg of the chosen GFP-tagged Ciz1 DNA construct (Table 2.11) together with 50µl of Opti-MEM I Reduced-Serum Medium (Cat. No.31985) and 1.5µl of the transfection reagent TransIT-LT1 Reagent (Cat. No. MIR2300).

Construct	Origin	Amino acids present	Total fragment size (aa)	Provided by
Full-length GFP-CIZ1	Mouse	Met1-845	845aa	Ainscough, J.
GFP-C275 CIZ1	Mouse	570-845	275	Ainscough, J.
GFP-I122 CIZ1	Mouse	708-830	122	Ainscough, J.
GFP-N572 CIZ1	Mouse	Met1-572	572	Ainscough, J.
Full-length GFP eCIZ1	Mouse	CIZ1 Δ exons 2,6,8 (Met84-196^202-342^391-845)	708	Ainscough, J.
GFP-N572 eCIZ1	Mouse	(Met84-196^202-342^391-572)	435	Ainscough, J.
Full-length GFP-CIZ1	Human	Full-length CIZ1 (898aa)	898	Higgins, G.
b-variant GFP-CIZ1	Human	Full-length CIZ1 (Δ VEEELCKQ, exon 14)	890	Higgins, G.
F-variant GFP-CIZ1	Human	Full-length CIZ1 Δ exons 9-11, Δ part of exons 8 and 12	315	Swarts, D.

Table 2.11. Summary table of mouse and human constructs used for transfection. Table has been adapted from Table 1 in Ainscough 2007. Numbers refer to amino acids (aa) at fragment boundaries, relative to full- length Ciz1 (Accession number NP_082688). In all ECiz1-derived fragments residues 1-83 (exon 2), 197-201 (exon 6) and 343-390 (exon 8) are absent and indicated by chevrons. The total number of residues in each fragment is also indicated.

CHAPTER 3. IMPORTANCE OF CIZ1 DOMAINS FOR ITS TARGETING TO THE INACTIVE X

3.1 Introduction

The N-terminal part of the CIZ1 protein is required for its function in DNA replication (Coverley et al, 2005), while its C-terminal domain is important for binding to the nuclear matrix (Ainscough et al, 2007). However, the importance of these two domains in the targeting and binding of CIZ1 to the inactive X chromosome remains uninvestigated. Some of the motifs within the protein include three zinc fingers, an acidic domain and a matrin 3 domain (Figure 3.1). Matrins are a set of highly conserved proteins which are localized to the nuclear matrix. The matrin 3 domain (MH3), contains a matrin 3 type zinc finger (Belgrader et al, 1991) (Figure 3.1). Also indicated below are the fragments that are spliced out in a mouse embryonic *Ciz1* splice variant (eCIZ1), which will be of interest in this chapter, and which lack parts of exons 2, 6 and 8 (Coverley et al, 2005).

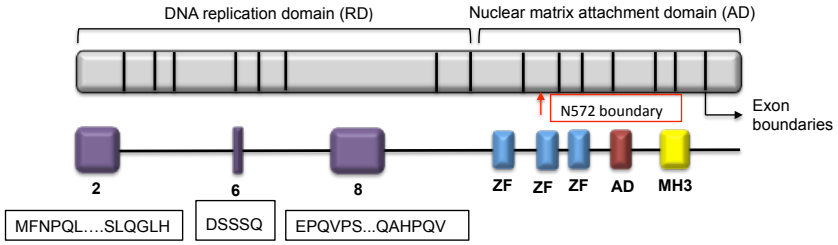


Figure 3.1. Schematic of full-length CIZ1 mouse protein. Indicated in purple are the sequences, which are spliced out in eCIZ1 variant. Indicated in blue are the three zinc fingers within the protein; in red, the acidic domain and in yellow the matrin 3 (MH3) domain within the C-terminus. The red arrow indicates the end of the region that spans the N572 construct.

The use of CIZ1 specific antibody 1794 against endogenous CIZ1 and RNA FISH against the long non-coding RNA Xist have shown that in fact both entities colocalise in the nucleus. Since Xist can be used as a marker for the location of the inactive X chromosome this demonstrates that CIZ1 does in fact form high-density patches at the Xi, which in turn also colocalise with the Barr body (Ridings-Figueroa, unpublished).

3.2 Aims

The aim in this chapter is to investigate which domains of CIZ1 are required for:

i) Transport of CIZ1 to the inactive X region.

ii) Binding and accumulation of CIZ1 at the inactive X region.

3.3 Experimental design

A set of already existing mouse and human GFP-tagged CIZ1 constructs of different lengths were transfected into wild-type female murine cells (3T3) in order to investigate their ability to form high-density patches at the Xi (Ainscough et al, 2007). Scoring was based on two parameters: i) ability of the expressed fragment to form discrete sub-nuclear patches at the Xi, ii) ability to interfere with accumulation of endogenous CIZ1 at the Xi.

3.4 Results

3.4.1 Full-length GFP-CIZ1 forms high-density patches at the Xi with high efficiency

Previous work has shown that endogenous full-length CIZ1 accumulates at the inactive X chromosome forming high-density patches (Ridings-Figueroa, unpublished). The aim of this experiment is to introduce a positive control showing that full-length mouse GFP-CIZ1 (FL-CIZ1), like endogenous CIZ1, can form high-density patches with high efficiency that colocalise with the Barr body and therefore mark the inactive X chromosome (Xi) (Figure 3.2). In fact, in approximately 81% of successfully transfected cells (Figure 3.2A) highly visible patches were scored. Three independent experiments were carried out, where n=100 for each biological replicate.

A) Full-length GFP-CIZ1

B)

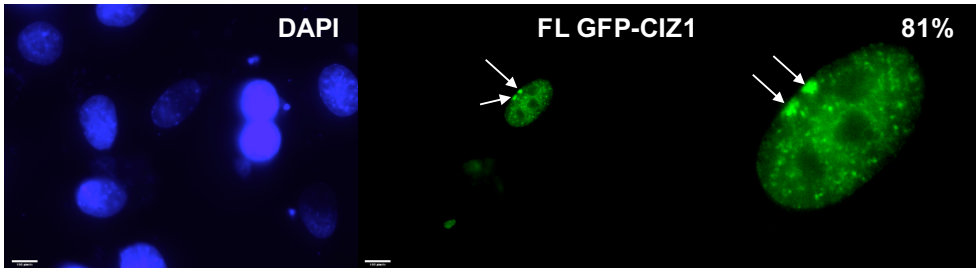


Figure 3.2A. Summary of transfection of full-length mouse GFP-CIZ1. Schematic of the full-length GFP-CIZ1 construct. Green structure represents the GFP tag. **Figure 3.2 B. Full-length CIZ1 forms high-density CIZ1 patches at the inactive X.** Transfection of GFP-tagged full-length CIZ1 into wild-type 3T3 cells leads to formation of CIZ1 high-density patches at the Xi. Image on the far-right shows

zoom in of GFP-CIZ1 positive cell. DNA was stained with DAPI (blue). White arrows indicate location of high-density CIZ1 patches at the inactive X chromosome. Number on the top right hand-side corner indicates the percentage of transfected cells with observed GFP-positive CIZ1 patches at the inactive X chromosome, obtained from the average of three separate experiments (biological replicates) where n=100 for each replicate. Scale bars represent 10μm.

3.4.2 C-terminus analysis shows binding at the Xi

3.4.2.1 The C-terminal end of CIZ1 contains the sequences required for accumulation of CIZ1 at the Xi

When transfecting the C-terminal end fragment (C275) into wild-type 3T3 mouse cells, CIZ1 high-density patches can be observed at the Xi (Figure 3.3B), although at very low frequencies, 14%, compared to full-length CIZ1 (Figure 3.3C). To confirm that the observed CIZ1 patches are in fact at the inactive X chromosome region, endogenous CIZ1 was stained with anti-CIZ1 1794 (see Methods, Figure 2.2). The 1794 polyclonal antibody binds to the N-terminal end of CIZ1 and therefore does not bind to the expressed construct. Since previous evidence has shown that endogenous CIZ1 localises to the inactive X, staining endogenous CIZ1 introduces a control for the real location of the inactive X chromosome. Co-localisation between the GFP-C275 high-density patches and 1794 CIZ1 patches confirmed the ability of the C-term to accumulate at Xi region. The ability of the GFP-C275 construct to interfere with recruitment of endogenous CIZ1 to the Xi was also investigated (Figure 3.3A). The ability of the 1794 antibody to bind endogenous CIZ1 at the Xi was not significantly different between transfected and non-transfected cells, suggesting there is no interference between the C275 construct and endogenous CIZ1 (Figure 3.3D).

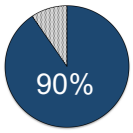
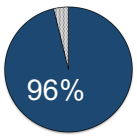
A shorter version of the C275 construct, the 'internal' I122 construct, containing amino acids 708-830 of the C-terminal end of full-length CIZ1 was transfected into wild-type 3T3 cells and endogenous CIZ1 was stained using 1794 anti-CIZ1 antibody (Figure 3.4). This experiment was only performed in one occasion, and only one I122 CIZ1 high-density patch was observed at the Xi. This patch co-localised with endogenous CIZ1 patch at the Xi, suggesting the result is real. However, to confirm the validity of this finding, further repetitions of this experiment will be necessary in the future.

3.4.2.2 Investigating dependence of the C-terminus construct on endogenous CIZ1 to accumulate at the Xi

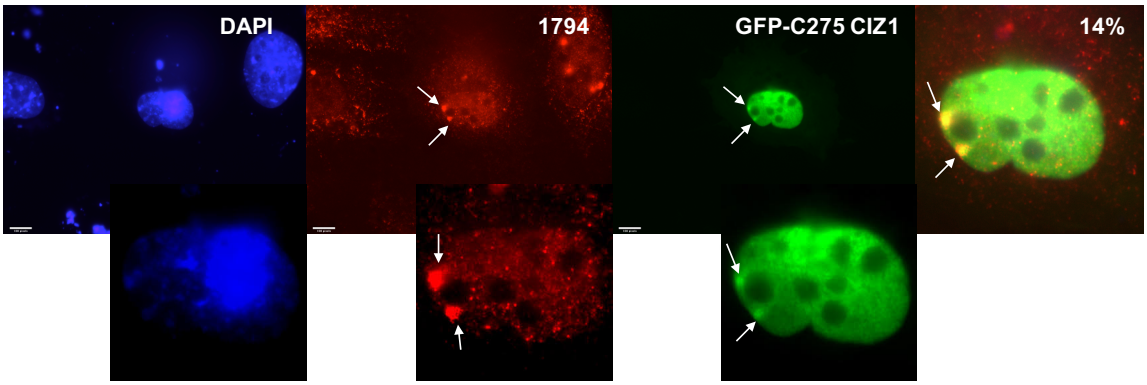
To test whether the ability of the C-terminus construct to bind at the Xi is dependent on the presence of endogenous CIZ1, GFP-C275 construct was transfected into e13.17 *Ciz1* KO cells (Ainscough, unpublished). No C275 high-density patches were observed (Figure 3.5) (n=200, two technical replicates), suggesting that in wild-type cells this construct has been dependent, either directly or indirectly, on endogenous CIZ1 to bind at the Xi. It is possible to argue that the 3T3 fibroblast cell line and the e13.17 *Ciz1* KO PEF cell line are not directly comparable and for this reason, a summer student transfected the GFP-C275 construct into e13.1 wild-type female primary embryonic fibroblasts and observed that the construct is also able to form high-density patches at the Xi in this cell line as well as confirmed its inability to do the same in the e13.17 *Ciz1* KO cell line (Cecere, personal communication).

A) GFP-C275 CIZ1

B)

Competitive assay	
% Transfected cells with 1794 X patches	**% Non-transfected cells with 1794 X patches**
	

C)



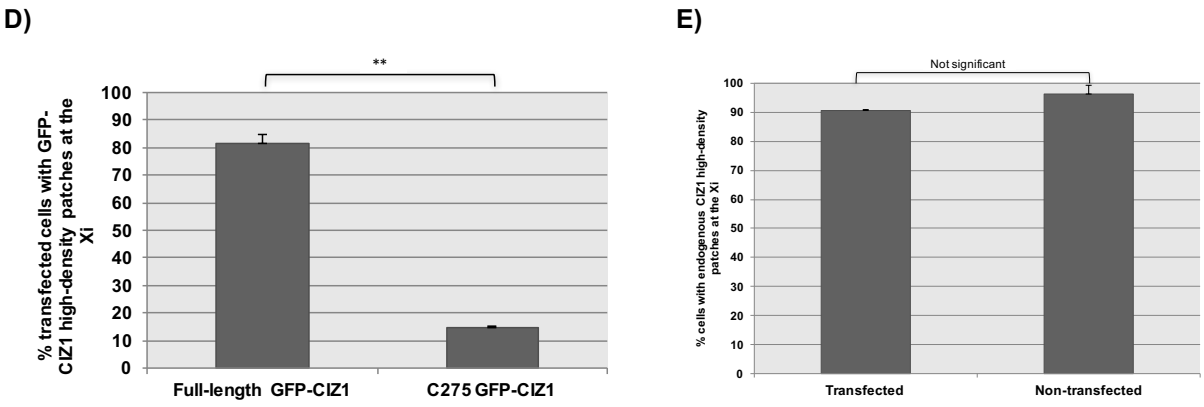


Figure 3.3A. Summary of transfection of mouse C275 GFP-CIZ1. Schematic of the CIZ1 GFP-C275 construct indicates regions present (grey) and absent (white) from this construct. **Figure 3.3B.** Pie charts show the percentage of cells transfected with C275 GFP-CIZ1 construct that show 1794 patches at the Xi compared to the percentage of non-transfected cells with 1794 patches at the Xi. **Figure 3.3C. The C-terminal fragment of CIZ1 contains the sequences required for binding at the Xi.** Transfection of the GFP-tagged C275 (green) terminal fragment of CIZ1 into wild-type 3T3 cells shows that this construct can bind at the Xi region. Endogenous CIZ1 was stained using anti-CIZ1 1794 antibody (red). Colocalisation of the GFP high-density patch with the 1794 high-density patch, shows localization of the GFP construct at the Xi. DNA was stained with DAPI (blue). White arrows indicate location of high-density CIZ1 patches at the inactive X chromosome. Number on the top right hand-side corner indicates the percentage of transfected cells with observed GFP-positive CIZ1 patches at the inactive X chromosome. The percentage is the average obtained from three separate experiments (biological replicates) where n=100 cells for each replicate. Scale bars represent 10μm. **Figure 3.3D. Efficiency of CIZ1 high-density patch formation at the inactive X chromosome.** Transfection of full-length GFP-tagged CIZ1 leads to formation of CIZ1 high-density patches at a significantly higher efficiency compared to transfection of GFP-C275 CIZ1 construct ($F=0.07$, $t=4.69951E-05$, $p<0.001$). *** used to indicate $p<0.001$. Error bars represent standard error of the mean from three independent experiments. **Figure 3.3E. Efficiency of endogenous CIZ1 detection with 1794 antibody in C275 transfected versus non-transfected cells.** Transfection of GFP-C275 CIZ1 construct into wild-type 3T3 does not cause interference with the presence of endogenous CIZ1 at the inactive X. Numbers are obtained from the average of two biological replicates where total n= 113 for transfected cells and n= 110 for non-transfected cells. Error bars represent standard error of the mean of two independent experiments.

A) GFP-I122 CIZ1 (708-830)

B)

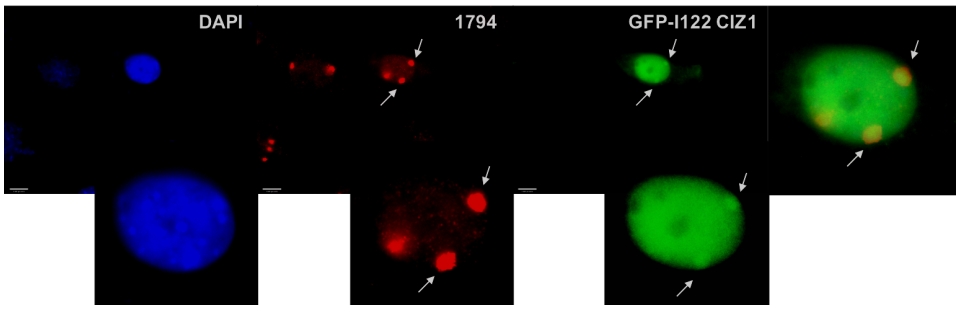


Figure 3.4A. Schematic of the 'internal' 122 GFP-tagged CIZ1 construct. Schematic of the construct, which contains 122 amino acids from the C-terminus end of CIZ1, indicates regions present (grey) and absent (white) from this construct. **Figure 3.4B. The GFP-tagged I22 CIZ1 fragment contains the sequences required for binding at the inactive X region.** Transfection of I122 GFP-CIZ1 construct into wild-type 3T3 cells shows binding at the inactive X region. Endogenous CIZ1 is stained using anti-CIZ1 1794 antibody. White arrows indicate location of high-density CIZ1 patches at the inactive X chromosome. DNA was stained with DAPI (blue). Scale bars represent 10μm. This experiment was only performed once.

A) GFP-C275 CIZ1

B)

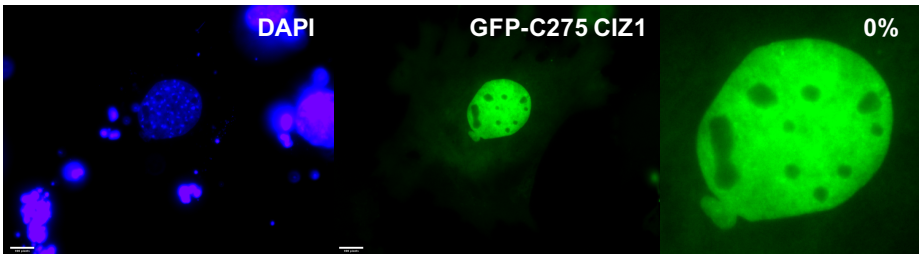


Figure 3.5A. Summary of transfection of mouse C275 GFP-CIZ1 in *Ciz1* KO cells. A schematic of the CIZ1 GFP-C275 construct indicates regions present (grey) and absent (white) from this construct. **Figure 3.5B. The C275 fragment of CIZ1 fails to form CIZ1 high-density patches in the absence of endogenous CIZ1.** Transfection of the GFP-tagged C275 terminal fragment of CIZ1 (green) into e13.17 Ciz1 KO cells shows failure of the construct to form patches at the inactive X chromosome in a Ciz1 KO context. DNA is stained with DAPI (blue). Number on the top right hand-side corner indicates the percentage of transfected cells with observed GFP-positive CIZ1 patches at the inactive X chromosome in *Ciz1* KO cells. The percentage represents the average obtained from two separate experiments (biological replicates) where n=100 cells for each replicate. Scale bars represent 10μm.

3.4.3 N-terminus analysis shows dominant negative effect on endogenous CIZ1

3.4.3.1 N572 GFP-CIZ1 is unable to form high-density patches at the Xi but interferes with the presence of endogenous CIZ1 at the Xi

Following analysis of possible functions of the C-terminus end of CIZ1 in the binding of CIZ1 at the Xi I proceeded to analyse possible functions of the N-terminal end of the protein. When transfecting an N-terminal fragment, the N572 construct, into 3T3 cells, no high-density CIZ1 patches are observed at the Xi region (Figure 3.6A) (Figure 3.6B). Location of the inactive X chromosome was determined by staining endogenous CIZ1 with Nov4 antibody, which binds the C-terminal end of the protein and therefore does not bind to the expressed fragment (see Methods, Figure 2.2). Interestingly it was observed that in non-transfected cells (Figure 3.6C) 72% of cells showed Nov4 X patches, compared to transfected cells where only a 31% had observable endogenous CIZ1 patches (Figure 3.6A). This significant drop suggests that transfection of the N572 construct interferes with the mechanism that allows endogenous CIZ1 to accumulate at the Xi patch. The reason why interference does not occur in 100% of cells remains unclear.

3.4.3.2 The GFP-eCIZ1 variant of N572 construct lacks the sequences involved in the interference mechanism

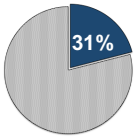
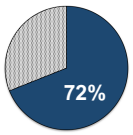
Alternatively spliced embryonic CIZ1 (eCIZ1) (Coverley et al, 2005) is missing sequences from exon 2, 6 and 8, which are present in full-length CIZ1. Since important functions have been linked to some of these sequences, eCIZ1 was picked as a good candidate for transfection into wild-type 3T3 cells. This will allow me to determine whether any of these sequences, also contained within the N-terminal end of the protein, could have important functions in the interference previously observed. I initially determined whether full-length eCIZ1 can form high-density patches at the Xi. Results showed that the construct can form CIZ1 high-density patches but at a low frequency of 15% (Figure 3.7A) (Figure 3.7B), significantly different to full-length CIZ1 (Figure 3.7E), but comparable to the percentage obtained when transfecting the

C275 fragment, which is also contained within eCIZ1 variant. This suggests that eCIZ1 does have the sequences required for binding at the inactive X region but could be missing sequences required for efficient targeting.

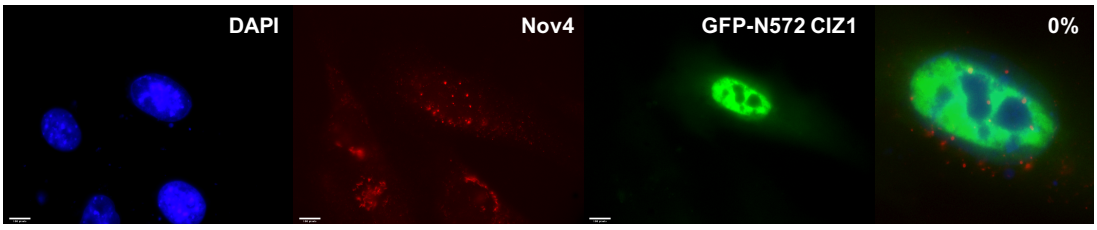
A shorter version of eCIZ1, eN572, (Figure 3.7C) missing the C-terminal end of eCIZ1 (C275) was also transfected into 3T3 cells and endogenous CIZ1 was stained using Nov4 antibody (Figure 3.7D). The aim of this experiment is to test whether this construct also shows the interference reported with the full-length version of N572. Results show that in fact, there is no significant difference between the number of cells with Nov4 X patches in transfected and non-transfected cells (Figure 3.7F), and therefore there is no interference with endogenous CIZ1. This implicates the sequences that are spliced out of eCIZ1 in the interference.

A) GFP-N572 CIZ1

B)

Competitive assay	
% Transfected cells with Nov4 X patches	**% Non-transfected cells with Nov4 X patches**
	

C)



D)

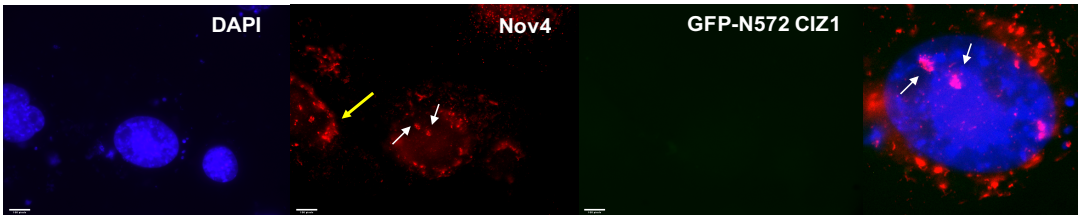
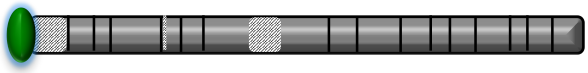
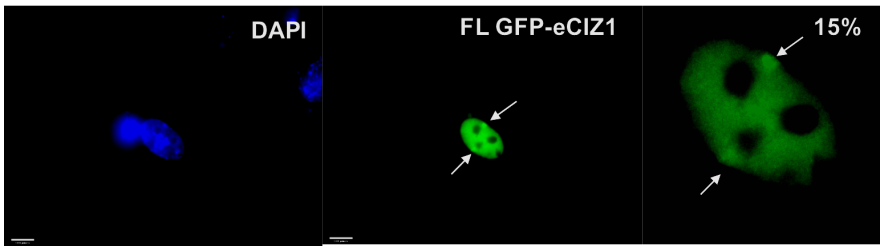


Figure 3.6A. Schematic of N572 mouse GFP-CIZ1. A schematic of the CIZ1 GFP-N572 construct indicates the regions which are present (grey) and absent (white) from the construct. **Figure 3.6B. Competition assay.** Pie charts reflect the ability of the construct to interfere with recruitment of endogenous CIZ1 at the Xi in transfected and non-transfected cells. Percentages are obtained from the average of three biological replicates, where total n=226 for transfected cells and n= 307 for non-transfected cells. **Figure 3.6C. The N-terminal fragment of CIZ1 lacks sequences required for formation of CIZ1 high-density patches at the Xi and interferes with the presence of endogenous CIZ11 at the site.** Transfection of a GFP-tagged N572 fragment (green) into wild-type 3T3 cells shows inability of this construct to form high-density patches at the inactive X region. Number on the top right hand-side corner indicates the percentage of transfected cells with observed GFP-positive CIZ1 patches at the inactive X chromosome in wild-type 3T3 cells. The percentage represents the average obtained from three separate experiments (biological replicates) where n_1=100 cells, n_2= 65 cells, n_3= 61 cells. Endogenous CIZ1 (red) fails to accumulate at the Xi in N572 transfected cells. Scale bars represent 10μm. **Figure 3.6D. Endogenous CIZ1 high-density patches form at the Xi in non-transfected cells.** Endogenous CIZ1 (red) accumulates at the inactive X in non-transfected cells. White arrows indicate location of high-density endogenous CIZ1 patches at the inactive X chromosome. Endogenous CIZ1 is detected with Nov4 antibody which also binds antigens external to the cell, leading to formation of a characteristic 'antibody cap' (indicated with yellow arrow). Scale bars represent 10μm.

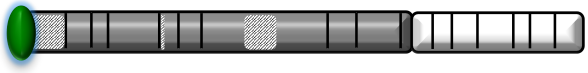
A) Full-length GFP eCIZ1 (Δ exons 2,6,8)



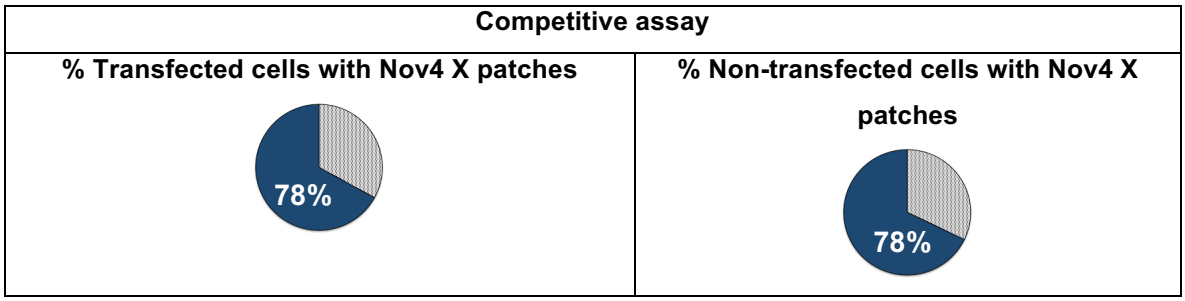
B)



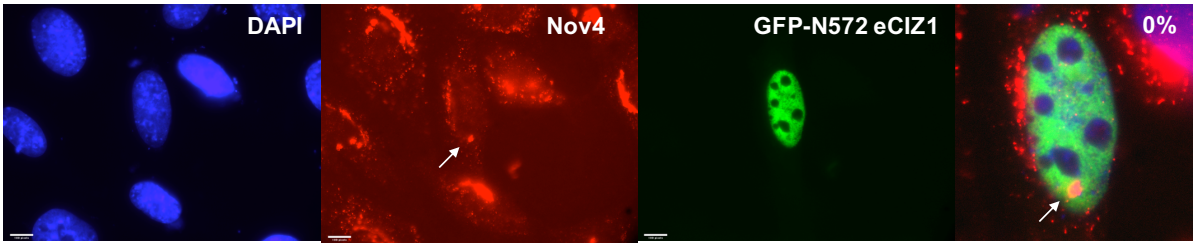
C) GFP-N572 eCIZ1 (Δ exons 2,6,8)



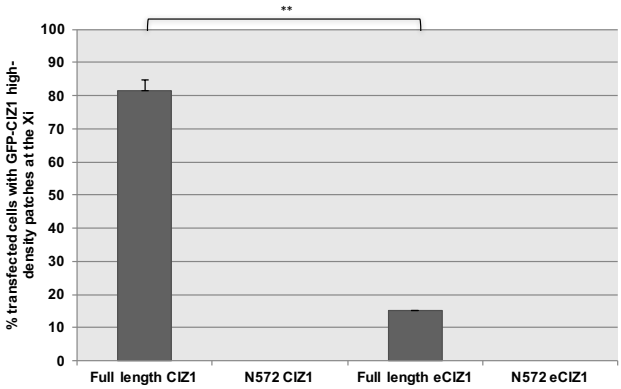
D)



E)



F)



G)

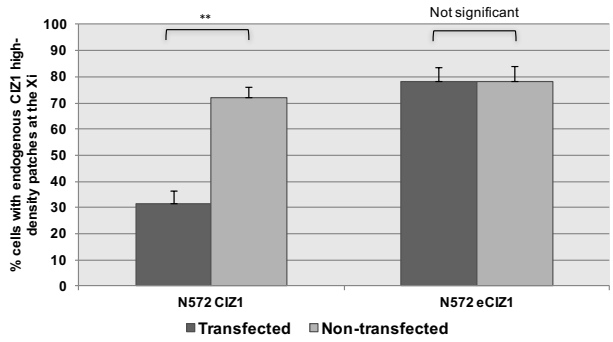


Figure 3.7A. Schematic of full-length eCIZ1 variant. A schematic of the full-length eCIZ1 variant (which is missing fragments from exon 2,6 and 8, crossed regions) construct indicates regions present (grey) and absent (white) from this construct. **Figure 3.7B. Full-length eCIZ1 variant forms high-density CIZ1 patches at the Xi with low frequency.** Transfection of full-length eCIZ1 (green) variant

into 3T3 wild-type cells shows ability of the construct to form patches at the inactive X region. DNA was stained with DAPI (blue). White arrows indicate location of high-density CIZ1 patches at the inactive X chromosome. Number on the top right-hand side corner indicates the percentage of transfected cells with observed GFP-eCIZ1 patches at the inactive X chromosome in wild-type 3T3 cells. The percentage represents the average obtained from three separate experiments (biological replicates) where n=100 for each replicate. Scale bars represent 10μm. **Figure 3.7C. Schematic of N572 mouse GFP-eCIZ1 variant construct.** A schematic of the eCIZ1 variant (which is missing fragments from exon 2,6 and 8) GFP-N572 construct indicates regions present (grey) and absent (white) from this construct. **Figure 3.7D**. Competitive assay. Table shows the results of the competition assay, where pie charts represent the percentage of cells that show endogenous CIZ1 patches at the Xi, both in transfected and non-transfected cells. The percentage represents the average obtained from three separate experiments (biological replicates) where n=100 for each replicate. **Figure 3.7E. The N-terminal fragment of eCIZ1 variant fails to form high-density patches at the Xi.** Transfection of the N572 fragment of eCIZ1 (green) into wild-type 3T3 cells shows inability of the construct to form CIZ1 patches at the inactive X region. Number on the top right-hand side corner shows the percentage of transfected cells with observed GFP-eCIZ1 patches at the inactive X chromosome in wild-type 3T3 cells. The percentage represents the average obtained from three separate experiments (biological replicates) where n=100 for each replicate. Endogenous CIZ1 (red) is stained using anti-CIZ1 Nov4 antibody. DNA was stained with DAPI (blue). White arrow indicates location of high-density endogenous Nov4 CIZ1 patches at the inactive X chromosome. Scale bars represent 10μm. **Figure 3.7F. Efficiency of CIZ1 and eCIZ1 variant constructs to form CIZ1 patches at the Xi.** Transfection of full-length CIZ1 as well as eCIZ1 variant leads to formation of CIZ1 patches at the inactive X chromosome. CIZ1 forms patches at a significantly higher efficiency compared to eCIZ1 variant ($t=0.002$, $p<0.01$). Transfection of the N-terminal end of CIZ1 as well as eCIZ1 variant shows inability of both constructs to form high-density patches at the Xi. Error bars represent standard error of the mean from three independent experiments. ** used to indicate $p<0.01$. **Figure 3.7G. Differential ability of CIZ1 and eCIZ1 N-terminus constructs to interfere with endogenous CIZ1.** Wild-type 3T3 cells, transfected with GFP-N572 CIZ1 construct, show a significantly lower percentage of cells with 1794 endogenous CIZ1 at the inactive X compared to non-transfected cells ($F=0.92$, $t=0.0035$, $p<0.01$). The percentage of cells with 1794 endogenous CIZ1 high-density patches at the Xi is not significantly different between cells transfected with GFP-N572 eCIZ1 construct and non-transfected cells. Error bars represent standard error of the mean. ** used to indicate $p<0.01$.

3.4.4 Transfection of human constructs

3.4.4.1 CIZ1 b-variant does not display a significantly different ability to form high-density patches at the Xi compared to full-length human CIZ1

A human CIZ1 splice-variant, known as b-variant, has an in-frame deletion of eight amino acids (VEEELCKQ) from exon 14 (Higgins et al, 2012). This variant is of special interest since it has been defined as a circulating biomarker for lung cancer

(Higgins et al, 2012). When transfecting a GFP-tagged construct of this variant into wild-type female 3T3 mouse cells, approximately 91.5% of transfected cells (GFP positive) (Figure 3.8A) showed high-density GFP patches at the Xi (Figure 3.8B). This is not significantly different from the 91.3% of transfected cells that show CIZ1 high-density patches at the Xi when transfecting full-length human CIZ1 (Figure 3.8A) (Figure 3.8B). This suggests that the sequences missing from this construct do not have any important functions in recruitment or binding at the Xi.

3.4.4.2 F-variant CIZ1 is unable to form high-density CIZ1 patches at the inactive X

F-variant is a CIZ1 splice variant in which all of exons 9 to 11 and part of exons 8 and 12 are removed (Figure 3.9A) (Swarts et al, in preparation) (Rahman et al, 2007). This splice event leads to a frame shift, which generates a short alternative reading frame and a premature stop codon. Transfection of a GFP-tagged F-variant construct into wild-type 3T3 cells shows inability of the construct to form high-density patches at the inactive X chromosome (Figure 3.9B). While conserving exons 2 and 6, F-variant is missing part of exon 8. This makes it an ideal candidate to test for competition with endogenous CIZ1. In an initial experiment no interference was observed. Although the number of endogenous CIZ1 patches in transfected cells is lower than in non-transfected cells the difference is much smaller than in other transfections where clear interference has been observed. This suggests that exon 8 might be responsible for interference, however, repetitions of this experiment would be necessary to confirm the result.

A) Full-length GFP-Human CIZ1

b-variant human CIZ1

B)

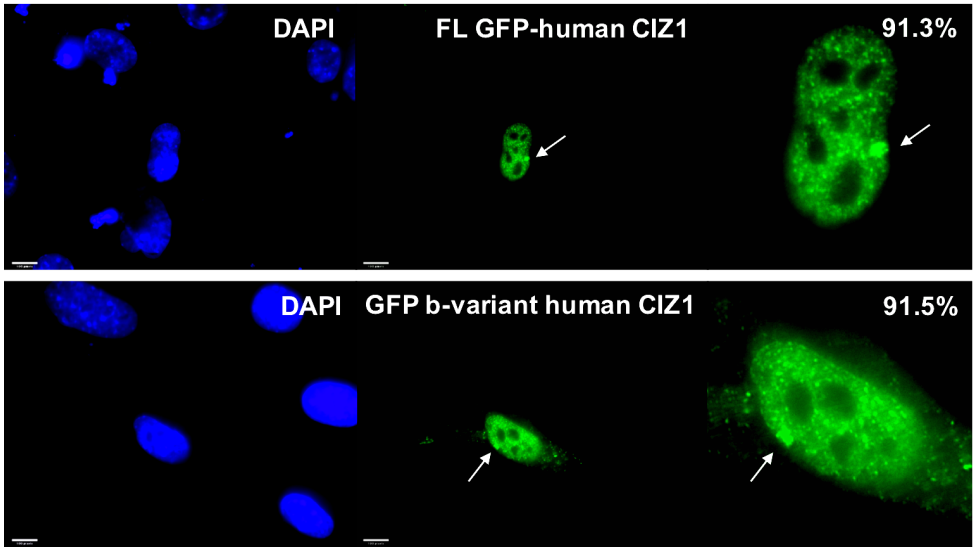
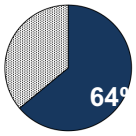
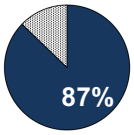


Figure 3.8A. Summary of transfection of full-length human GFP-CIZ1 and human b-variant. A schematic of the full-length human CIZ1 and b-variant human CIZ1 indicates regions present (grey) and absent (hatched region) from this construct. **Figure 3.8B. Full-length GFP-tagged human CIZ1 and human b-variant form high-density patches at the Xi.** Upper images show transfection of full-length GFP-tagged human CIZ1 (green) into wild-type mouse 3T3 cells, which can form patches at the Xi. Number on the top right-hand side corner indicates the percentage of transfected cells with observed GFP-positive human CIZ1 patches at the inactive X chromosome in wild-type 3T3 mouse cells. The percentage represents the average obtained from three separate experiments (biological replicates) where n=100 cells for each replicate. Bottom images show transfection of human b-variant CIZ1 (green) into wild-type 3T3 mouse cells and formation of GFP high-density patches at the Xi. DNA was stained with DAPI (blue). White arrows indicate location of high-density CIZ1 patches at the inactive X chromosome. Number on the top right-hand side corner indicates the percentage of transfected cells with observed GFP-positive human CIZ1 patches at the inactive X chromosome in wild-type 3T3 mouse cells. The percentage represents the average obtained from two separate experiments (biological replicates) where n=100 cells for each replicate. Scale bars represent 10μm.

A) F-variant human CIZ1

B)

Competitive assay	
% Transfected cells with Nov4 X patches	**% Non-transfected cells with Nov4 X patches**
	

C)

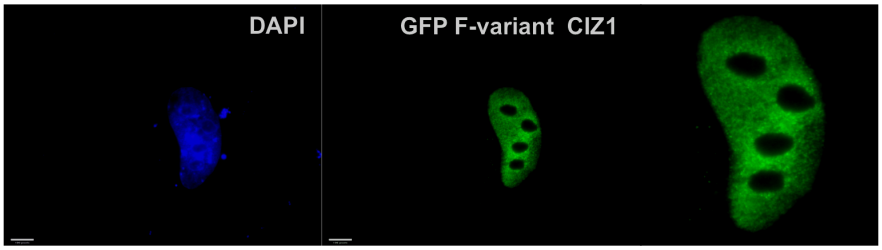


Table 3.9A. Summary of transfections of human GFP-F-variant into wild-type 3T3 mouse cells. A schematic of the GFP-F-variant CIZ1 construct indicates regions which are present (grey) and absent (white) in this construct. Red region indicates the region of the protein generated because of the premature stop codon. **Figure 3.9B**. Competition assay. Table shows the results of the competition assay using pie charts which reflect the percentage of transfected and non-transfected cells with endogenous CIZ1 patches at the Xi. **Figure 3.9C. GFP-tagged human CIZ1 F-variant fails to form high-density patches at the Xi.** Transfection of GFP-tagged CIZ1 F-variant (green) into wild-type 3T3 cells shows inability of the construct to form high-density patches at the Xi region. DNA was stained using DAPI (blue). Number on the top right-hand side corner indicates the percentage of transfected cells with observed GFP positive CIZ1 patches at the inactive X chromosome. Scale bars represent 10μm.

3.5 Discussion

The ability of the C275 construct to form high-density patches at the Xi suggests that it contains the sequences required for stable binding of CIZ1 to the Xi (Figure 3.10). However, the low frequency could suggest that it lacks the sequences for efficient targeting to the site, only accumulating passively or by a less efficient delivery process than full-length CIZ1. It is also possible that the construct can efficiently bind but not stabilize at the site. To test this hypothesis, it would be possible to transfect the construct for varying periods of time and ask whether the percentage of cells with Ciz1 high-density patches decreases with time. The number of full-length CIZ1 and C275 transfected cells does not appear to be different, just the number of transfected cells with high-density CIZ1 patches at the Xi, suggesting there is no difference in the speed of expression of these constructs. The C275 construct contains two important functional domains, which could be responsible for the binding function: the acidic domain, within exon 14, and the MH3 domain, which spans part of exon 15 and 16. The acidic domain has been suggested to have a function in protein interactions (Mitsui et al, 1999). Although of unknown function, the MH3 domain (Belgrader et al, 1991) has been suggested to allow binding of CIZ1 to DNA (Mitsui et al, 1999)

(Warder and Keherley, 2003) or the nuclear matrix (Ainscough et al, 2007). In addition, the zinc finger spanning amino acids 626 to 660, referred to as JAZ ds-RNA-binding zinc finger protein, has been predicted to contain sequences, which may function in binding of double-stranded RNA (InterPro, [Accessed 11th June 2016).

The 'internal' I122 construct, which spans the MH3 domain but lacks the acidic domain, is also able to form high-density patches at the Xi. Although at very low frequency, they are verified as Xi by colocalisation with endogenous CIZ1. This suggests that the MH3 domain is sufficient for stable recruitment of CIZ1 to the Xi. The significantly decreased efficiency suggests however, that other sequences contained within the C275 construct, and missing from the shorter I122 construct, contribute to stable binding. Due to time constraints, I was unable to obtain replicates of this experiment so to verify the result it would be necessary to repeat the experiment.

Failure of the C275 construct to form high-density patches at the Xi in a *Ciz1* KO context shows that binding is dependent, either directly or indirectly, on endogenous CIZ1. One possible hypothesis is that the construct is binding to CIZ1 itself. However, it is also possible that the construct is binding to an entity or structure that is either absent or disrupted in the absence of CIZ1. A likely candidate would be the long non-coding RNA *Xist*, which becomes delocalised from the inactive X chromosome in the absence of CIZ1 (Ainscough et al, in preparation). This would likely cause ectopically expressed CIZ1 to also become delocalised.

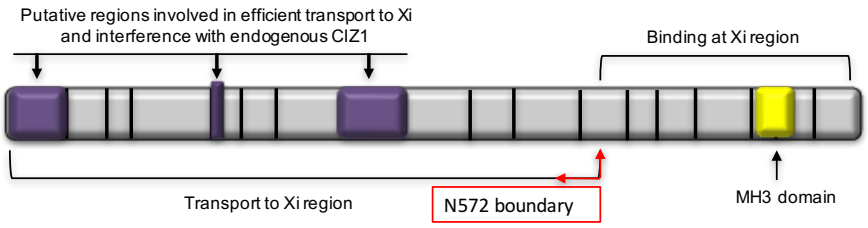


Figure 3.10. Schematic of CIZ1 regions implicated in transport and binding to the Xi. The N-terminus region, and more specifically the regions contained within the N-terminus that are absent from eCIZ1 variant (indicated in purple), appear to be responsible for efficiently targeting to the Xi region, thus interfering with the presence of endogenous CIZ1 at the site. The C-terminus region, although unable to allow efficient targeting to the Xi, appears to be implicated in binding at the Xi.

Regarding the N-terminus function of CIZ1, transfection of full-length embryonic CIZ1 (eCIZ1) shows a significantly lower ability to form high-density patches at the Xi

compared to full-length CIZ1. In fact, the percentage of transfected cells with CIZ1 high-density patches is very similar for both eCIZ1 variant and the C275 CIZ1 construct. These observations suggest that it is eCIZ1 key functional domains, which contribute to efficient targeting (Figure 3.10). Since the C-terminal fragment is identical between CIZ1 and eCIZ1 variant this also suggests that those sequences are located at the N-terminus. We are the developing and testing the hypothesis that function encoded by the exons that are spliced out of eCIZ1 (2,4,8) support binding to an active transport mechanism that delivers CIZ1 to the Xi. This hypothesis is supported by the ability of the CIZ1 N572 fragment to interfere with the targeting of endogenous CIZ1. This dominant negative effect suggests that N572 could have the sequences required for transport to the Xi region, therefore displacing endogenous CIZ1. Due to the lack of C-terminal sequences it would be unable to bind and form patches itself. This also provides further support that sequences contained within the C-terminus are the ones responsible for successful binding of CIZ1 at the Xi. Notably, transfection of the C-terminus construct does not cause any interference with the presence of endogenous CIZ1, controlling for the possibility that transfection itself is responsible for the dominant negative effect on endogenous CIZ1.

Importantly, the N572 fragment of eCIZ1 (lacking parts of exon 2,6,8) does not cause interference with localization of endogenous CIZ1 at the Xi, further implicating these sequences in the dominant negative effect. The next step to determine what sequences of CIZ1 are required for successful recruitment to the Xi would be to investigate the functions of exon 2, the DSSQ fragment within exon 6 and the fragment of exon 8 missing from eCIZ1. All three sequences could be of potential importance. Exon 2 contains a glutamine-rich region. Although the function of these glutamine rich regions is still uncertain it is known that abnormal expansion of this residues has been linked with the generation of protein aggregations related to neurodegenerative conditions (Michelitsch et al, 2000). The DSSSQ fragment is identical between human and mouse protein and it has been reported that naturally expressed CIZ1 proteins can either contain or not this sequence, suggesting its presence or absence might be an important for the regulation of protein activity or function. Important functions have also been attributed to exon 8. A CIZ1 splice variant, which lacks a glutamine-rich region within exon 8, is overexpressed in patients with Alzheimer's disease (Dahmcke et al, 2008).

Human CIZ1 b-variant is lacking eight amino acids from exon 14, which cause a one amino acid deletion from the acidic domain, and has been implicated in lung cancer. The ability of full-length human CIZ1 and b-variant to form high-density patches at the inactive X with a very similar efficiency suggests this domain might not have an important role in this process. In addition, the inability of F-variant CIZ1 to form CIZ1 patches at the Xi, supports the hypothesis that the C-terminus, which is missing from F-variant is required for successful binding at the Xi region. In addition, the initial observation that F-variant does not interfere with recruitment of endogenous CIZ1 to the Xi suggests that exon 8, which is missing from this construct, could be responsible for this function. However, further repetition of the experiments as well as further investigation of exon 2 and 6 would be necessary to confirm this result.

The results shown in this chapter begin to provide some information on a possible model that describes localisation and binding at the inactive X region. The N-terminal end, and most probably one or more of the sequences spliced out in eCIZ1 variant, could contain the sequences required for delivery to the X chromosome localisation. The C-terminus end of the protein on the other hand, contains sequences that are sufficient for binding of the protein at the inactive X, which is likely to occur through the MH3 domain. Possible new experiments would involve the generation of a series of constructs missing the fragments from exon 2, 6 and 8. These would then be tested either for their ability to form high-density patches at the same frequency as full-length CIZ1, or if missing the C-terminal domain, for their ability to interfere with the presence of endogenous CIZ1 at the inactive X region.

CHAPTER 4. ANALYSIS OF THE RELATIONSHIP BETWEEN CIZ1 AND H3K27ME3

4.1 Introduction

Methylation of histone 3 at lysine 27 (H3K27) is one of the events that occurs during X chromosome inactivation. This repressive mark, laid down by the PRC2 complex, contributes to maintenance of epigenetic silencing. An RNA FISH experiment targeting lnc-RNA *Xist*, one of the major players in X inactivation, showed that *Xist* becomes delocalized from the inactive X in a *Ciz1* KO background unlike in a wild-type scenario where Xist forms dense patches at the Xi. It is therefore of interest to assess whether other players such as the H3K27me3 mark are also affected by absence of CIZ1.

4.2 Aims

In this chapter I aim to investigate three main questions:

- Test whether there is disruption of the H3K27me3 methylation mark from the inactive X chromosome in a *Ciz1* KO background.
- Using an inducible gene-expression system, test whether re-expression of *Ciz1* in a KO context, is capable of reinstating the methylation mark and determine the kinetics.
- In case of reinstatement of the methylation mark, test whether passage through the cell cycle is required.

4.3 Experimental design

These questions will be answered by using a *Ciz1* KO and WT cell lines derived from PEFs, and the e13.17 *Ciz1* KO cell line, which has a transgene capable of re-expressing *Ciz1* when doxycycline is added to the cell culture media (see Methods, Figure 2.1). By exposing the e13.17 KO cell line to doxycycline for different periods of time ranging from 6 to 48 hours, I will test the effects of exposure time on the reinstatement of the methylation mark at the Xi. In addition, thymidine will be used to

arrest cells at G1-S phase to allow me to investigate whether blocking passage through the cell cycle prevents or in any way affects, reinstatement of the H3K27 methylation mark at the Xi.

4.4 Results

4.4.1 *Ciz1* KO causes disruption of the H3K27me3 methylation mark at the Xi

Detection of the H3K27 methylation mark in a WT and *Ciz1* KO cell line by means of immunofluorescence, revealed loss of the methylation mark at the inactive X chromosome in the e13.15 *Ciz1* KO cell line (Figure 4.1). As shown in Table 4.1, in the wild-type cell line approximately 98% of the cells have CIZ1 high-density patches at the Xi, and 97% of these show a H3K27 patch at the same location. However, in *Ciz1* null cells (e13.15 KO cell line) CIZ1 patches are absent from all cells and the H3K27 methylation mark is also lost from the Xi in 100% of cells.

As seen in Figure 4.1, loss of CIZ1 in the KO cell line seems to not just disrupt the H3K27 methylation mark at the Xi but seems to lead to an overall decrease in the H3K27 staining across the nucleus. This observation suggests that loss of CIZ1 could have a much wider effect on the methylation mark than originally expected.

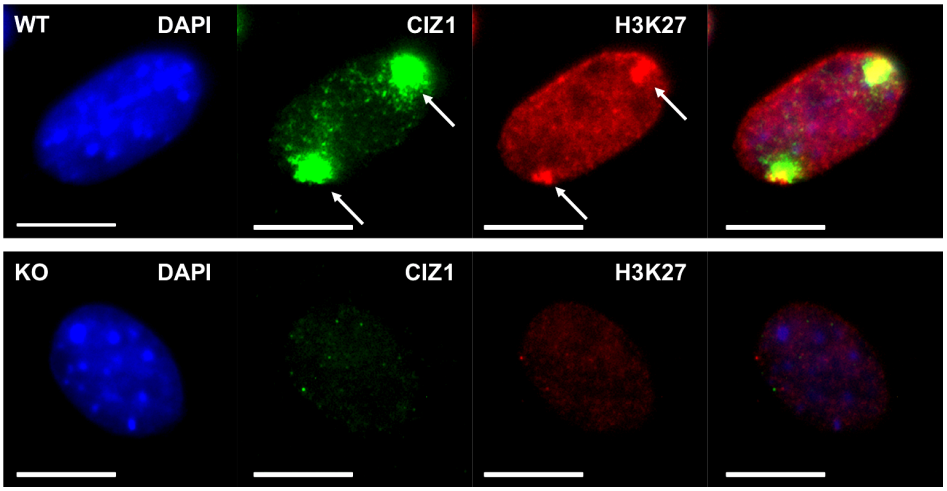


Figure 4.1. Knockout of *Ciz1* leads to disruption of the H3K27me3 methylation mark. Top images show immunofluorescence experiment on e13.1 WT PEF cells. CIZ1 high-density patches (green) at the Xi colocalise with H3K27me3 methylation mark patches (red) at the Xi. Bottom images show immunofluorescence experiment on e13.15 *Ciz1* KO PEF cells. No CIZ1 (green) high-density patches

are observed at the Xi although some residual staining is observed in the nucleus. In the absence of CIZ1 the H3K27me3 mark (red) is lost from the Xi and decreased in the rest of the nucleus. White arrows indicate location of high-density CIZ1 patches and H3K27me3 patches at the inactive X chromosome, respectively. Scale bars represent 10μm.

Immunofluorescence results	Wild-type	Ciz1 KO
% cells with CIZ1 patches at the Xi	98%	0
% cells with H3K27me3 patches at the Xi	97%	0

Table 4.1. Table showing differences in percentage of cells with H3K27me3 mark at the Xi in wild-type versus *Ciz1* KO cell lines. In dark grey, percentage of cells with CIZ1 patches at the Xi in wild-type versus *Ciz1* KO cells. In light grey, percentage of cells with H3K27me3 patches at the Xi in wild-type versus *Ciz1* KO cells. A total of 100 cells was counted for both the wild-type and *Ciz1* KO cell lines.

4.4.2 Re-expression of *Ciz1* in a *Ciz1* KO background leads to reinstatement of the methylation mark

An important question to ask is whether re-expression of *Ciz1* in a KO background can reverse the effects that the absence of CIZ1 has on the H3K27 methylation mark. By exposing the e13.17 *Ciz1* KO cells, which contains a doxycycline inducible Ciz1 transgene, to doxycycline during varying time points, ranging from 6 hours to 48 hours I was able to ask not only whether the methylation mark is reinstated at the inactive X chromosome, but also how long after re-expression of *Ciz1*.

Results show that the methylation mark is in fact present following re-expression of *Ciz1*, and that, as the time of exposure to doxycycline is lengthened, the number of cells with observable CIZ1 high-density patches at the Xi increases, starting at 52% at 6h and then reaching a plateau at approximately 73% at 48h (Figure 4.3). The number of cells with H3K27me3 patches, which colocalise with the CIZ1 patches at the Xi (Figure 4.2), also increases with time although at a slower rate, going from 24% at 6 hours to 60% at 48 hours (Figure 4.3). This result suggests that re-expression of *Ciz1* and recruitment of the protein to the Xi precedes reinstatement of the H3K27 methylation mark.

However, one important anomaly was observed in this experiment. Unlike for the e13.15 *Ciz1* KO cells, where the H3K27me3 mark is lost from the Xi in 100% of cases, in the e13.17 Ciz1 KO cell line, which contains the *Ciz1* transgene, 55% of cells

showed H3K27me3 mark at the Xi in the absence of doxycycline (Figure 4.3), that is in the absence of *Ciz1* re-expression. This could be explained if the *Ciz1* transgene in the e13.17 cell line showed some leaky expression, however as shown later in this thesis, qPCR analysis shows there is no *Ciz1* expression at 0 hours in the absence of doxycycline (Figure 5.2). Interestingly, the number of cells showing H3K27me3 marks at the Xi dropped by more than half immediately following induction of *Ciz1* and then gradually increased again. This data does not fit the simple model suggested by previous data and implies a CIZ1-independent mechanism for deposition of the H3K27m3 mark at the Xi, which is switched off when *Ciz1* is expressed.

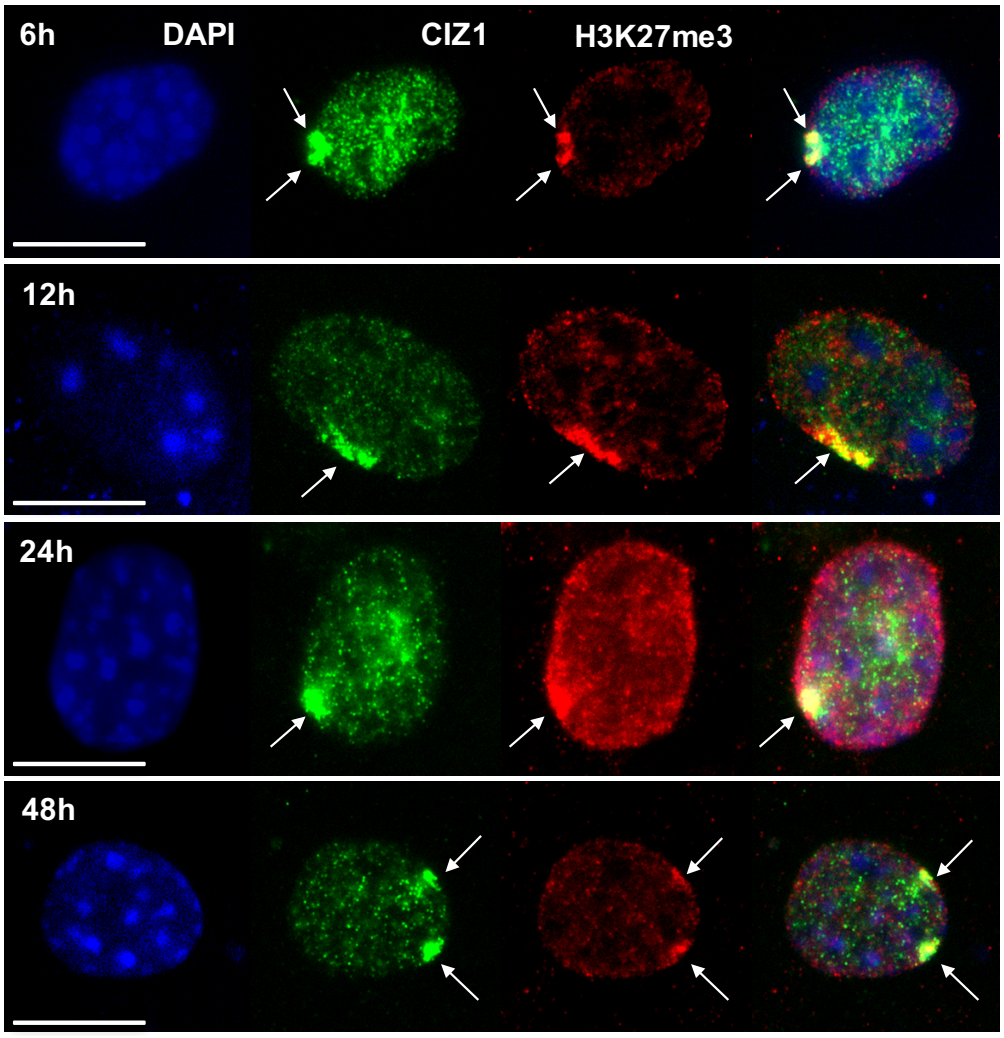


Figure 4.2. Immunofluorescence experiment showing presence of the H3K27me3 mark at the Xi following re-expression of *Ciz1*. Immunofluorescence experiment shows re-expression of CIZ1 (green) and accumulation at the Xi following exposure to doxycycline for periods ranging from 6 to 48 hours. The H3K27me3 mark (red) colocalises with the CIZ1 high-density patch at the Xi. White arrows indicate location of high-density CIZ1 patches and H3K27me3 patches at the inactive X chromosome. Scale bars represent 10μm.

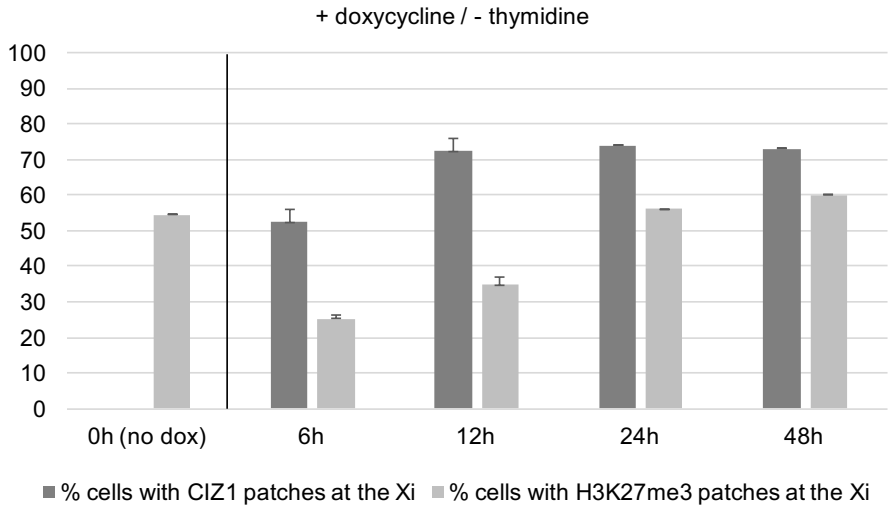


Figure 4.3. Histogram showing percentage of cells with CIZ1 and H3K27me3 density patches at the Xi following re-expression of *Ciz1* in e13.17 *Ciz1* KO cell line. In dark grey, percentage of cells with CIZ1 patches at the Xi. In light grey, percentage of cells with H3K27me3 patches at the Xi. The e13.17 *Ciz1* KO cell lines was exposed to 10μg/mL of doxycycline for time points ranging from 6 to 48 hours. A 0-hour time point, with no doxycycline, was also included. At 0 hours 0% of the cells showed CIZ1 patches at the Xi. Numbers of the 0-hour time point represent an average of two technical replicates with n=100 of each replicate. Numbers of the 6 and 12-hour time point represent an average of three technical replicates with n=100 of each replicate. Error bars represent standard error of the mean. Numbers of the 24 and 48-hour time point represent data from one single replicate with n=100.

4.4.3 Reinstatement of the H3K27me3 mark is initially delayed in non-cycling cells

Another question of interest is whether reinstatement of the H3K27 methylation mark at the inactive X chromosome is dependent on progression through the cell cycle. To address this question, the same experiment as above was repeated but in this occasion the cells were prevented from cycling by adding thymidine to the cell culture media. Again, increasing exposure to doxycycline results in an increase of cells with CIZ1 patches at the inactive X (Figure 4.4), going from 55% at 6 hours to 70% at 48 hours (Figure 4.5). These results do not differ from the minus thymidine experiment, suggesting that recruitment of CIZ1 to the Xi is not a cell cycle dependent phenomenon. Also, as seen previously, this is followed by reinstatement of the H3K27 methylation mark at the Xi. However, in this occasion, the rate at which the methylation mark is laid down is significantly delayed compared to the minus thymidine experiment (Figure 4.5). At 6 hours, with thymidine, only 2% of cells show

H3K27 methylation marks at the Xi, compared to 24% at the same time point without thymidine. Despite this initial delay, by the end of the time course, the total percentage of cells with H3K27me3 marks is not very different between the cycling, 60%, and non-cycling cells, 55%.

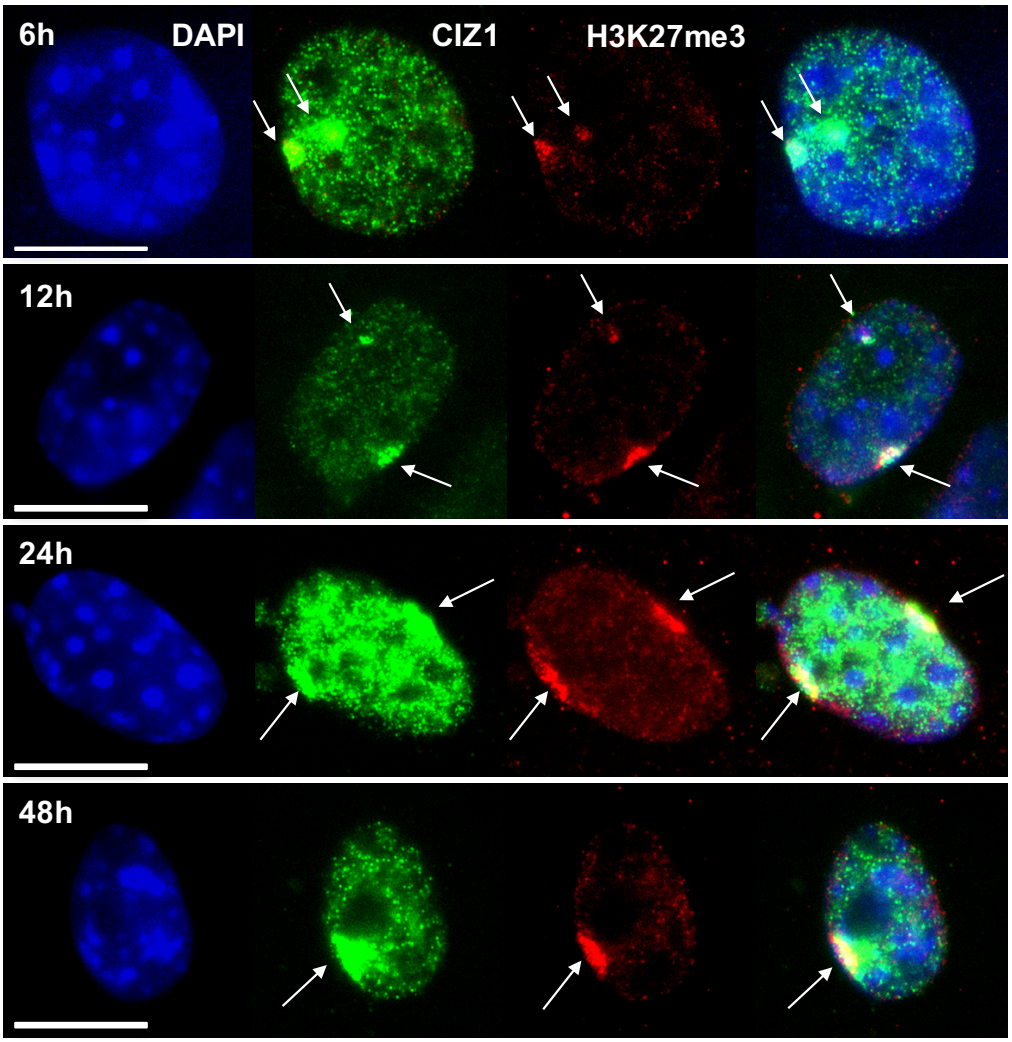


Figure 4.4. Immunofluorescence experiment showing reinstatement of the H3K27me3 mark at the Xi following re-expression of CIZ1 in cell cycle arrested cells. Immunofluorescence experiment shows re-expression of CIZ1 (green) and accumulation at the Xi following exposure to doxycycline for periods ranging from 6 to 48 hours. The H3K27me3 mark (in red) colocalises with the CIZ1 high-density patch at the Xi. White arrows indicate location of high-density CIZ1 patches and H3K27me3 patches at the inactive X chromosome. Scale bars represent 10μm.

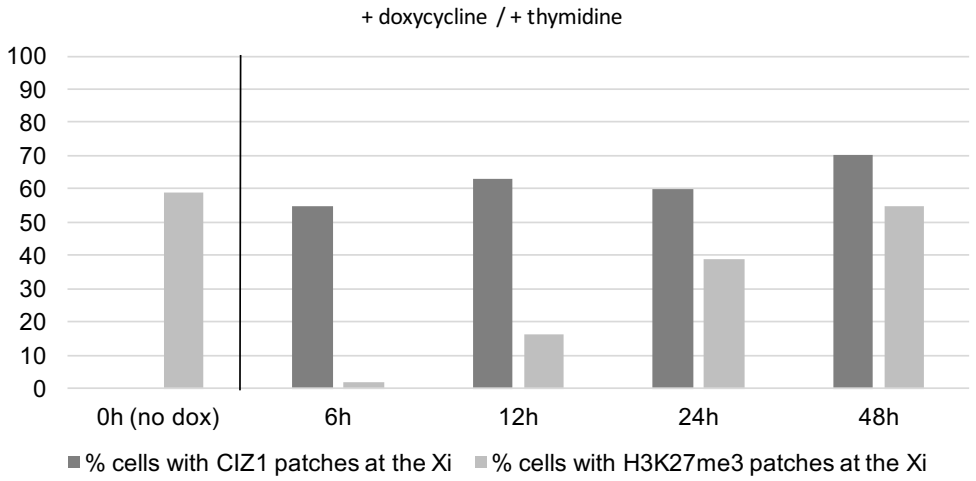


Figure 4.5. Histograms showing reinstatement of the H3K27me3 mark following re-expression of CIZ1 in thymidine treated cells. In dark grey, percentage of cells with CIZ1 patches at the Xi. In light grey, percentage of cells with H3K27me3 patches at the Xi. The e13.17 *Ciz1* KO cell lines was exposed to thymidine and then exposed to 10μg/mL of doxycycline for time points ranging from 6 to 48 hours. A 0-hour time point, with no doxycycline, was also included. At 0 hours 0% of the cells showed CIZ1 patches at the Xi. Numbers showed represent one replicate per time point with n= 100 cells for each time point.

4.5 Discussion

Experiments in this chapter have confirmed disruption of the H3K27me3 mark from the Xi in the absence of CIZ1 protein for one cell line, while in the transgene-containing line the data is more complicated. Nevertheless, my results, together with previous evidence that loss of CIZ1 results in delocalization of Xist from the inactive X, argue that in the absence of CIZ1 important changes occur at the inactive X. Given the anomaly observed at the 0-hour time point in the CIZ1 re-induction experiment it is important to take into account that such drastic disruption of H3K27me3 signal from the Xi might not occur in all cell types. It is possible that in some cell types alternative pathways are in place that compensate for the absence of the protein. Given the fact that the H3K27me3 patches that are observed at the 0-hour time point disappear following induction of CIZ1 with doxycycline it is possible that this putative alternative pathway, is overridden by a CIZ1-dependent mechanism when expression of the protein is restored. However, this is just a possible hypothesis, and the mechanism for H3K27me3 at the Xi in the absence of CIZ1 will have to be further investigated. A possible approach would be to test for absence or presence of the H3K27me3 mark at the Xi in other *Ciz1* KO cell lines, with no transgene, to see whether the methylation

mark is ever present in the total absence of CIZ1. This could be done through immunofluorescence as well as by looking at protein levels on a Western blot. These results could be compared to those obtained when analysing the e13.17 cell line with the transgene. Another possibility that must be considered is the fact that the growth media used to grow the cells could contain traces of tetracycline antibiotics which could be sufficient to induce re-expression of *Ciz1* in the e13.17KO cell line and therefore responsible for the observed methylation mark at a 0h time point.

In addition, loss of CIZ1 also seems to lead to a general reduction of H3K27me3 signal in the rest of the nucleus, at least in the e13.15 *Ciz1* KO cells. This unexpected result suggests that the impact that loss of CIZ1 protein has on the methylation mark might not be specific to the inactive X but that it might have a more widespread effect on the mechanism that allows deposition of the epigenetic mark by PRC2 elsewhere in the nucleus. However, it is crucial to confirm this observation by means of other experiments, such as for example, Western blot.

Regarding the CIZ1 re-induction experiment, I have confirmed that expression of CIZ1 in a *Ciz1* KO background is capable of reinstating the H3K27me3 mark at the Xi. This experiment argues that loss of the methylation mark is a CIZ1 specific phenomenon. Interestingly, blockage of the cell cycle does not affect formation of CIZ1 high-density patches at the Xi, suggesting that movement of CIZ1 to the site occurs through a mechanism that is independent of progression through the cell cycle. Blockage of the cell cycle does not prevent reinstatement of the H3K27 methylation mark either but rather just delays it. This delay occurs between the 0 and 6-hour time point, where the rate of H3K27me3 mark reinstatement is significantly higher in the non-arrested cells. This result could suggest that following re-expression of CIZ1 a CIZ1-dependent pathway comes into place in order to reinstate the H3K27me3. However, it is possible that this pathway is at the same time dependent on progression through the cell cycle, and since the cells have been arrested with thymidine, a second pathway, which is CIZ1- and cell-cycle-independent comes into place and allows fast reinstatement of the methylation mark. It is also possible that this is the same pathway that allows deposition of the mark at the 0-hour time point in the absence of CIZ1. Nevertheless, this is one of many other possible scenarios that could explain this phenomenon and further research will have to address this question in a more specific manner. It is also of interest to investigate the reasons why the e13.15 *Ciz1* KO cell line is able to grow and survive with a disrupted

H3K27me3, which is unusual, given the importance of this repressive epigenetic mark on basic processes such as cell proliferation and differentiation (Margueron and Reinberg, 2011). It is possible that in this cell line levels of the methylation mark have been dramatically reduced but are still sufficient to allow survival of the cell line so it would be interesting to compare the levels of protein by Western blot between the e13.15 cell line to those of a wild-type. It could also be of interest to test whether the e13.15 cell line shows any anomalies in cell proliferation or markers of differentiation.

CHAPTER 5. GENOME-WIDE EXPRESSION ANALYSIS

5.1 Introduction

Previous work has shown that loss of *Ciz1* leads to changes at the inactive X, including delocalisation of *Xist* and disruption of the H3K27 methylation mark. However, whether these events lead to alterations at the level of gene expression from the Xi remains unknown. In addition, the effects that knockout of *Ciz1* has on the rest of the genome are also not known.

5.2 Aims

The aims of this experiment are divided into two main sections involving i) analysis of gene expression from the X chromosome and ii) analysis of gene expression from the whole genome.

- X chromosome analysis:
- Obtain an overview of gene expression from the X chromosome in wild-type cells compared to *Ciz1* knockout cells.
- Determine whether there is any evidence of gene reactivation from the Xi as a consequence of loss of *Ciz1*.
- Investigate whether, in the absence of *Ciz1*, expression from genes at the X inactivation centre is affected.

- Whole-genome analysis
- Investigate effects of *Ciz1* KO on genes from the whole genome.
- Identify potential gene sets that are significantly affected by *Ciz1* KO
- Investigate possible links between loss of *Ciz1* and the female specific lymphoma-like phenotype observed in *Ciz1* KO mice.

5.3 Experimental design

These questions will be addressed by means of a high-throughput RNA sequencing experiment, which involved the following steps (Figure 5.1):

- Extraction of RNA from a set of selected wild-type and Ciz1 KO cell lines (see below).
- Analysis of integrity of the RNA samples.
- Preparation of cDNA libraries for sequencing.
- Sequencing by Illumina 3000 system.
- Analysis of results by me with support from bioinformatics specialist in York Biology Technology Facility, Dr Peter Ashton and Dr. Richard Randle-Boggis.

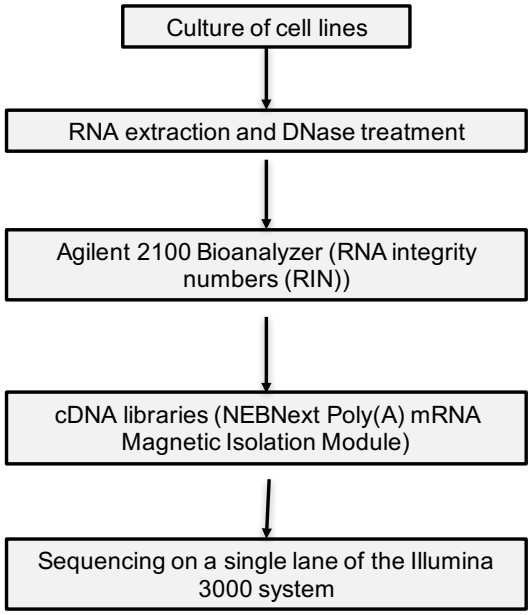


Figure 5.1 Schematic showing laboratory work-flow for the RNA sequencing experiment.

5.4 Results

5.4.1 Preliminary work for RNA sequencing analysis

5.4.1.1 Selection of cell lines for RNA sequencing samples

The main aim of the RNA sequencing experiment is to compare levels of gene expression between a wild-type and a *Ciz1* knockout condition, using primary embryonic mouse female fibroblasts (PEFs). In order to obtain statistically significant results three different cell lines of each genotype where selected for the experiment (see Methods, Table 2.1). Each cell line within the wild type and knockout selection

is derived from a different mouse (cell lines generated by Dr. J Ainscough), which will allow accounting for biological differences in gene expression between individuals.

In order to begin to address whether re-expression of *Ciz1* (in a *Ciz1* KO background) is able to reverse any effects on gene expression, the e13.17 *Ciz1* KO cell line, which contains a doxycycline inducible *Ciz1* transgene, was also included (see Methods, Figure 2.1). In addition, in order to account for any possible direct effects that doxycycline could have on gene expression, a normal *Ciz1* KO cell line, e13.15, was also exposed to doxycycline and included in the RNAseq analysis (see Methods, Table 2.1).

Another question of interest is whether there are any differences in gene expression from the X chromosome between female and male cell lines and whether these differences could help explain the female specific lymphoma-like phenotype. If results do not show that the lymphoma phenotype could be linked to X inactivation, other differences between female and male gene expression in the rest of the genome could also be of interest for the same reason. In order to begin to investigate these questions, a single cell line each derived from a wild-type and a knockout male mouse were included in the samples for sequencing (see Methods, Table 2.1).

Finally, it is also of interest to begin to assess whether the gene expression obtained in cultured cell lines does in any way compare to that in tissue samples obtained directly from the mice. For this reason, a tissue sample from the spleen of two female mice, one wild-type and one knockout, was also included. The knockout female mouse presented a lymphoma-like phenotype (see Methods), in which one of the affected organs is the spleen (Dr. Justin Ainscough, personal communication).

In order to ensure minimal variability in gene expression levels due to differences in treatment of the cells all selected cell lines were plated and grown to 80% density and no further than passage number 4 before homogenization with TRIzol (see Methods).

5.4.1.2 Testing ability of the e13.17 KO cell line to express GFP-CIZ1 upon induction with doxycycline

A quantitative (real-time) PCR experiment was performed preliminary to the RNA sequencing experiment. One of the aims of this experiment is to confirm that addition of doxycycline to the media of e13.17 KO cells line does in fact result in expression of *Ciz1.* Although expression of GFP-CIZ1 has been previously confirmed by means of immunofluorescence experiments it is crucial to look at levels of expression. In addition, I will investigate the effect that different exposure times to doxycycline, ranging from six to seventy-two hours, has on relative levels of *Ciz1* transcript, in order to determine the most appropriate time point for RNAseq analysis. Results of this experiment will also provide useful information on levels of *Ciz1* expression in relation to *Xist* expression, in the same time frame.

Results of the qPCR experiment show significant expression of *Ciz1* in the e13.17 KO cell line following addition of doxycycline, compared to the e13.17 KO cell line with no doxycycline, which shows no *Ciz1* expression (Figure 5.2). This shows that the cell line is suitable for use in the RNA sequencing experiment. It is also important however to choose a time point, because financial constraints limited my RNAseq analysis to 12 samples in total. Levels of *Ciz1* expression reach a peak at 24 hours, then decrease at 48, and then increase again at 72 hours (Figure 5.2). It is important to note the overwhelming levels of *Ciz1* compared to wild-type levels (e14.4), over 90 times higher at 24 hours, so when analysing the RNA sequencing data, it will be necessary to check whether the same phenomenon is observed. Levels of expression of *Xist* slightly decrease across the time course from 6 to 72 hours. However, the two controls with no dox at the beginning (0i) and end (0f) of the experiments do not remain constant, suggesting this decrease might be related to other factors such as for example, cell density. Although these might be interesting observations, results from the RNA sequencing experiment will provide further and more useful information on the levels of expression of *Xist* in the presence and absence of *Ciz1* expression.

Although there is an apparent drop at the 48-hour time point, I chose this for my RNA sequencing as it represents a time after induction at which all the cells in the analysis can be expected to have passed through at least one cell cycle, and so have implemented any nuclear organizational changes that might result from the presence of CIZ1.

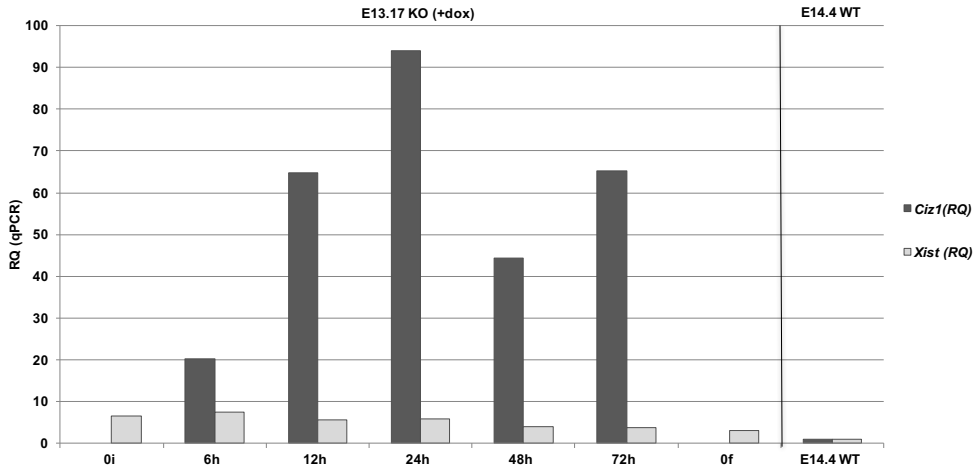


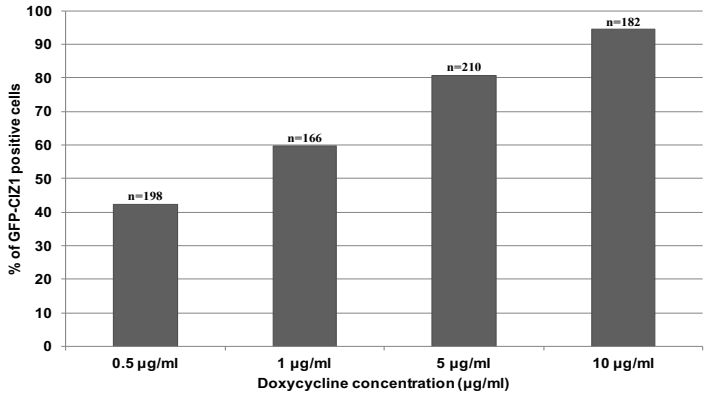
Figure 5.2. *Ciz1* is highly expressed following addition of doxycycline to the media of e13.17 *Ciz1* KO cells. Histogram showing levels of *Ciz1* expression (dark grey) following induction with doxycycline compared to no expression of *Ciz1* in un-induced controls (0i, 0f). Levels of *Xist* expression are indicated by light grey bars. Levels of expression are expressed relative to wild-type levels, e14.4 WT, which was used as the calibrator.

5.4.1.3 Investigating the effects of doxycycline concentration on induction of Ciz1 expression

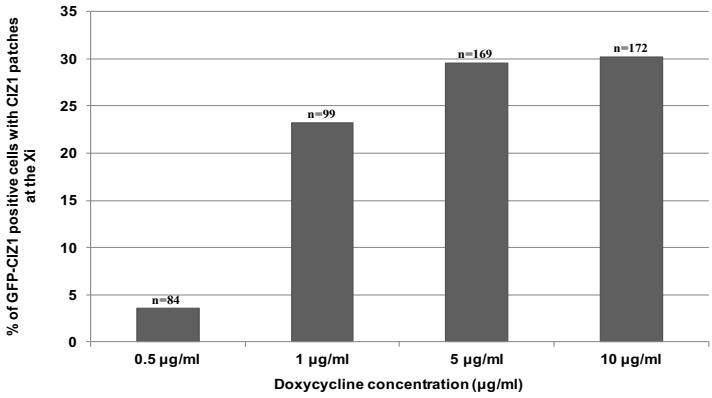
Doxycycline, which is a tetracycline antibiotic commonly used in inducible gene expression systems, has been previously reported to have off-target effects on gene expression and metabolism (Ahler et al, 2013). For this reason, I tested a range of doxycycline concentrations on the e13.17 KO inducible cells, ranging from 0.5µg/mL to 10µg/mL. The aim of this experiment is to identify the lower concentration of doxycycline that can efficiently induce *Ciz1* expression. Output was quantified based on the efficiency of formation of GFP-CIZ1 high-density patches at the inactive X, and was also scored for overall expression of GFP-CIZ1.

A concentration of 10µg/mL of doxycycline achieved induction of *Ciz1* expression in over 90% of the cells while 5µg/mL of doxycycline induced approximately 80% of the cells to express GFP-CIZ1 (Figure 5.3A). The percentage of GFP positive cells with correspondent GFP-CIZ1 high-density patches at the Xi however does not greatly differ between both concentrations (Figure 5.3B) (Figure 5.3C). For this reason, and to minimize possible direct side effects of doxycycline on gene expression the lower, sub-saturating, concentration of 5µg/mL was finally selected for use in the RNA samples for sequencing.

A)


% of GFP-CIZ1 positive cells
100
90
80
70
60
50
40
30
20
10
0
n=198
n=166
n=210
n=182
0.5 µg/ml
1 µg/ml
5 µg/ml
10 µg/ml
Doxycycline concentration (µg/ml)

B)


% of GFP-CIZ1 positive cells with CIZ1 patches at the Xi
35
30
25
20
15
10
5
0
n=84
n=99
n=169
n=172
0.5 µg/ml
1 µg/ml
5 µg/ml
10 µg/ml
Doxycycline concentration (µg/ml)

C)

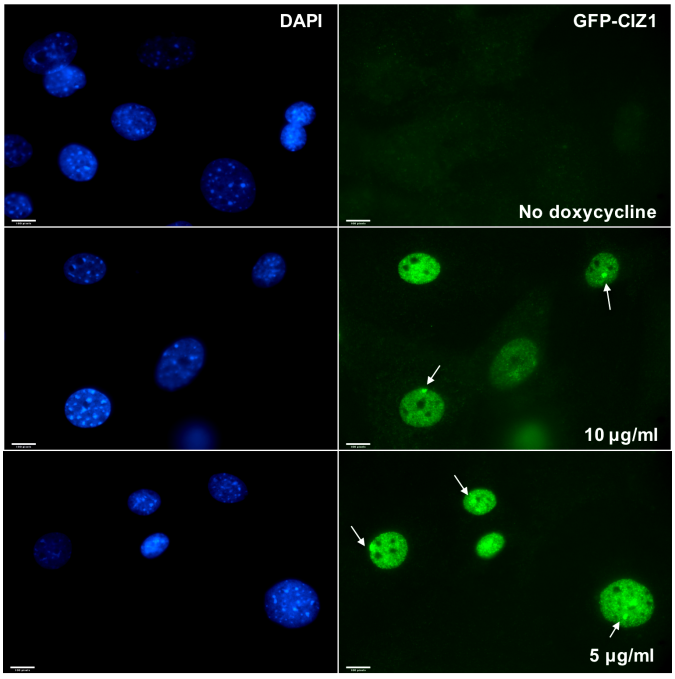

DAPI
GFP-CIZ1
No doxycycline
10 µg/ml
5 µg/ml

Figure 5.3A. Histograms showing ability of varying concentrations of doxycycline to induce expression of GFP-CIZ1 in e13.17 *Ciz1* KO cells. The e13.17 *Ciz1* KO cell line was exposed to concentrations of doxycycline varying from 0.5µg/mL to 10µg/mL and the percentage of GFP-CIZ1 positive cells was assessed for each concentration. Data labels (n = x) indicate the total number of cells that was counted for each concentration. **Figure 5.3B. Histograms showing ability of varying concentrations of doxycycline to induce expression of Ciz1 and formation of GFP-CIZ1 high-density patches at the Xi in e13.17 *Ciz1* KO cells.** The e13.17 *Ciz1* KO cell line was exposed to concentrations of doxycycline varying from 0.5µg/mL to 10µg/mL and the percentage of GFP-CIZ1 positive cells with GFP-CIZ1 patches at the Xi was assessed for each concentration. **Figure 5.3C. Example images of e13.17 *Ciz1* KO cell lines after exposure to doxycycline.** Top images show control experiment of e13.17 *Ciz1* KO without doxycycline exposure. Middle and bottom images show e13.17 *Ciz1* KO cell lines after being exposed to 5µg/mL and 10µg/mL of doxycycline, respectively. GFP-CIZ1 (green) is expressed following induction with doxycycline. DNA was stained with DAPI (blue). White arrows indicate GFP-CIZ1 patches at the inactive X. Scale bars represent 10µm.

5.4.1.4 RNA integrity analysis

Following RNA extraction and DNase treatment (see Methods), I checked the integrity of my DNase treated RNA samples by running an aliquot of the RNA samples on an agarose gel, stained with ethidium bromide (Figure 5.4). The presence of two clear bands, corresponding to the 28s and 18s rRNA bands, together with the absence of smears, indicates that the RNA is, in fact, intact. RNA samples were further analysed using a NanoDrop spectrophotometer in order to obtain the concentration of RNA (Figure 5.5) and the A260/280 values (Figure 5.4), which ideally should be as close to 2 as possible for pure RNA. All 12 samples showed good concentration and integrity values and were therefore transferred to the Genomics facility at University of York, where they were run through an Agilent 2100 Bioanalyzer (see Methods). RNA integrity numbers (RIN) (Sup. info) confirmed the quality of the samples and so the cDNA libraries were prepared and then sent for sequencing (see Methods).

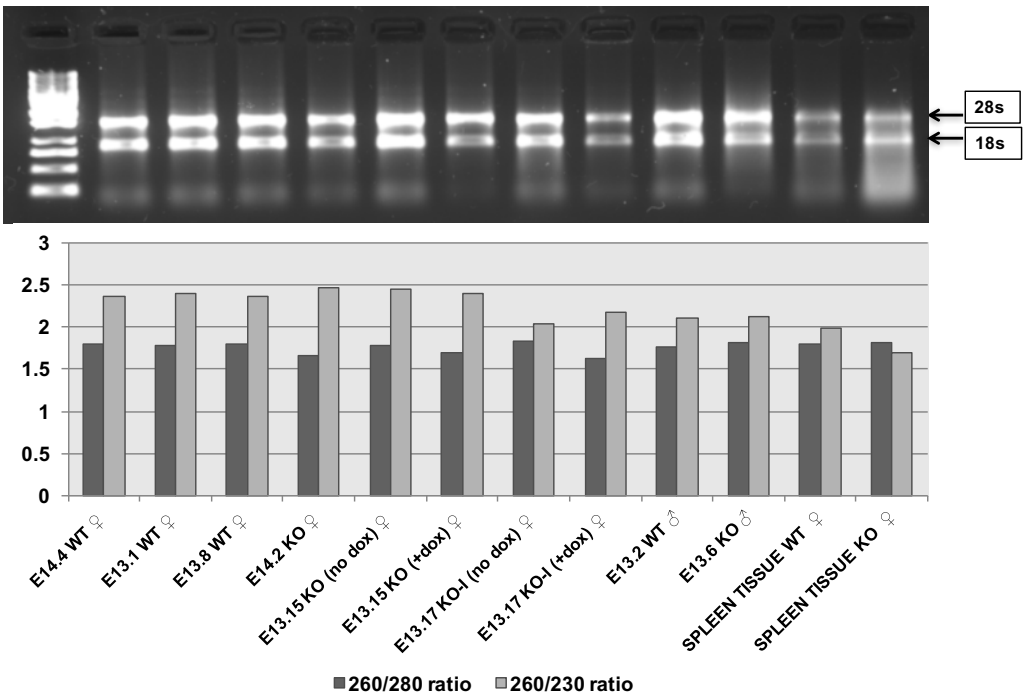


Figure 5.4 Analysis of RNA integrity of RNA samples for sequencing. Top image shows RNA samples run on an agarose gel. Bands corresponding to the two ribosomic RNA bands are indicated. Order of the samples corresponds with the order indicated in the histogram below. Histograms show the 260/280 ratio (dark grey) and the 260/230 ratio (light grey) as indicated by the NanoDrop spectrophotometer.

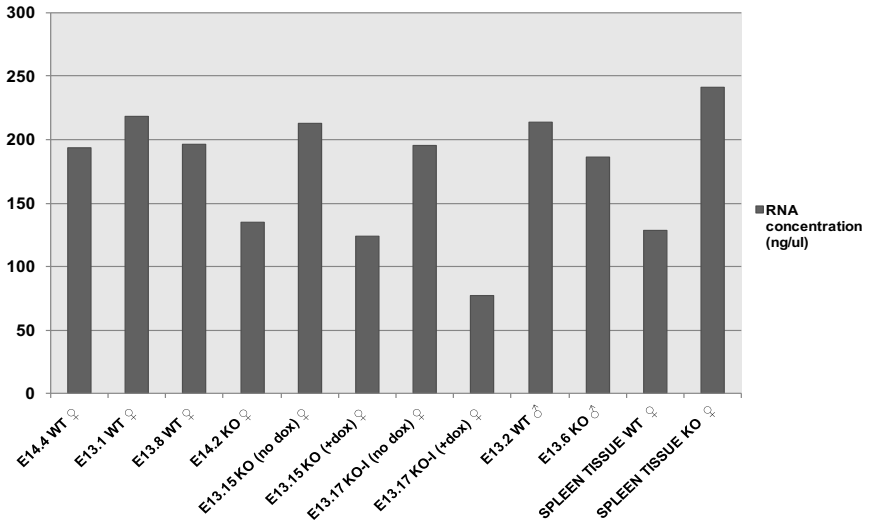


Figure 5.5. Final RNA concentrations of samples for RNA sequencing. Histograms show final concentrations of RNA of all the RNA sequencing samples after DNase treatment.

5.4.2 Analysis of *Ciz1* expression

Since the investigation in this thesis revolves around *Ciz1* it is important to first analyse the levels of gene expression of *Ciz1* across the samples submitted for sequencing. This is important to control for possible anomalies and to ensure that the genetically modified cell lines have been successfully modified, given the design of the KO model (exon 1 highjack), and the propensity for alternative splicing in the *Ciz1* gene. Although the *Ciz1* KO cell lines and the *Ciz1* KO cell line with the inducible transgene (e13.17 *Ciz1* KO), underwent checks when they were generated it is important to confirm it before starting to analyse the RNA sequencing results. As seen in Table 5.1, the three wild-type cell lines show very similar levels of expression which shows that *Ciz1* expression is relatively consistent across these three individual wild-type mice. In all three of the *Ciz1* KO cell lines there is some *Ciz1* expression but at significantly lower levels (p= 5.00E-05; q = 0.00515895), which are again similar across all three individuals. This shows that the *Ciz1* gene, or most of it has been successfully suppressed with only a very small amount of leaky expression (Table 5.1). This may also be 'noise' in the system and represent inaccurately mapped transcripts.

Regarding the e13.17 *Ciz1* KO inducible cell line, following induction with doxycycline, used at a concentration of 5µg/mL and for a period of 48 hours, Ciz1 is expressed (Table 5.1). However, levels of expression are overwhelmingly higher compared to the wild-type levels of expression, in agreement with prior qRT-PCR analysis.

In the male cell lines, levels of *Ciz1* expression are very similar to the female wild-type (Table 5.1), suggesting there is no significant differences between the sexes. The same occurs for the female and male *Ciz1* KO cell lines which again have very similar levels of low-level leaky *Ciz1* expression.

Finally, in the wild-type female spleen tissue sample the level of expression of *Ciz1* is almost double that observed in the female cell lines (Table 5.1). However, it is not possible to make a direct comparison between the two as they represent unrelated cell types. The spleen tissue sample from the *Ciz1* KO female mouse however, has very similar low level expression compared to the female *Ciz1* KO cell lines (Table 5.1).

	Replicate no.	Sample	FPKM	Average	p-value	q-value
Wild-type	0	E14.4 WT ♀	19.5346	21.9871	5.00E-05	0.00515895
	1	E13.1 WT ♀	23.2416			
	2	E13.8 WT ♀	23.1852			
Ciz1 KO	0	E14.2 Ciz1 KO ♀	0.270572	0.322298		
	1	E13.15 Ciz1 KO ♀	0.362798			
	2	E13.17 Ciz1 KO ♀	0.333524			
+ doxycycline	-	E13.15 Ciz1 KO (+dox) ♀	0.250912	0.250912	-	-
	-	E13.17 Ciz1 KO (+dox) ♀	330.331	330.331	-	-
Male samples	-	E13.2 WT ♂	22.1556	22.1556	-	-
	-	E13.6 KO ♂	0.242394	0.242394	-	-
Spleen tissue	-	Spleen WT ♀	41.3976	41.3976	-	-
	-	Spleen KO ♀	0.250324	0.250324	-	-

Table 5.1. Table showing levels of gene expression of *Ciz1* across all RNA sequencing samples. For cases where more than one replicate was used, such as comparison of wild-type versus *Ciz1* KO female cell lines, the number assigned to each replicate in the RNA sequencing is indicated (0-2). Levels of gene expression are indicated in Fragments Per Kilobase of transcript per Million mapped reads (FPKM values). p and q values (see Methods) for the WT vs *Ciz1* KO female comparison are also indicated.

5.4.3 Analysis of RNA sequencing data from X chromosome

5.4.3.1 Analysis of gene expression from the X chromosome

RNA sequencing data was initially processed by members of the bioinformatics group at the technology facility, which involved alignment of the reads to the genome of the C57BL/6 mouse, transcriptome assembly and analysis of differential gene expression (see Methods). Using the output of this analysis I created a summary table of results from the X chromosome (Table 5.2). This shows that gene expression between the wild-type and *Ciz1* KO conditions remains predominantly unchanged, but that 2% of X-linked transcription units are de-regulated (Figure 5.6). From a total of 4273 transcription units 85 have a p-value equal to or less than 0.05 (Table 5.2). A total of 23 genes were manually excluded from this list, where no gene expression was seen in any of three replicates of one of the conditions, and two of the replicates of the other condition. Although a p and q value (see Methods) can still be calculated by the Cuffdiff software these genes were filtered out since, from a biological point of view, they are of no interest. Following this filtering, a total of 62 transcription units with p-

value equal or less than 0.05 remained, of which 6 have positive q-values (Figure 5.7).

From these 62 transcription units, 34 are downregulated in the *Ciz1* KO cell lines, while 28 are upregulated (Figure 5.7). This almost even distribution suggests that changes occurring across the X chromosome may not be directly linked to reactivation of the Xi, but are more likely to represent genome-wide changes occurring as a consequence of loss of *Ciz1*. If gene reactivation from the inactive X chromosome was occurring we would expect most genes to become upregulated in the *Ciz1* KO condition, as a consequence of loss of repression. Even for those transcription units that are up-regulated no strong conclusions can be derived, as it is not possible, in this experiment, to specifically look at transcription from the inactive X chromosome in isolation from overall X chromosome expression. Furthermore, no specific regions within the X chromosome appear to be enriched within this list of 62 affected transcription units (Figure 5.7).

X chromosome
x3 Wild-type vs x3 *Ciz1* Knockout ♀
Total number of transcription units: 4273
Status: OK (1720); NOTEST (2552); FAIL (1)
Significant q-value: 9 , up-regulated (7)/down-regulated (2)
P<0.05: 85; up-regulated (42)/down-regulated (43)
P< 0.01: 41; up-regulated (22)/down-regulated (19)
P<0.005: 29; up-regulated (17)/down-regulated (12)
P<0.001: 14; up-regulated (9)/down-regulated (5)
x1 Wild-type vs x1 *Ciz1* KO spleen ♀
Total number of transcription units: 4273
Status: OK (2209); NOTEST (2063); FAIL (1)
Fold change > 0.3, < -0.3: 1986; up-regulated (960)/ down-regulated (1026)
Fold change > 2, < -2**:** 827: up-regulated (420)/ down-regulated (407)
x1 E13.17 *Ciz1* KO (+dox) vs x1 E13.17 *Ciz1* KO (-dox) ♀
Total number of transcription units: 4273
Status: OK (1731); NOTEST (2542); FAIL (0)
Fold change > 0.3, < -0.3: 1458 ; up-regulated (733)/ down-regulated (725)

Fold change > 2, < -2: 487: up-regulated (246)/ down-regulated (241)
x1 E13.15 *Ciz1* KO (+dox) vs x1 E13.15 *Ciz1* KO (-dox) ♀
Total number of transcription units: 4273
Status: OK (1843); NOTEST (2417); FAIL (13)
Fold change > 0.3, < -0.3: 1725; up-regulated (840)/ down-regulated (885)
Fold change > 2, < -2: 629: up-regulated (266)/ down-regulated (363)
x1 E13.2 WT vs x1 E13.6 *Ciz1* KO ♂
Total number of transcription units: 4273
Status: OK (1708); NOTEST (2555); FAIL(10)
Fold change > 0.3, < -0.3: 1385; up-regulated (683)/ down-regulated (702)
Fold change > 2, < -2: 410: up-regulated (186)/ down-regulated (224)

Table 5.2. Table showing overview of RNA sequencing results for the X chromosome. For all comparisons where statistical results cannot be taken into account due to only having one replicate of each condition a threshold of +/- 0.3 fold change was applied to determine those genes with possible significant changes. This threshold was obtained by looking at the minimum fold change that rendered a significant p-value in the X chromosome and whole genome WT vs *Ciz1* KO comparison, which have three replicates. In addition, the number of genes that are changed greater than 2 fold are indicated. For definitions of status terms (OK, NOTEST, FAIL) as well as transcriptional units see Methods.

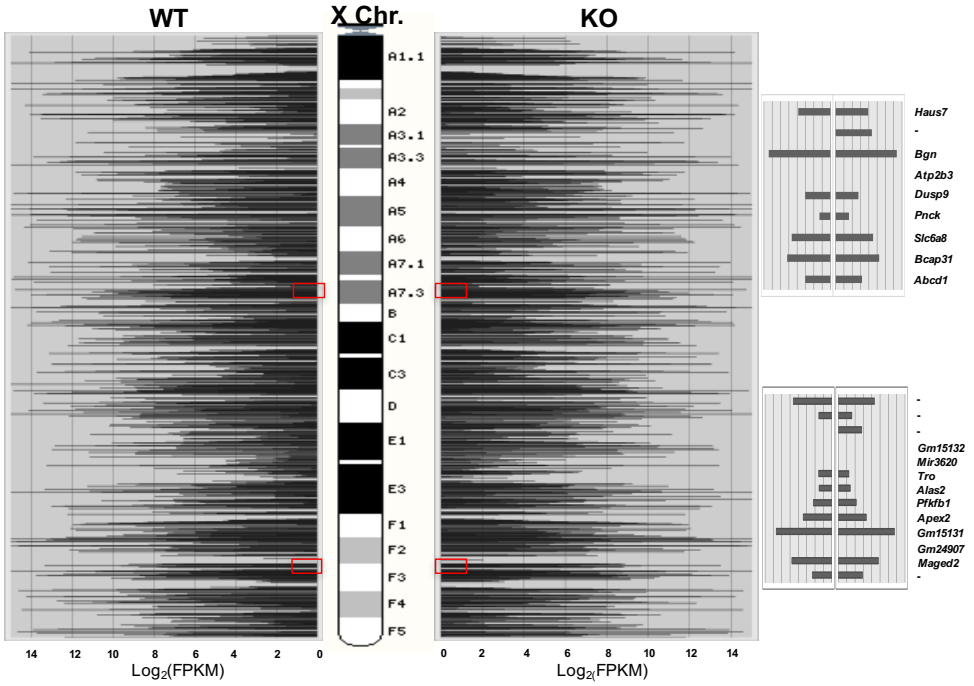


Figure 5.6. Histograms showing mean levels of gene expression in log scale, for all transcription units on the X chromosome, mapped against a schematic of the X chromosome, for wild-type cells (left) and *Ciz1* KO cells (right). Small histograms on the right hand side represent examples of regions where changes can be seen between the wild-type and *Ciz1* KO means. These areas are marked in red in the main figure. Hyphens are used to indicate non-annotated genes. Genes where no expression is shown are examples of NO TEST outcomes.

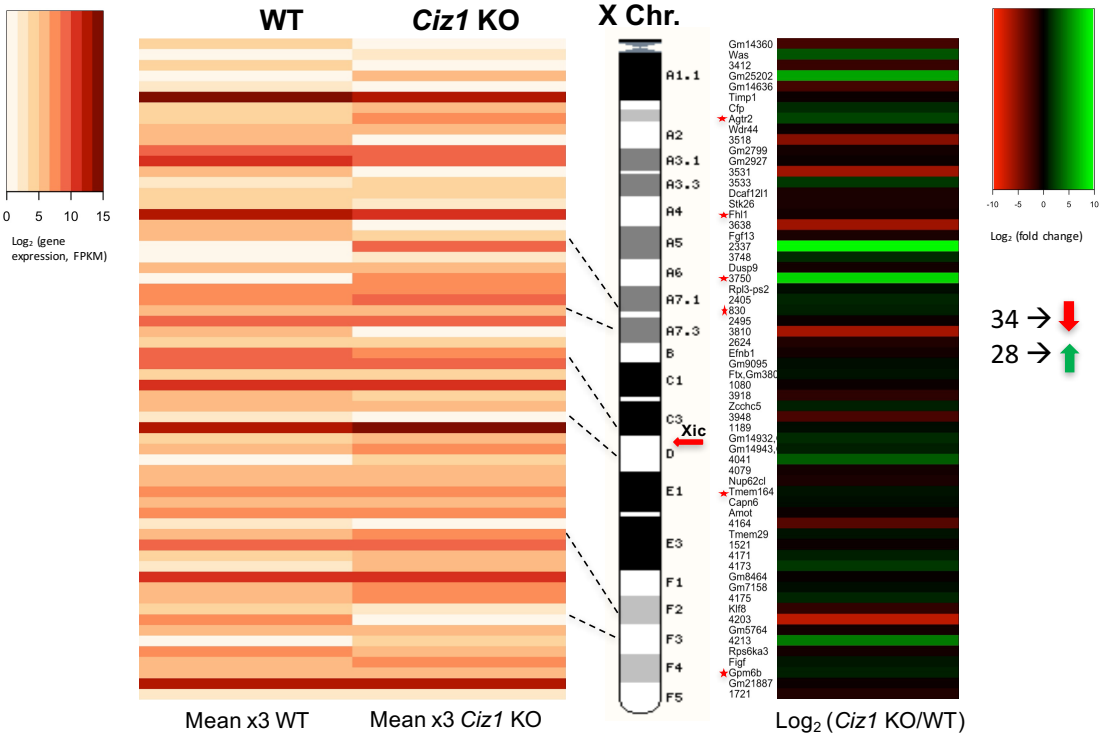


Figure 5.7. Heatmaps showing gene expression data for the 62 X-linked transcription units that are significantly changed (p<0.05) between wild-type and *Ciz1* KO cell lines. Left, levels of gene expression (FPKM, with orange colour coded key), in log scale, plotted against an annotated schematic of the X chromosome (middle). Levels of gene expression represent the average of three replicates both for the WT and *Ciz1* KO. Right, heat map shows fold-change in gene expression with red-green colour scale (red: downregulated; green: upregulated). Genes with a significant q-value are starred. Hyphens are used to indicate non-annotated transcription units.

5.4.3.2 Genes from the X inactivation centre are not significantly affected as a consequence of Ciz1 KO

The state of X chromosome repression was further analysed by looking at genes within the X inactivation centre (XIC), with known functions in the X inactivation process (Figure 5.8). No significant changes can be seen, except for *Ftx*, which does have a significant p-value (p = 0.02) but does not have a significant q value (q = 0.603433). It is therefore reasonable to conclude that loss of CIZ1 does not have a direct effect on the well-established repressive mechanism originating at the XIC.

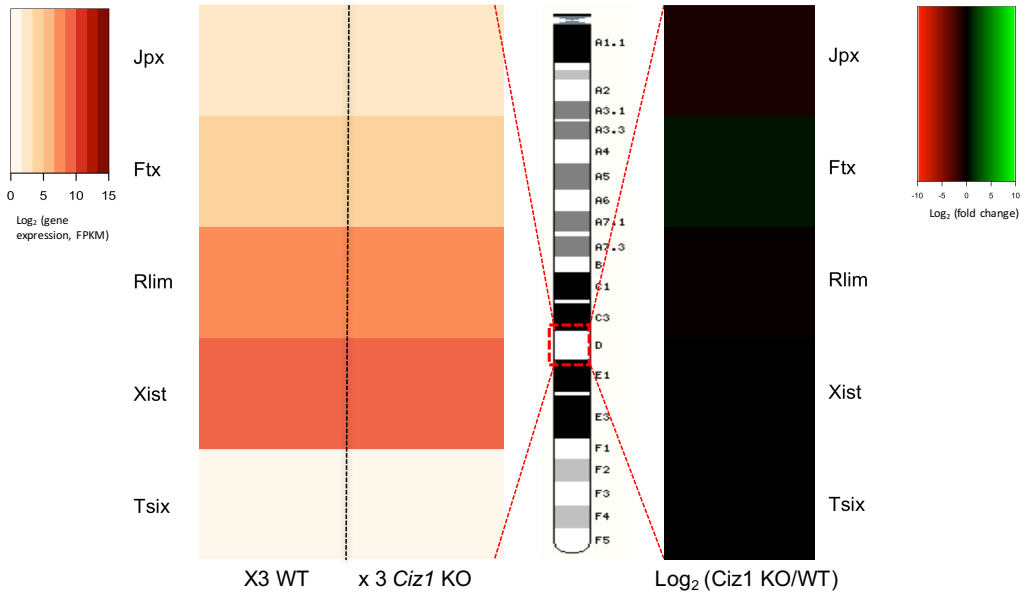


Figure 5.8. Analysis of expression of genes within the X inactivation centre. Left, levels of gene expression (FPKM, with orange colour coded key), in log scale, plotted against an annotated schematic of the X chromosome (middle), where the X inactivation centre is indicated with a red square. Levels of gene expression represent the average of three replicates both for the WT and Ciz1 KO. Right, heat map shows fold-change in gene expression with red-green colour scale (red: downregulated; green: upregulated).

5.4.3.3 Female and male mouse cell lines show different response to Ciz1 KO

The 62 transcription units that are significantly changed ($p<0.05$) in the female cell line following *Ciz1* KO were further analysed by looking at their levels of expression in the male WT and male *Ciz1* KO cell line (Figure 5.9). A total of 33 transcription units are differentially affected by loss of *Ciz1* in female compared to male cells. These were selected based on opposite patterns of gene expression in the female compared to the wild-type, for example a transcription unit that is significantly upregulated in the female C*iz1* KO but downregulated in the male *Ciz1* KO, or vice versa. A further 9 genes were added to the list of differentially expressed genes since although the fold change occurred in the same direction, the difference is at least three times higher in one of the sample sets. That makes a total of 42 genes out of 62 that respond differently to knockout of *Ciz1* in female and male samples. The six genes that have a significant q-value in the female list, and could therefore be of most interest were picked as examples to illustrate this outcome (Figure 5.9). Four out of the six q-positive genes showed differences between female and male cells; *Agtr2,*

Gpm6b, Tmem164 and a non-annotated gene with locus X: 079926530-079993730. These results suggest that although the female specific lymphoma might not be related to X chromosome inactivation other secondary changes that occur as a consequence of the loss of CIZ1, and that affect the female and male cells differently could underlie the phenotype.

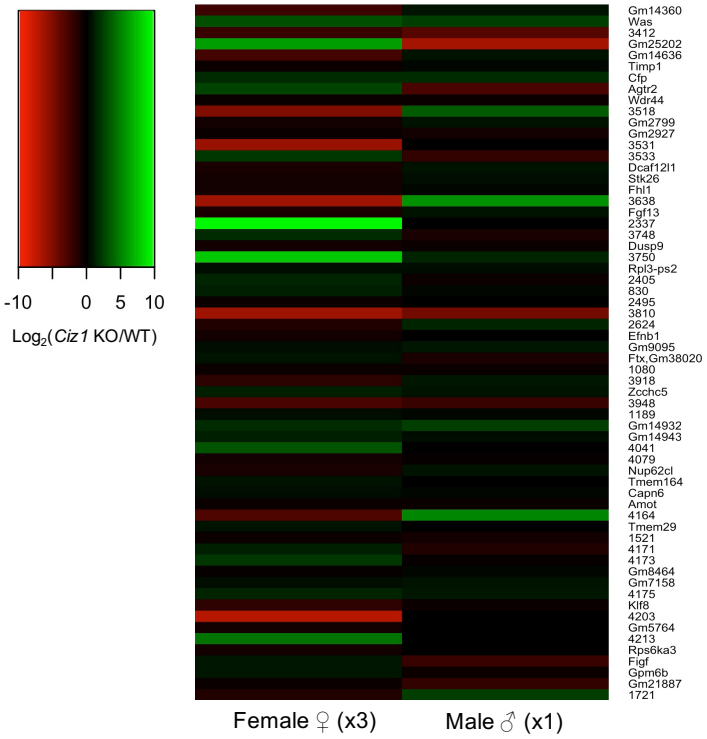


Figure 5.9. Heatmap represents fold-change of gene expression of the 62 genes with significant p-values (p<0.05) in the *Ciz1* KO compared to the wild-type cell lines in female and male samples. Left, heatmap represents fold-change in female samples obtained from the average of three replicates for both WT and *Ciz1* KO cell lines. Right, data for the same genes derived from one single male replicate each for the WT and *Ciz1* KO. A red-green colour scale represents levels of fold change (red: downregulated; green: upregulated). Non annotated genes are indicated with a hyphen.

5.4.3.4 Re-expression of Ciz1 in a Ciz1 KO cell line restores wild-type levels of gene expression for some genes

To test whether re-expression of *Ciz1* in a *Ciz1* KO background is capable of restoring wild-type levels of expression the previously described e13.17 *Ciz1* KO cell line was used. Upon addition of doxycycline to culture media GFP-CIZ1 is expressed and targeted to the Xi, as shown previously for transcript and fluorescent protein (Figure 5.2 and Figure 5.3). The six genes with significant q-values were picked as examples to investigate this question, and further compared to the effect of doxycycline on a

Ciz1 KO cell line that does not contain the transgene (e13.15), to introduce a control for possible direct effects of doxycycline. Any gene with a change in gene expression in the same direction, and greater than 0.5% of that in the transgenic e13.17 *Ciz1* KO cell line was ruled out of further analysis. This affected 44% of the 62 genes but none of the six genes with significant q-values, (data not shown). Since it is also of interest to look for genes that are differentially affected by knockout of *Ciz1* between female and male in order to try and find links with the lymphoma-like phenotype these same six genes were also picked as candidates to look for differences between female and male cells (Figure 5.10).

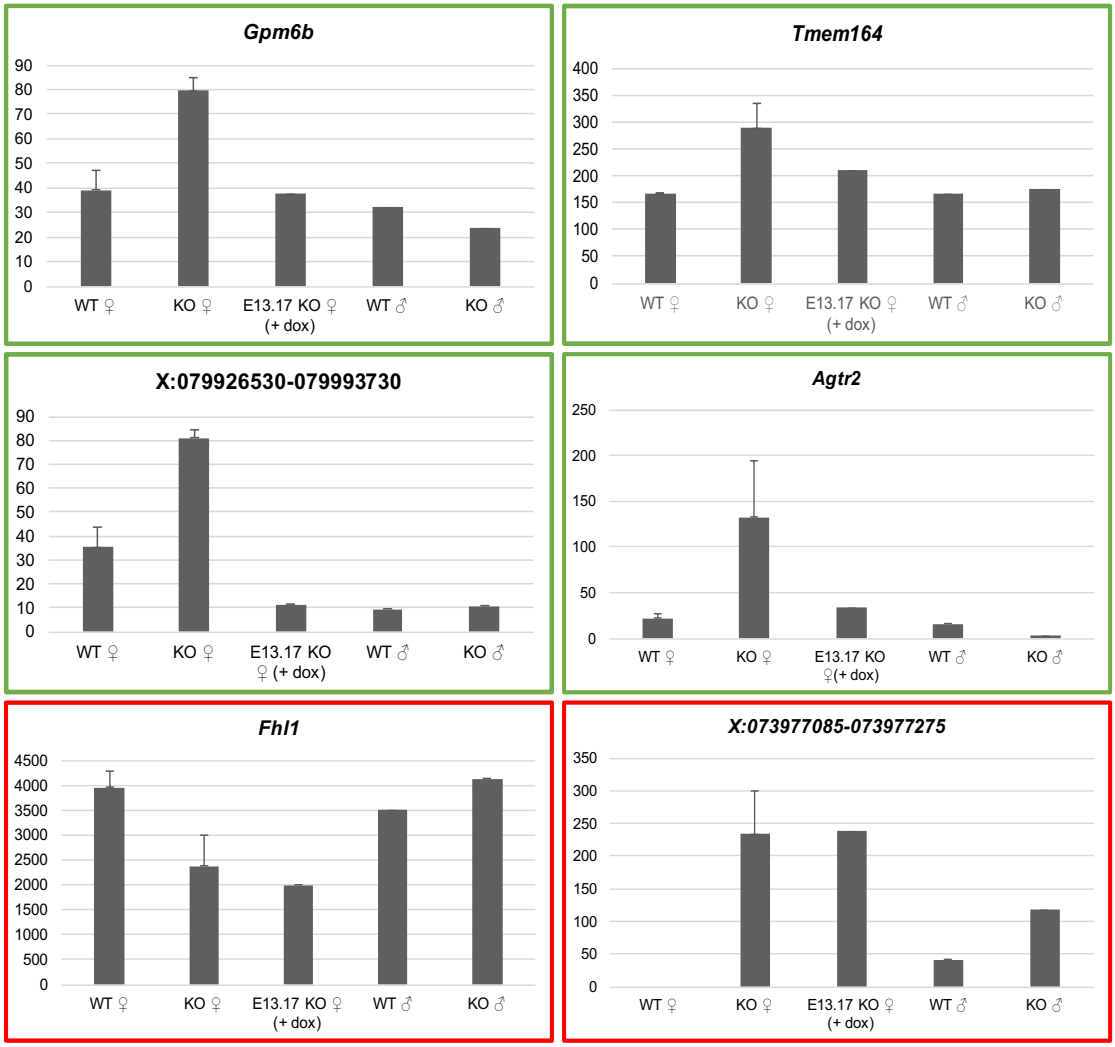


Figure 5.10. Histograms showing analysis of female versus male and *Ciz1* re-induction analysis for the six q-significant transcription units from the X chromosome. For each gene the first bar represents mean female wild-type levels of expression. The next two bars represent levels of expression in *Ciz1* KO female cell lines (mean) and induced e13.17 *Ciz1* KO cell line, respectively. On the right-hand side of each histogram the first bar shows male wild-type levels of expression. The next bar represents levels of expression in *Ciz1* KO male cell line. The female WT and *Ciz1* KO bars represent the average of three replicates while all other bars represent data from one single replicate. Error bars

on the first to bars of each histogram represent standard error of the mean calculated from three replicates.

5.4.3.5 Comparative analysis of X-linked genes affected by loss of CIZ1 with existing literature on X inactivation

In the mouse approximately 3.3% of genes escape XCI and remain transcriptionally active, (Yang et al, 2010), many of which are thought to have female-specific functions that contribute to sexual dimorphism (Berletch et al, 2011). The subset that escape silencing vary between cell type and developmental stage (Gendrel et al, 2012), with some genes (constitutive escapees) avoiding silencing in most cell types, and others (facultative escapees) only in specific contexts. Although no significant overlap of our list with constitutive escapees has been observed other comparisons of interest have been made.

Direct comparison of a subset of our X-linked genes significantly affected by loss of Ciz1 (Table 5.3) with the list of genes obtained from studies of escape from X inactivation revealed that in fact, a total of five genes, have been described as facultative escapees, only escaping X inactivation in specific mouse tissues (Berletch et al, 2015). It has been suggested that genes that escape X inactivation are likely those involved in female specific functions (Berletch et al, 2011). In addition, escape of X inactivation might specifically occur in those tissues where the gene exerts its functions. An example is the Gpm6b gene, which escapes X inactivation in brain tissue, and has been reported to have nervous system related functions, namely in the myelination of the central nervous system (Werner et al, 2013). Interestingly, expression changes in this gene have also been detected in the cerebellum of Alzheimer's mouse models (Bao, 2008), suggesting a possible link with the disease. Interestingly CIZ1 itself has also been implicated in the disease (Dahmcke et al, 2008). Moreover, Alzheimer's is more likely to develop in females than in males (Viña and Lloret, 2010), revealing a sex-bias. This was originally attributed to hormonal differences, however trials involving hormone replacement therapy have not shown any evidence for it (Brinton, 2004). It is therefore possible that other female-specific differences such escape of X inactivation might be accountable for this risk factor difference between the sexes. Another example gene is Cfp which encodes the properdin protein and is involved in positive regulation of complement activation in the immune system (Lesher et al, 2013). This gene, which is significantly affected by loss of CIZ1, escapes X inactivation in the spleen, which is also the tissue from which

the properdin protein is produced (Maves and Weiler, 1992). Interestingly, Cfp has been implicated in several autoimmune disorders (Holers, 2008), which are known to affect women more than men (Fairweather et al, 2008). Importantly, this gene, which is significantly upregulated following Ciz1 KO, also overlaps with the list of genes that are known to be significantly upregulated upon knockdown of Xist in the blood compartment of mice (Yildirim et al, 2013), revealing a possible common mechanism that upon disruption leads to upregulation of the gene.

Two transmembrane protein coding genes, Tmem164 and Tmem29, also overlap with the list of genes of the same study, being defined as genes that escape X inactivation in the ovary (Berletch et al, 2015). This is again interesting, since both genes are expressed in an organ with female specific functions. In addition, Tmem164 gene, which is overexpressed following knockout of Ciz1, is also overexpressed following knockdown of Xist (Yildirim et al, 2013) and Smchd1 (Gendrel et al, 2013).

Genes	Escapees of X inactivation			Other knockdown studies		
	(Berletch et al, 2015)	Others	Studies with no overlap	*Xist* knockdown (Yildrim et al, 2013)	*Firre* knockdown (Yang et al, 2015)	*Smchd1* (Gendrel et al, 2013)
Was			(Yang et al, 2010);PMID: 20363980			
Timp1			(Anderson and Brown, 2002);PMID: 11935340			
Cfp	Tissue specific (spleen)		(Calabrese et al, 2012);PMID: 23178118			
Agtr2			(Li et al, 2012); PMID: 22384067			
Wdr44						
Dcaf12						
Stk26						
Fhl1		(Finn et al, 2014);PMID: 24594094				
Fgf13						
Dusp9						
Rpl3-ps2						
Efnb1						
Ftx	Tissue specific (brain, spleen, ovary)	(Pinter el al, 2012);PMID: 22948768				
Zcchc5						
Nup62cl						
Tmem164	Tissue specific (ovary)					
Capn6						
Amot						
Tmem29	Tissue specific (ovary)					
Klf8						
Rps6ka3						
Figf						
Gpm6b	Tissue specific (brain)					

Table 5.3. Table comparing a subset of X-linked genes significantly affected by Ciz1 KO with results from other X-inactivation studies. First half of the table aims to assess whether the X-linked

gene list is enriched for genes known to escape X inactivation. A list of studies that do not show overlap with my list are also listed (these studies do not correlate to the genes on the first column). Second half of the table compares the list of X-linked genes known to be affected by loss of *Ciz1* with the list of genes that are known to be affected by knockdown of other players of X inactivation such as *Xist*, *Firre*, and *Smchd1* genes. Grey shaded areas indicate overlapping genes in the two lists.

5.4.4 Analysis of RNA sequencing data from the whole genome

5.4.4.1 Overview of gene expression from the whole genome

For the whole genome RNA sequencing data, I also generated a summary table of results (Table 5.3). Initial analysis shows that 2.6% of transcription units are significantly changed between the wild-type and *Ciz1* KO conditions. From a total of 56417 transcriptional units 1480 genes have a p-value equal or less than 0.05 (Table 5.3). This will be referred to as CIZ1-1480. It is important to note that the overall rate of significance is similar between the whole genome, 2.6% and the X chromosome, 2%, suggesting that the previous effect observed at the X chromosome is likely caused via the same mechanisms observed in the whole genome.

Whole genome
x3 Wild-type vs x3 Knockout ♀
Total number of transcription units: 56417
Status: OK (16715); NOTEST (39700); FAIL (2)
Significant q-value: 286, up-regulated (136)/down-regulated (150)
P<0.05: 1480; up-regulated (821)/down-regulated (659)
P< 0.01: 914; up-regulated (564)/down-regulated (350)
P<0.005: 788; up-regulated (500)/down-regulated (288)
P<0.001: 298; up-regulated (142)/down-regulated (156)
x1 Wild-type vs x1 KO spleen ♀
Total number of transcription units: 56417
Status: OK (24869); NOTEST (31539); FAIL (2); LOWDATA (7)
Fold change > 0.3, < -0.3: 24585; up-regulated (13629)/ down-regulated (10956)
x1 E13.17 KO (+dox) vs x1 E13.17 KO (-dox) ♀
Total number of transcription units: 56417

Status: OK (17360); NOTEST (39057); FAIL (0)
Fold change > 0.3, < -0.3: 15722; up-regulated (8621)/ down-regulated (7101)
x1 E13.15 KO (+dox) vs x1 E13.15 KO (-dox) ♀
Total number of transcription units: 56417
Status: OK (18244); NOTEST (38164); FAIL (9)
Fold change > 0.3, < -0.3: 18201; up-regulated (8946)/ down-regulated (9255)
x1 E13.2 WT vs x1 E13.6 KO ♂
Total number of transcription units: 56417
Status: OK (17190); NOTEST (39221); FAIL (6)
Fold change > 0.3, < -0.3: 15220; up-regulated (7712)/ down-regulated (7508)

Table 5.4. Overview of RNA sequencing results for the whole-genome. For all comparisons where statistical results cannot be taken into account due to only having one replicate of each condition a threshold of +/- 0.3 fold change was applied to determine those genes with possible significant changes. This threshold was obtained by looking at the minimum fold change that rendered a significant p-value in the X chromosome and whole genome WT vs KO comparison, which have three replicates. For definitions of status terms (OK, NOTEST, FAIL) as well transcriptional units see Methods, Table 2.7.

5.4.4.2 The list of genes that are significantly affected by knockout of Ciz1 in the whole genome is enriched for PRC1 and PRC2 target genes

Initial analysis of the gene expression data from the whole genome used the Gene Set Enrichment Analysis (GSEA) software. The aim is to look for enriched gene sets within the list of genes with significant changes in the wild-type versus *Ciz1* KO comparison, where $p<0.05$. The GSEA software also applies statistical tests which return both p and q values that reflect the degree of overlap between CIZ1-1480 and specific gene sets. Of the 1480 CIZ1-related transcription units, 821 are up-regulated and 659 are down-regulated (Table 5.3).

CIZ1-1480 was run first through the GSEA computational software, selecting the 'curated' gene set. As defined by the Broad Institute this represents 'gene sets collected from various sources such as online pathway databases, publications in PubMed, and knowledge of domain experts' (Subramanian et al, 2005) (Mootha et al, 2003). This initial analysis returned some striking results, which show that CIZ1-1480 is enriched for genes that are targets of both PRC1 and PRC2 subunits (Table 5.4). The extremely high statistical significance provided by both the p and q values

emphasizes this finding. This link to PRC1 and PRC2 subunits was also obtained when running GSEA 'oncogenic signature' gene set. As defined by the Broad Institute this gene set represents 'signatures of cellular pathways which are often dis-regulated in cancer'. This analysis shows CIZ1-1480 is enriched for genes that are known to be either up-regulated or downregulated upon knockdown of genes that encode subunits of the PRC1 (BMI1 and MEL18), and PRC2 (EZH2) complexes (Table 5.5).

CURATED GENE SETS			
Gene set name	**Description**	**Genes in overlap/ genes in gene set**	**p/q value**
BMI1_TARGETS_UP (Douglas et al, 2008)	Genes up-regulated in A4573 cells (Ewing's sarcoma, ESFT) after knockdown of BMI1 [GeneID=648] by RNAi.	64/566	p =1.26E-29 q = 7.84E-28
SUZ12_TARGETS (Ben-Porath et al, 2008)	Set 'Suz12 targets': genes identified by ChIP on chip as targets of the Polycomb protein SUZ12 [GeneID=23512] in human embryonic stem cells.	120/1038	p = 2.70E-56 q = 1.06E-53
EED_TARGETS (Ben-Porath et al, 2008)	Set 'Eed targets': genes identified by ChIP on chip as targets of the Polycomb protein EED [GeneID=8726] in human embryonic stem cells.	113/1062	p = 2.72E-49 q = 7.57E-47
EZH2_TARGETS_UP (Nuytten et al, 2008)	Genes up-regulated in PC3 cells (prostate cancer) after knockdown of EZH2 [GeneID=2146] by RNAi.	125/1037	p = 9.11E-61 q = 4.78E-58
ES_WITH_H3K27ME3 (Ben-Porath et al, 2008)	Set 'H3K27 bound': genes posessing the trimethylated H3K27 (H3K27me3) mark in their promoters in human embryonic stem cells, as identified by ChIP on chip.	111/1118	p = 1.86E-45 q = 4.00E-43

Table 5.5. Table showing overlap with curated gene sets returned by GSEA. These gene sets include targets of known PRC1 subunits, such as BMI1, as well as targets of known PRC2 subunits, such as SUZ12, EED and EZH2. For a detailed description of GSEA terms see Methods, Table 2.8.

ONCOGENIC SIGNATURES			
Gene set name	**Description**	**Genes in overlap/ genes in gene set**	**Statistical test**
BMI1_DN.V1_UP (Wiederschain et al, 2007)	Genes up-regulated in DAOY cells (medulloblastoma) upon knockdown of BMI1 [Gene ID=648] gene by RNAi.	37/147	p = 2.67E-30 q = 5.05E-28
BMI1_DN.V1_DN (Wiederschain et al, 2007)	Genes down-regulated in DAOY cells (medulloblastoma) upon knockdown of BMI1 [Gene ID=648] gene by RNAi.	13/144	p = 5.44E-06 q = 1.08E-05
MEL18_DN.V1_UP (Wiederschain et al, 2007)	Genes up-regulated in DAOY cells (medulloblastoma) upon knockdown of PCGF2 [Gene ID=7703] gene by RNAi.	36/141	p = 9.34E-30 q = 8.82E-28
MEL18_DN.V1_DN (Wiederschain et al, 2007)	Genes down-regulated in DAOY cells (medulloblastoma) upon knockdown of PCGF2 [Gene ID=7703] gene by RNAi.	30/148	p = 7.23E-22 q = 1.14E-20
PRC2_EZH2_UP.V1_UP (Bracken et al, 2006)	Genes up-regulated in TIG3 cells (fibroblasts) upon knockdown of EZH2 [Gene ID=2146] gene.	16/195	p = 1.66E-06 q = 3.48E-06
PRC2_EZH2_UP.V1_DN (Bracken et al, 2006)	Genes down-regulated in TIG3 cells (fibroblasts) upon knockdown of EZH2 [Gene ID=2146] gene.	23/194	p = 6.34E-12 q =2.49E-11

Table 5.6. Table showing overlap with oncogenic signatures returned by GSEA. These gene sets include genes that are known to be either upregulated or downregulated following knockdown of PRC1 subunits, such as BMI1 and MEL18, as well as PRC2 subunits, such as EZH2. For a detailed description of GSEA terms see Methods, Table 2.8.

The GSEA analysis also returned highly significant overlaps with gene sets implicated in other pathways and diseases such as cancer, however given the vast amount of data derived from this RNAseq data set it is not possible for me to analyze all of it in the course of one year. In addition, the relevance of some of these results in the context of CIZ1 biology is not clear and will require intense research. Furthermore, the hits implicating PRC complexes are of more relevance in this thesis due to their role in the process of X inactivation. In the following analysis, I have chosen to focus on one subunit from the PRC1 complex, BMI1, and one subunit from the PRC2 complex, EZH2, looking at data from both the curated gene set analysis and oncogenic signature analysis.

The curated gene set analysis shows that CIZ1-1480 is enriched for genes that are known to be upregulated in a female cancer cell line (A4573, Ewing sarcoma) following knockdown of BMI1 (Table 5.4). It also shows it to be enriched for genes that are known to be upregulated in prostate cancer cells following knockdown of EZH2 (Table 5.4). These were plotted in a heat map showing expression in the wild-type versus *Ciz1* KO female cells (Figure 5.11). A total of 12 genes appear in both lists, suggesting there is some degree of overlap between these pathways. Interestingly, even though the genes are known to be upregulated upon knockdown of BMI1 and EZH2, respectively, the majority are downregulated upon knockdown of *Ciz1* for both gene sets, suggesting the expression profile that we are observing might be more characteristic of an overexpression of these two subunits rather than a knockdown effect.

The oncogenic signature analysis shows that CIZ1-1480 is enriched for genes that are upregulated in DAOY cells (medulloblastoma) upon knockdown of BMI1. From a total of 147 genes that appear in this gene set, 37 overlap with CIZ1-1480 (p= 2.67E-30, q = 5.05E-28) (Table 5.5). In addition, CIZ1-1480 is also enriched for genes that are downregulated in DAOY cells upon knockdown of BMI1, with a total of 13 genes overlapping out of 144 (p=5.44E-06, q=1.08E-05) (Table 5.5). When looking at levels of expression of the overlapping genes in the wild-type versus *Ciz1* KO, results show that those genes in the downregulated category are upregulated in CIZ1-1480 and vice versa (Figure 5.12A-B). This begins to suggest that the expression profile observed in *Ciz1* KO cells is more characteristic of an overexpression of BMI1 rather than a knockdown or downregulation, and is an unexpected finding.

The oncogenic signature analysis also shows that CIZ1-1480 is enriched for genes that are upregulated in TIG3 cells (fibroblasts) upon knockdown of EZH2. From a total of 195 genes in this gene set, 16 overlap with CIZ1-1480 (p= 1.66E-06, q = 3.48E-06) (Table 5.5). CIZ1-1480 is also enriched for genes that are downregulated in TIG3 cells upon knockdown of EZH2 with a total of 23 genes overlapping out of 194 total genes in the gene set (p= 6.34E-12, q = 2.49E-11) (Table 5.5). When looking at their levels of expression, in both categories the majority of genes are downregulated following *Ciz1* KO. (Figure 5.12A-B).

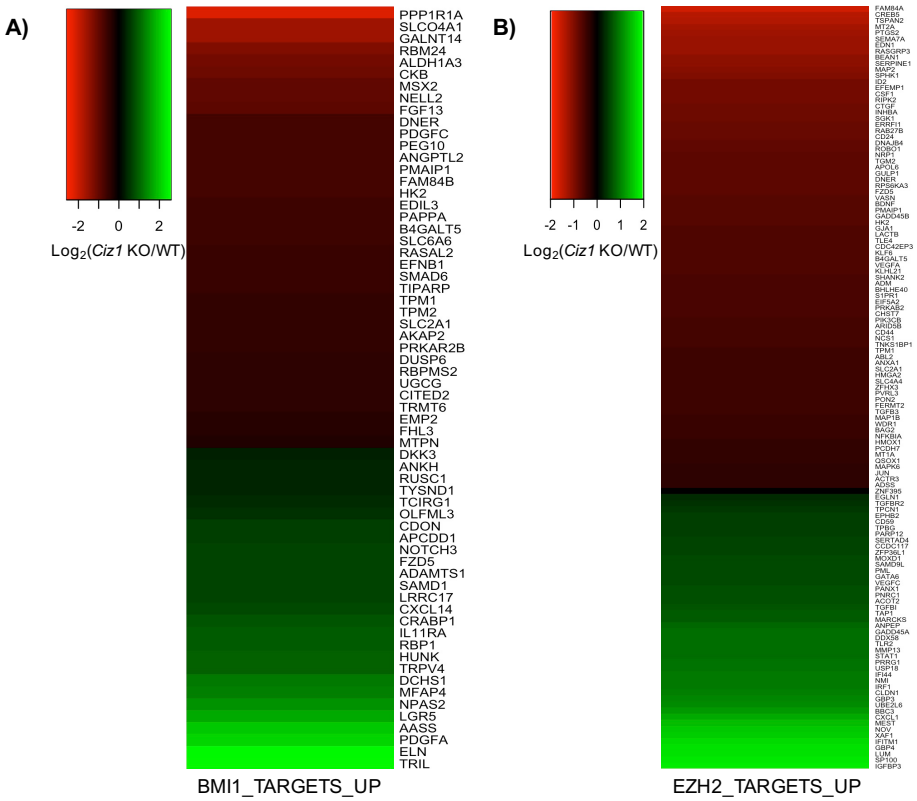


Figure 5.11. Heatmaps showing wild-type versus Ciz1 KO fold change of gene expression of genes *upregulated* upon knockdown of A) BMI1 or B) EZH2, that overlap with CIZ1-1480. Data has been obtained from the average of three replicates for both WT and Ciz1 KO cell lines. A red-green colour scale represents levels of fold change (red: downregulated; green: upregulated).

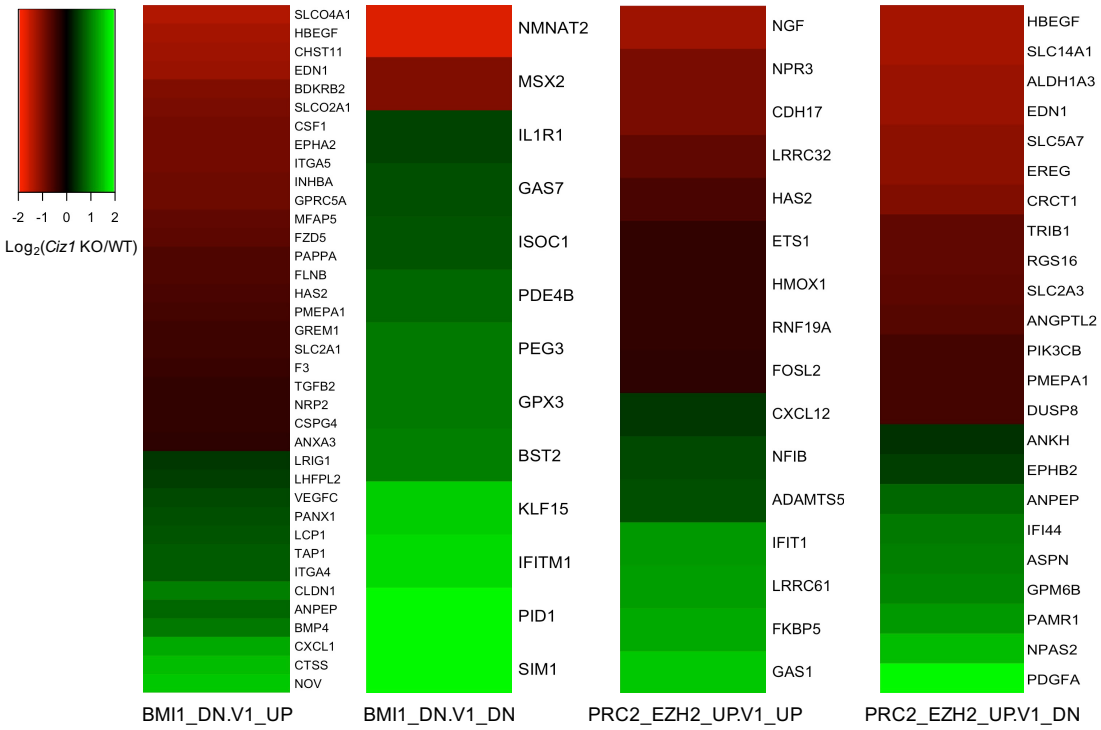


Figure 5.12. Heatmaps showing wild-type versus Ciz1 KO fold change of gene expression of genes that overlap between CIZ1-1480 and some of the gene sets returned by the oncogenic signatures analysis. From left to right: genes known to be *upregulated* upon knockdown of BMI1, genes known to be *downregulated* upon knockdown of BMI1, genes known to be *upregulated* upon knockdown of EZH2 and genes known to be *downregulated* upon knockdown of EZH2. Data has been obtained from the average of three replicates for both WT and Ciz1 KO cell lines. A red-green colour scale represents levels of fold change (red: downregulated; green: upregulated.

With so many target genes of PRC1 and PRC2 subunits affected by loss of CIZ1, we might expect to see significant changes in gene expression of the subunits themselves. Remarkably, gene expression levels of the PRC1 and PRC2 subunits is not affected as a consequence of *Ciz1* KO (Table 5.6).

PRC1/PRC2 genes				
Gene	**WT**	***Ciz1* KO**	**p-value**	**q-value**
BMI1	15.9396	14.2735	0.6403	0.999858
MEL18	14.9691	18.4002	0.24695	0.999858
EZH2	12.1375	12.0017	0.94195	0.999858
SUZ12	12.0546	11.9928	0.964	0.999858
EED	36.1997	34.9442	0.86835	0.999858

Table 5.7. Expression of genes encoding PRC1 and PRC2 subunits are not affected by knockout of *Ciz1*. Table shows mean gene expression in FPKM of genes encoding PRC1 subunits (BMI1, MEL18) and PRC2 subunits (EZH2, SUZ12, EED), in the wild-type female cells versus *Ciz1* KO female cells. p and q values indicate that levels of expression are not significantly different between the two conditions.

5.4.4.3 Genes affected by knockdown of BMI1 are differentially affected in female and male cell lines

Since analysis of gene expression from the X chromosome has not shown strong evidence that it could underlie the female gender-bias of the lymphoma-like phenotype observed in *Ciz1* KO mice, I investigated whether some of the autosomal gene targets of PRC1 and PRC2 are differentially affected by *Ciz1* KO in female versus male cell lines. Since BMI1 has established links with lymphoma (Raaphorst, 2005) I will look specifically at the genes within the BMI1 gene sets, more specifically those that are upregulated by knockdown of BMI1 (BMI1_DN.V1_UP) in the oncogenic signature analysis, which includes a total of 37 genes (Table 5.5). Analysis of the fold change in gene expression between *Ciz1* KO and WT cells for these 37 genes (Figure 5.13), reveals 18 that are differentially affected between female and male cells. This includes 15 genes with opposite patterns of expression, that is where a gene is upregulated in the female as a consequence of *Ciz1* KO but downregulated in the male, or vice versa. It also includes three genes where the change occurs in the same direction but the fold change is at least three times greater in one of them. It is important to take into account that this analysis only includes one replicate for each of the male samples and therefore no statistical analysis can be carried out. However, all three replicates for the female samples are internally consistent which

suggests the same might occur for the males. If any further work was to be derived from this analysis female versus male comparisons would need repeating with enough replicates to obtain statistical significance or validation of candidate genes by qPCR.

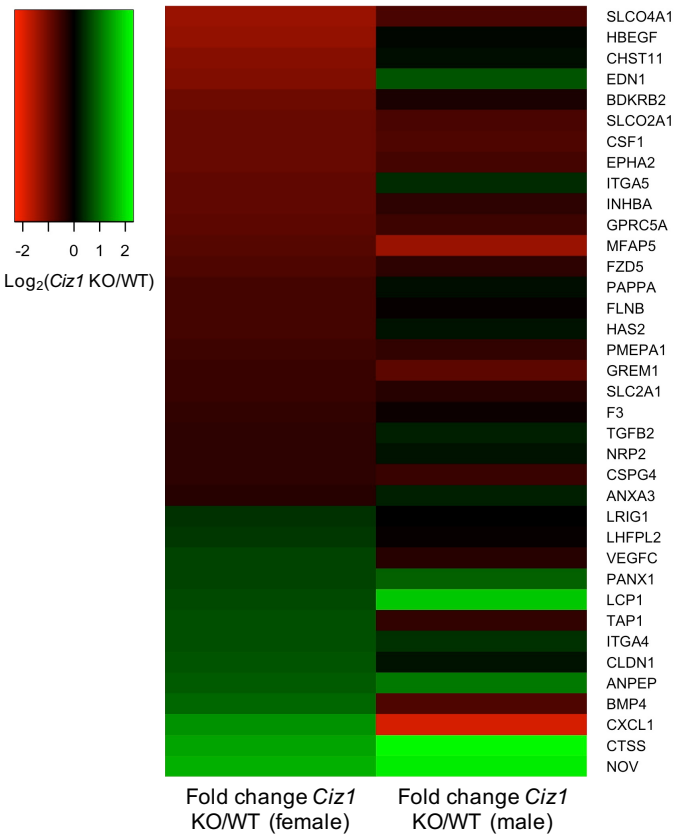


Figure 5.13. Analysis of female versus male differences following *Ciz1* knockout for genes known to be affected by knockdown of BMI1. Left, heatmap shows fold change between wild-type and *Ciz1* KO female cells. Right, heatmap shows fold change for the same genes between wild-type and *Ciz1* KO male cells. The genes shown are the 37 that overlap between the oncogenic signature gene set that are upregulated upon knockdown of BMI1 and CIZ1-1480. Data has been obtained from the average of three replicates for both WT and Ciz1 KO female cell lines and from one single replicate for the male data. A red-green colour scale, indicated by the colour key, represents levels of fold change (red: downregulated; green: upregulated).

5.4.4.4. Re-expression of Ciz1 in a Ciz1 KO background reinstates wild-type levels of gene expression for some genes

A *Ciz1* KO inducible cell line (13.17) exposed to 5μg/mL of doxycycline for a period of 48 hours was included in the RNA sequencing in order to test whether re-expression of *Ciz1* is capable of restoring wild-type levels of expression. Once again,

I will focus my analysis on the same small subset of genes, the 37 genes that overlap with the BMI1_DN.V1_UP gene set. As shown in Figure 5.14 reinstatement to wild-type levels occurs following re-expression of *Ciz1* for at least six example genes. In other cases, slight reinstatement occurs but fails to reach wild-type levels. It is possible that in these cases a longer exposure could completely reinstate wild-type levels. In addition, for some genes no reinstatement is observed or in fact addition of doxycycline causes levels of expression to drift even further from wild-type status (Figure 5.15). A total of 14 genes were discarded from the analysis since the control cell line, e13.15 KO exposed to dox, showed gene expression to be affected by the presence of doxycycline. From the remaining candidate genes 15 showed no reinstatement while 8 did show reinstatement of wild-type levels.

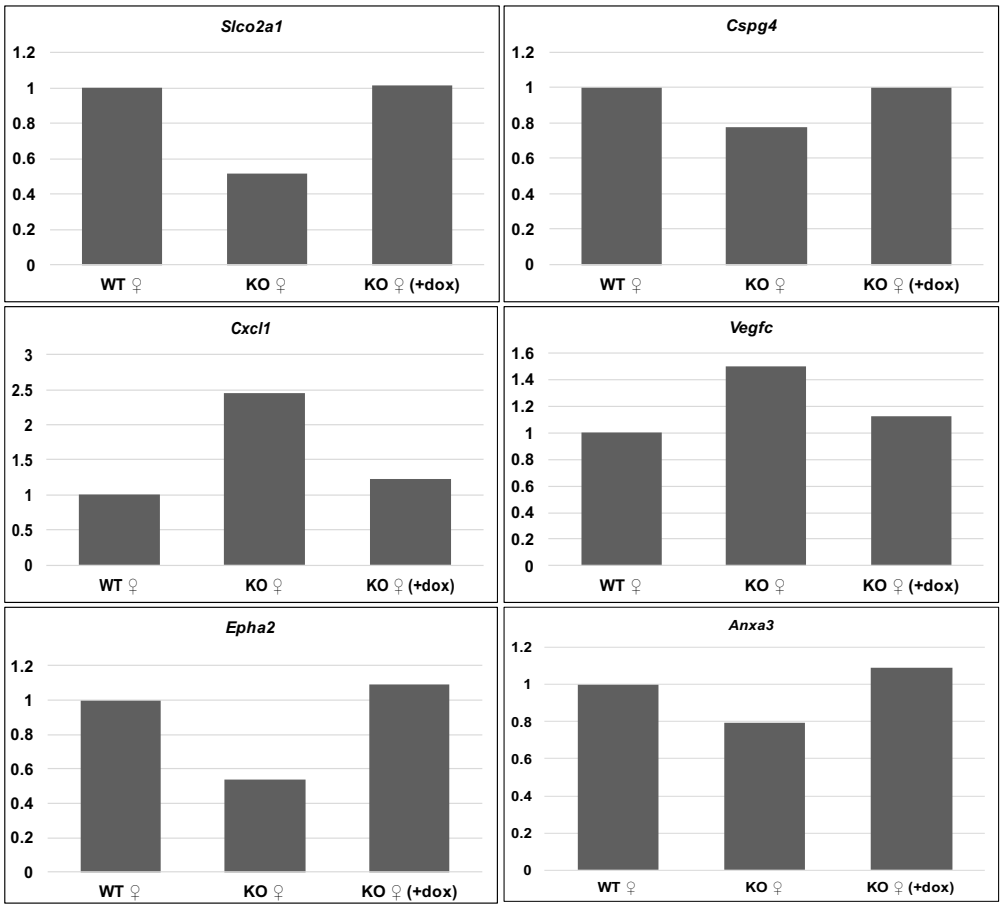


Figure 5.14. Histograms showing six example genes where re-expression of *Ciz1* in a *Ciz1* KO background reinstates wild-type levels of gene expression. The first bar in each histogram represents wild-type levels of expression, which have been normalised to 1. Levels of expression in the *Ciz1* KO cell line and the induced e13.17 *Ciz1* KO cell line, showed in the second and third bar, respectively, are expressed relative to wild-type levels. The WT and *Ciz1* KO bars represent the average of three replicates while the *Ciz1* KO (+dox) cell line represents data from one single replicate.

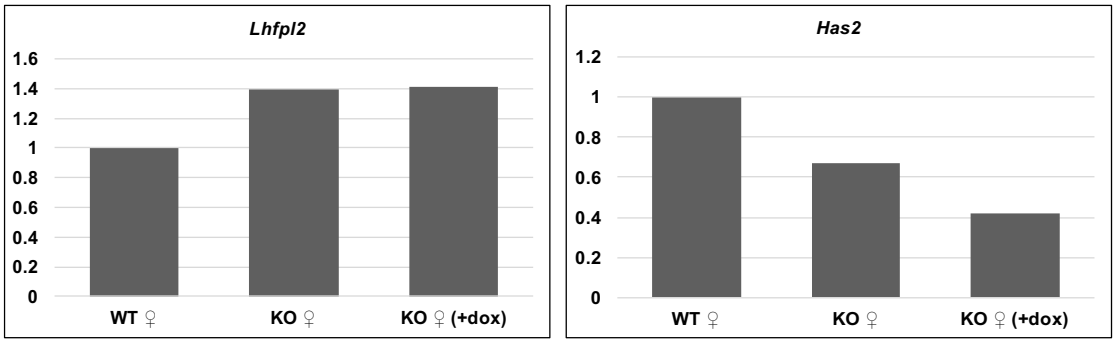


Figure 5.15. Example genes where re-expression of *Ciz1* in a *Ciz1* KO background fails to reinstate wild-type levels of expression. The first bar in each histogram represents wild-type levels of expression, which have been normalised to 1. Levels of expression in the *Ciz1* KO cell line and the induced e13.17 *Ciz1* KO cell line, showed in the second and third bar, respectively, are expressed relative to wild-type levels. The first histogram shows levels of expression of the *Lhfpl2* gene, an example where re-expression of *Ciz1* does not cause any changes in gene expression compared to the *Ciz1* KO cell line. The second histogram shows levels of expression of the *Has2* gene and constitutes an example where re-expression of *Ciz1* cause levels of expression to drift even further from wild-type levels. The WT and *Ciz1* KO bars represent the average of three replicates while the *Ciz1* KO (+dox) cell line represents data from one single replicate.

5.4.4.5 Genes affected by knockdown of BMI1 are good candidates for the female specific lymphoma

It is possible that the female-specific lymphoma-like phenotype could be linked to some of the genes that are known to be affected by knockdown of BMI1. Good candidate genes would likely be differentially affected in female and male cells lines, although these results must be interpreted with caution due to the absence of replicates. Genes which are capable of reaching wild-type levels of expression upon re-expression of Ciz1 could also be of interest since they are most likely those that are more directly affected by loss of the protein and therefore more dependent of it. Three genes from the list of 37 genes meet all of these criteria: *Cxcl1*, *Vegfc*, and *Anxa3* (Figure 5.16). All three genes have significant p-values in the wild-type versus *Ciz1* KO comparison and significant p and q values reported by the GSEA software make them of great interest. Cxcl1 and Vegfc are upregulated upon loss of *Ciz1*, with a fold change of 1.29 and 0.58 respectively. Anxa3 on the other hand is downregulated, with a fold change of -0.33.

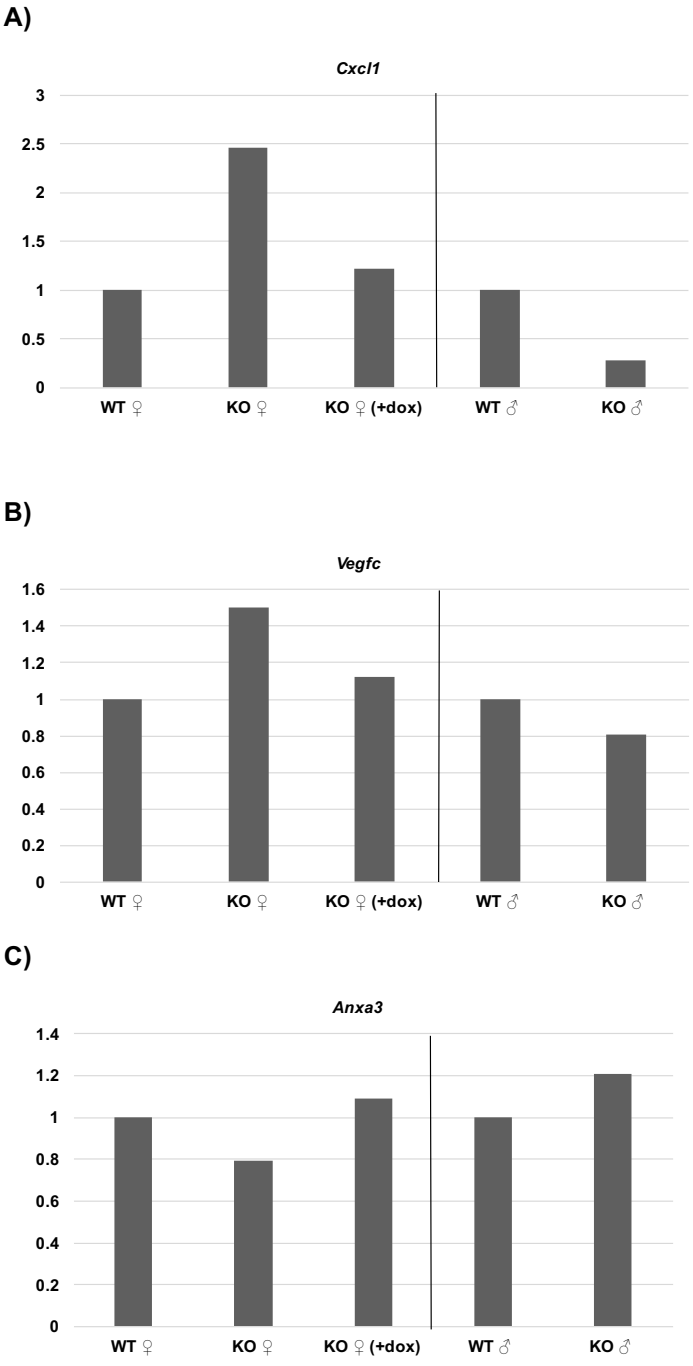


Figure 5.16. Histograms showing analysis of female versus male and *Ciz1* re-induction analysis for three genes that are candidate mediators of the lymphoma-like phenotype. For each gene the first bar represents female wild-type levels of expression, which have been normalised to 1. The next two bars represent levels of expression in *Ciz1* KO female cell lines and induced e13.17 *Ciz1* KO cell line, respectively, which have been expressed relative to the normalised wild-type level. On the right-hand side of the histograms the first bar shows male wild-type levels of expression, which have also been normalised to 1. The next bar represents levels of expression in *Ciz1* KO male cell line, expressed relative to the normalised male wild-type level. The female WT and female *Ciz1* KO bars represent the average of three replicates while all other bars represent data from one single replicate.

5.4.4.6 *Immune function related genes are also affected by knockout of Ciz1*

Since 100% of *Ciz1* KO female mice that were analysed in detail were found to have developed a lymphoma-like phenotype, we might expect to see changes in genes with functions linked to the immune system. In fact, when analysing the list of 1,480 genes that are differentially expressed between WT and *Ciz1* KO female cell lines using the GSEA Gene Ontology (GO) biological process gene set, a total of 25 genes were found to overlap with the immune system process gene set, which has a total of 332 genes (GO:0002376). This 7.5% representation returned a p value of 1.23E-08, and a q value of 1.39E-07, indicating biological significance. Since the lymphoma-like phenotype is specific to female mice we would also expect to see that some of these genes are differentially affected between female and male cells (Figure 5.17). In fact, a total of seven of these 25 immune-linked genes have opposite patterns of expression between female and male *Ciz1* KO cells, and one further gene (*Rsad2*), is of interest since, although upregulated in both, it is more than three times higher in the female.

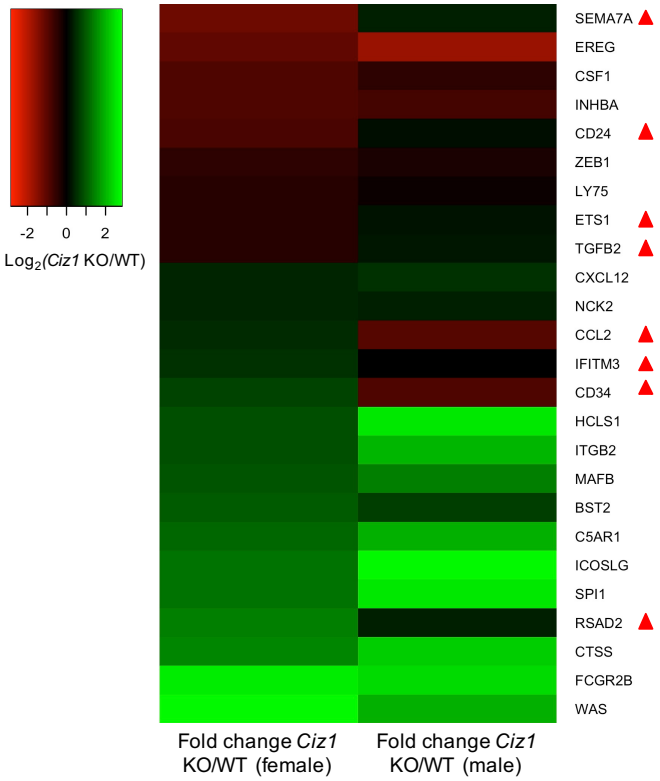


Figure 5.17. Analysis of female versus male differences in gene expression following *Ciz1* knockout for immune-linked genes. The genes shown on the heatmap are genes that overlap between CIZ1-1480 and the list of genes that are included in the immune system functions gene set as

defined by biological process gene set in the GSEA software. Data has been obtained from the average of three replicates for both WT and *Ciz1* KO female cell lines and from one single replicate for the male data. A red-green colour scale, indicated by the colour key, represents levels of fold change (red: downregulated; green: upregulated).

5.4.4.7 Summary section of other interesting results of gene ontology analysis

In addition to the results described above, the gene set enrichment analysis also returned other interesting results, which will possibly become subject of further research in the future. However, with the intention of keeping a written record I have gathered some of these results and compiled a table. Some of the hits of the GSEA software have not been included since they are not relevant to my investigation, e.g: genes consistently up-regulated in mammary stem cells both in mouse and human species.

The first result in Table 5.8. indicates that CIZ1-1480 is enriched for genes that are known to be upregulated in the brain of patients with Alzheimer's disease (p= 1.93E-40, q = 3.14E-38). This constitutes a striking result since a splice variant of Ciz1 has been reported to be upregulated in the brain of Alzheimer's patients (Dahmcke et al, 2008).

CIZ1-1480 is also enriched for genes that are known to be upregulated in breast cancer cells after knockdown of SATB1. This is of interest because SATB1 is a scaffold or matrix-associated region (S/MAR) binding protein which has been linked with chromatin organization and is thought to have a role in Xist-mediated silencing although its exact role is still unclear (Tattermusch and Brockdorff, 2011).

All other results in Table 5.8 show that CIZ1-1480 is also enriched for genes that are known to be either upregulated or downregulated in several types of cancer including urothelial carcinoma, breast cancer, prostate cancer, papillary thyroid carcinoma and pediatric adrenocortical tumours. These are just some of the examples of gene set enrichment results that link loss of CIZ1 to cancer. These observations are of great interest because CIZ1 has already been linked with cancers such as breast cancer (den Hollander et al, 2006), lung cancer (Higgins et al, 2012), hepatocellular carcinoma (Wu et al, 2015) and Ewing sarcoma (Rahman et al, 2007). These results

could shed some light on the CIZ1-related pathways that are being affected during the development of cancer.

CURATED GENE SETS			
Gene set name	**Description**	**Genes in overlap/ genes in gene set**	**Statistical test**
BLALOCK_ALZHEIMERS_DISEASE_UP	Genes up-regulated in brain from patients with Alzheimer's disease.	129/1691	p = 1.93E-40 q = 3.14E-38
HAN_SATB1_TARGETS_UP	Genes up-regulated in MDA-MB-231 cells (breast cancer) after knockdown of SATB1 [GeneID=6304] by RNAi.	53/395	p = 2.47E-28 q = 1.44E-26
LINDGREN_BLADDER_CANCER_CLUSTER_2B	Genes specifically up-regulated in Cluster IIb of urothelial cell carcinoma (UCC) tumors.	74/392	p = 7.09E-50 q = 2.09E-47
SMID_BREAST_CANCER_LUMINAL_B_DN	Genes down-regulated in the luminal B subtype of breast cancer.	83/564	p = 4.91E-47 q = 1.22E-44
LIU_PROSTATE_CANCER_DN	Genes down-regulated in prostate cancer samples.	76/481	p = 2.23E-45 q = 4.58E-43
DELYS_THYROID_CANCER_UP	Genes up-regulated in papillary thyroid carcinoma (PTC) compared to normal tissue.	71/443	p = 7.28E-43 q = 1.32E-40
WEST_ADRENOCORTICAL_TUMOR_DN	Down-regulated genes in pediatric adrenocortical tumors (ACT) compared to the normal tissue.	75/546	p = 2.04E-40 q = 3.21E-38

Table 5.8. Examples of other results of interest obtained from the gene set enrichment analysis on genes that are significantly affected (p<0.05) by knockout of *Ciz1.*

5.5 Discussion

One of the aims of this work was to investigate, whether the changes that have been observed at the inactive X in the absence of CIZ1, such as delocalization of Xist and disruption of the H3K27 methylation mark, impact on X-linked gene expression. More specifically, whether there is evidence for gene reactivation from the inactive X. Results show that gene expression remains largely unchanged, which is in line with similar analyses of X-linked transcript reactivation following disruption of known players of X inactivation in somatic cells. Examples are derived from disruption of genes within the X inactivation centre (Brown and Willard, 1994) or nuclear matrix proteins with important functions in X inactivation such as *Saf-A* (Hasegawa et al, 2010) and *Firre* (Yang et al, 2015), knockout of which leads to disruption of the H3K27me3 but not reactivation of the inactive X. The small number of significant changes (%2 of transcription units) in gene expression that were observed following *Ciz1* knockout seem to be evenly distributed over the X chromosome and, given the design of this experiment, it is not possible to specifically identify X-linked expression from Xi, therefore strong conclusions about the 28 upregulated genes in relation to Xi

reactivation cannot be made. Furthermore, none of the genes at the XIC with important functions in the initiation of X chromosome inactivation, such as Xist (Plath et al, 2002), Tsix (Lee et al,1999), *Jpx* (Tian et al, 2010), *Ftx* (Chureau et al, 2011) are affected by knockout of *Ciz1*. This indicates that, if indeed CIZ1 has a role in X chromosome inactivation, it is likely not in the initiation stage of XCI.
Overall it is clear that widespread gene reactivation does not occur when CIZ1 is removed, nevertheless Ciz1 could still have a function in the maintenance of X inactivation, as the multiple layers of control are unlikely to be breached by loss of one factor. The results are consistent with the previous observation that *Ciz1* KO embryos develop normally (Ainscough, unpublished), and contrasts with the lethal phenotype of *Xist* knockout (Marahrens et al, 1997), which is known to be essential for initiation of X inactivation.

The changes that do occur could be secondary effects of changes elsewhere in the genome, or genes directly regulated by CIZ1 but not via XCI, and seem to occur with similar frequency to the rest of the genome. The data show that 2.6% of genes across the whole genome are significantly changed following *Ciz1* KO. Although these changes initially seemed to be unlinked, gene set enrichment analysis showed that many are targets of PRC1 and PRC2 subunits, or genes that are affected by knockdown of some of these subunits. This is an interesting observation since we also have evidence that the laying down of the H3K27me3 mark, for which PRC2 is responsible, is also disrupted in the absence of *Ciz1*. These observations begin to suggest that CIZ1 is involved in the normal function of the PRC1 and PRC2 complexes. Interestingly, the RNA sequencing results revealed that expression of the PRC1 and PRC2 complexes subunits themselves is not affected by loss of *Ciz1*. This raises a whole new set of questions about the exact function of CIZ1 in relation to the action of polycomb repressive complexes. Further work should try to confirm observations which suggest that PRC1 and PRC2 complexes are functioning normally in the absence of *Ciz1*, using protein expression and immunolocalization studies in the first instance, since transcript levels do not always correlate with protein. In addition, although we have evidence that loss of *Ciz1* disrupts the H3K27me3 mark, no work has been done on the epigenetic mark laid down by the PRC1 complex. It is therefore of interest to test whether ubiquitylation of H2AK119 (Chen et al, 2004) is in any way disrupted.

The lymphoma-like phenotype observed in female *Ciz1* knockout mice could be linked to the observation that PRC1 and PRC2 targets are affected by loss of *Ciz1*. Female mice with *Ciz1* knockout have enlarged spleens, liver and lungs (J. Ainscough, unpublished). Pathological characterisation (J. Roulsen, pathologist, unpublished data) of the lymphoid abnormalities led to description of the phenotype as consistent with follicular-type B cell lymphoma with reactive T cell infiltration. It is therefore of interest to investigate whether any of the affected PRC1 or PRC2 target genes could be linked to this phenotype. Of particular interest are the genes that are affected by knockdown of BMI1, since this PRC1 subunit and more specifically its overexpression, has been linked with lymphoma (Raaphorst, 2005). Although no overexpression of BMI1 itself is observed in our model, the majority of genes that are affected by knockdown of BMI1 show a downregulation of gene expression rather than the upregulation observed following knockout of BMI1 (Wiederschain et al, 2007) which, as previously mentioned, might suggest the gene expression observed following knockout of Ciz1 might be more characteristic of an overexpression of BMI1.

In order to understand what drives development of the lymphoma-like phenotype in *Ciz1* KO mice, analysis of the strong gender bias is likely to provide clues. Despite the fact that the lymphoma phenotype is observed exclusively in female mice, and as described above, this does not seem to be a consequence of compromised X chromosome inactivation. Nonetheless, there could be other intrinsic differences between female and male mice that could explain a female-only phenotype such as hormonal differences, for example. For this reason, BMI1 linked genes affected by *Ciz1* KO were investigated for differences between female and male cell lines. This analysis revealed that in fact, approximately half of the 37 analysed genes are differentially affected by loss of *Ciz1* in female cell lines compared to male. This is not a surprising result since even for humans significant differences in gene expression patterns have been reported for women and men (Zhang et al, 2009).

From the total of 37 genes three candidates were identified, as they showed differences between female and male and also showed restoration of wild-type levels of expression following re-expression of *Ciz1*. These genes are *Vegfc, Cxcl1 and Anxa3*.

Vascular endothelial growth factor C (*Vegfc*) is part of the VEGF family, which signal through receptor tyrosine kinases to induce angiogenesis, and lymphangiogenesis

(Ho and Kuo, 2007). In addition to playing a role in normal development, VEGFs also play crucial roles in the development and progression of cancer. Defined as one of the cancer hallmarks (Weinberg and Hanahan, 2000), tumours require a constant supply of oxygen and nutrients, and in order to ensure their supply they are capable of activating an "angiogenic switch". This leads to an increase in the presence of inducers of angiogenesis such as VEGF, while reducing presence of angiogenic inhibitors. In line with this observation, many solid tumours as well as haematological malignancies, including several types of lymphoma and leukemia, are characterised by overexpression of VEGF (Ho and Kuo, 2007) (Gerber et al, 2003). Expression of this gene has also been found in malignant lymphoma, correlating with increased lymphangiogenesis (Kadowaki et al, 2005). Interestingly, while expression of *Vegfc* is upregulated in the *Ciz1* KO female cell line, with increased lymphangiogenesis observed in the spleen of female *Ciz1* KO mice. The same gene is downregulated in the male cell line, and might therefore contribute to the gender-specific phenotype.

Cxcl1 encodes the chemokine ligand 1 protein, which belongs to the CXC chemokine family. It is expressed by macrophages, neutrophils and epithelial cells and it plays a role in inflammation, mainly through chemoattraction of neutrophils (De Filippo et al, 2013). There is also evidence that secretion of CXCL1 from epithelial and endothelial cells induces angiogenesis. Just like the *Vegfc* gene, *Cxcl1* is also upregulated in the female cell line but downregulated in the male. Increased expression of *Cxcl1* has also been reported in follicular lymphoma (Glas et al, 2007), coinciding with the pathologist report on the type of lymphoma observed in the *Ciz1* KO female mice.

Finally, *Anxa3* encodes the annexin 3 (ANXA3) protein, and belongs to the annexin family, which plays a role in cellular growth. No direct links have been established between this gene and lymphoma. It has however been linked with tumour growth related functions, specifically in hepatocellular carcinoma. However, given the fact that *Anxa3* is downregulated in female cells following knockout of *Ciz1* rather than upregulated which would be of interest, this gene is of no further interest for the purpose of this thesis.

In addition to the BMI1–linked genes analysis, the immune system gene ontology analysis provided additional candidate genes which might contribute to the lymphoma-like phenotype. A good example is the *Cd34* gene, which is upregulated in the female cell lines and downregulated in the male cells. *Cd34* is a cell surface

glycoprotein with a role in a cell-cell adhesion and is expressed on the surface of lymph node high endothelial venules (van Zante et al, 2003). It plays an essential role in the infiltration of T-cells into lymph nodes, which has been observed in the mouse *Ciz1* KO phenotype. The CD34 protein binds to the lymphocyte receptor L-selectin, expressed on the surface of T cells, allowing tethering and rolling along the endothelial wall and into the lymph node (Girard et al, 2012). Although no causative effects can be concluded from this gene expression analysis, these observations suggest that within the genes that are affected by knockout of *Ciz1*, there are some with known functions that could underlie the lymphoma-like phenotype and that in addition, show different expression patterns in female versus male cell lines. To understand cause and effect the candidate genes would need to be validated in animal models.

The RNA sequencing results have provided large amounts of information on the effects that loss of Ciz1 has on gene expression, and I have selected a subset to investigate here. Analysis of all of the results is beyond the scope of this thesis, however, I have found some very promising avenues, which could become the subject of further research. This data could help understand the specific biological functions of CIZ1 both in a wild-type scenario such as cell-cycle and possibly polycomb repressive complexes related functions, as well as in the context of pathological conditions such as Alzheimer's disease or cancer.

CHAPTER 6. DISCUSSION

The nuclear matrix protein CIZ1 has been linked with devastating diseases such as several types of cancer, including breast (den Hollander et al, 2006) and lung cancer (Higgins et al., 2012), as well as neurological conditions such as Alzheimer's disease (Dahmcke et al, 2008) and cervical dystonia (Xiao et al, 2012). It is therefore of importance to understand the biological functions of CIZ1, in addition to the already known role in the cell cycle (Coverley et al, 2005). By understanding the normal functions of the protein in different biological pathways we might obtain information on how disruption of the protein is linked with the development of disease.

Previous research has shown that CIZ1 is localized to the Xi and that in its absence, the long non-coding RNA, *Xist*, which is essential for initiation of X chromosome inactivation, becomes delocalised from the inactive X (Ridings-Figueroa, unpublished). The exact function that CIZ1 exerts at this site is unknown and therefore became the subject of my research. Investigating the function of CIZ1 in the context of X chromosome inactivation is scientifically very interesting, not just to the X chromosome inactivation field, but also more widely as X chromosome inactivation has long been used as a model for the study of chromatin remodeling and epigenetic silencing, processes which are essential for cell survival and are known to become disrupted in the context of disease (Moss and Wallrath, 2007). In addition, the importance of nuclear matrix proteins in X inactivation has also become a matter of interest in the field (Hasegawa et al, 2010) (Hasegawa et al, 2011). This research could therefore add CIZ1 to the list of nuclear matrix proteins with known functions in X inactivation.

Results from my investigation suggest that CIZ1 might have important links with the functions of PRC1 and PRC2, complexes which are central to the epigenetic silencing of the inactive X chromosome (Lund et al, 2004). This is further supported by observations that in the absence of CIZ1 the H3K27me3, laid down by the PRC2 complex, is disrupted, at least to a certain extent. It is however unclear how exactly this disruption occurs. Other studies suggest that maintenance of this epigenetic mark is coincident with movement of the inactive X chromosome to the perinucleolar ring during S-phase, and that the H3K27me3 is deposited on the newly synthesized DNA (Zhang et al, 2007). In the absence of Xist, movement to the perinucleolar ring fails to occur (Zhang et al, 2007). Initial studies suggested direct binding of Xist to PRCs

(Zhao et al, 2009) but more recent work argue against by showing spatial separation between Xist and polycomb repressive complexes (Cerase et al, 2014). In a similar way, knockdown of the long non-coding RNA *Firre,* results in disruption of perinucleolar localization of the Xi as well as disruption of the H3K27me mark (Yang et al, 2015).

Strikingly, evidence collected in the Coverley labs shows that movement of the Xi to the perinucleolar ring during S-phase is also dependent on the presence of CIZ1 (Scott, unpublished). By putting these observations together, it is possible to bring forward a hypothesis for the function of CIZ1 in this process. If one assumed that the PRC2 factories, which are responsible for the laying down of the H3K27me3 mark, are located at the perinucleolar ring then it could be possible that CIZ1, likely in conjunction with other players such as Xist or other nuclear matrix proteins, could mediate the movement of the inactive X chromosome to the site, and that, in its absence, the 'transport chain' becomes disrupted failing to deliver template to the PRC2 factory. Whether CIZ1 constitutes an adaptor or linker molecule between the Xi and this transport chain, which could possibly be the nuclear matrix itself, or whether it is part of the transport chain itself remains unknown. In addition, the type of interaction, direct or indirect, that exists between CIZ1 and the Xi is also unknown.

The proposed hypothesis is supported by my observation that although many of the PRC1 and PRC2 target genes are affected by loss of CIZ1, the PRC1 and PRC2 subunits themselves do not show any differences in gene expression in the absence of Ciz1. Future work would need to extend this analysis from transcript to protein, and ask whether PRC components, including those that lay down H3K27me3 (PRC2) and H2AK119ub1 (PRC1) are present at similar levels and similar locations in WT and *Ciz1* KO cells. If so, failure of template to visit those locations would be a viable hypothesis.

If this proposition is correct, one could expect that this function of CIZ1 in relation to polycomb repressive complexes might extend not just to the process of X inactivation but also to the rest of the genome, as suggested by my whole genome analysis. This could offer an explanation for the overall disruption of the H3K27me3 mark across the whole nucleus in one of the subject cell types in this thesis. It is however likely that in the absence of CIZ1 other players, possibly other nuclear matrix proteins, are

able to compensate or at least partially substitute CIZ1 in its absence, since both the *Ciz1* KO embryo and *Ciz1* KO derived cell lines are able to survive.

It is also possible that different mechanisms could be at work in the embryo compared to somatic cells, as well as between cell types. In the absence of CIZ1 100% of female mice develop a lymphoma-like phenotype, similar to B-cell follicular lymphoma, suggesting that cell types involved in this condition could be more susceptible to loss of CIZ1 compared to others. In fact, polycomb genes have been reported to be essential for B-cell development and differentiation and are suggested to have important roles in the initiation and progression of lymphoma (Wang et al, 2015).

Future experiments should aim to validate the observation that loss of CIZ1 affects levels of expression of PRC1 and PRC2 target genes. This could be done through qPCR and ideally should also be validated at the protein level through a Western blot. In addition, it would be of interest to use antibodies against different components of the PRC1 and PRC2 complexes and test whether they colocalise with CIZ1 at different stages of the cell cycle or at the perinucleolar ring and whether normal localization of the PRC components is in any way disrupted in the absence of CIZ1. Protein-protein interaction assays could also be performed to test for interactions between CIZ1 and PRC components. Furthermore, I would also test whether loss of CIZ1 does in fact only affect the H3K27me3 mark laid down by PRC2 or whether it also disrupts the H2AK119 ubiquitination mark laid down by PRC1. If the importance of CIZ1 in the normal functioning of PRC components was validated it would be important to then ask whether the numerous links that CIZ1 has with several types of cancer are underlined by this function. A more detailed analysis of the RNA sequencing results would be useful in this matter.

In conclusion, my analysis has opened up an exciting line of research which could shed new light on the molecular function of CIZ1, and with it, the mechanisms that are deregulated in its absence and which result in development of malignant conditions.

LIST OF ABBREVIATIONS

Abbreviation	Term
XCI	X chromosome inactivation
PRC	Polycomb repressive complexes
lncRNA	Long non-coding RNA
RNAseq	RNA sequencing
Xic	X inactivation centre
Xp	Paternal X chromosome
Xm	Maternal X chromosome
rXCI	Random X chromosome inactivation
Xi	Inactive X chromosome
mPEFs	Mouse primary embryonic fibroblasts
GSEA	Gene set enrichment analysis software
H3K27me3	Trimethylated lysine residue at position 27 in the protein histone H3

REFERENCES

Ahler, E. et al. (2013). Doxycycline Alters Metabolism and Proliferation of Human Cell Lines. Plos *8* (5), 1-7.

Ainscough, J.F., Rahman, F.A., Sercombe, H., Sedo, A., Gerlach, B., and Coverley, D. (2007). C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. Journal of cell science *120*, 115-124.

Alberts et al. (2008). Molecular Biology of the Cell 4th Ed. New York, Garland Science.

Anachkova, B., Djeliova, V., and Russev, G. (2005). Nuclear matrix support of DNA replication. Journal of cellular biochemistry *96*, 951-961.

Anderson, C.L. and Brown, C.J. (2002). Variability of X chromosome inactivation: effect on levels of TIMP1 RNA and role of DNA methylation. Human Genetics *110* (3), 271-8.

Augui, S., Filion, G.J., Huart, S., Nora, E., Guggiari, M., Maresca, M., Stewart, A.F., and Heard, E. (2007). Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. Science *318*, 1632-1636.

Bao, H. (2008). RNA profiling in an Alzheimer's disease mouse model. The University of Texas, ProQuest.

Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Molecular cell *53*, 965-978.

Belgrader, P., Dey, R. and Berezney, R. (1991). Molecular cloning of matrin 3. A 125-kilodalton protein of the nuclear matrix contains an extensive acidic domain. J Biol. Chem. *266*, 9893-9899.

Ben-Porath, I. et al. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nature Genetics *40* (5), 499-507.

Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. Human genetics *130*, 237-245.

Berletch, J.B. et al (2015). Escape from X inactivation varies in mouse tissues. Plos Genetics *11* (3), e1005079.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y*., et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Bracken, AP., Dietrich, N., Pasini, D., Hansen, KH., Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20* (9), 1123-1136.

Brinton, R.D. (2004). Impact of estrogen therapy on Alzheimer's disease: a fork in the road?. CNS Drugs *18* (7), 405-422.

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154-156.

Calabrese, J.M. et al. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell *151* (5), 951-963.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-404.

Cerase A., Smeets D., Tang YA., Gdula M., Kraus F., Spivakov M., et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc Natl Acad Sci U S A. *111*, 2235–2240.

Cerase, A., Pintacuda, G., Tattermusch, A., and Avner, P. (2015). Xist localization and function: new insights from multiple levels. Genome biology *16*, 166.

Chen et al., (2004). Emerging roles of polycomb silencing in X-inactivation and stem cells maintenance. Cold Spring Harb Symp Quant Biol. *69*, 319-326.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404-416.

Chureau, C. et al. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Human Molecular Genetics *20* (4), 705-718.

Copeland, N.A., Sercombe, H.E., Ainscough, J.F., and Coverley, D. (2010). Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. Journal of Cell Science *123*, 1108-1115.

Coverley, D., Marr, J., and Ainscough, J. (2005). Ciz1 promotes mammalian DNA replication. Journal of Cell Science *118*, 101-112.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nature Genetics *22*, 323-324.

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*., et al.* (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Molecular Cell *53*, 301-316.

Dahmcke, C., Büchmann-Møller, S., Jensen, N., Mitchelmore, C. (2008). Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution

and is associated with Alzheimer's disease. Molecular and Cellular Neuroscience *38* (4), 589–594.

De Filippo, K. et al. (2013). Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. Blood *121* (24), 4930-4937.

Den Hollander, P., Rayala, S.K., Coverley, D., and Kumar, R. (2006). Ciz1, a Novel DNA-binding coactivator of the estrogen receptor alpha, confers hypersensitivity to estrogen action. Cancer Research *66*, 11021-11029.

Dobin, A. et al, (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, *29* (1).

Douglas D., Hsu JH., Hung L., Cooper A., Abdueva D., van Doorninck J., et al. (2008). BMI-1 promotes ewing sarcoma tumorigenicity independent of CDKN2A repression. Cancer Research *68*, 6507–6515.

Fairweather, D., Frisancho-Kiss, S. and Rose, N. (2008). Sex differences in autoimmune disease from a pathological perspective. American Journal of Pathology *173* (3), 600-609.

Fawcett D.W. (1966). An atlas of fine structure: the cell, its organelles and inclusions. W.B. Saunders Co.

Finn, E.H. (2014). Maternal bias and escape from X chromosome imprinting in the midgestation mouse placenta. Developmental Biology *390* (1), 80-92.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Current Opinion in Genetics & Development *31*, 57-66.

Gendrel, AV., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental Cell *23* (2),265-279.

Gendrel, A.V. et al. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and Cell Biology *33* (16), 3150-3165.

Gerber, H.P. and Ferrara, N. (2003). The role of VEGF in normal and neoplastic hematopoiesis. Journal of Molecular Medicine *81* (1), 20-31.

Girard, J., Moussion, C., Forster, R. (2012). HEVs, lymphatics and homeostatic immune cells trafficking in lymph nodes. Nature Reviews *12*, 762-773.

Glas, A.M. et al. (2007). Gene-Expression and Immunohistochemical Study of Specific T-Cell Subsets and Accessory Cell Types in the Transformation and Prognosis of Follicular Lymphoma. Journal of Clinical Oncology *25* (4), 390-398.

Greaves, E.A., Copeland, N.A., Coverley, D., and Ainscough, J.F. (2012). Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. Journal of Cell Science *125*, 2466-2477.

Gribnau, J., and Grootegoed, J.A. (2012). Origin and evolution of X chromosome inactivation. Current Opinion in Cell Biology *24*, 397-404.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. Developmental Cell *19*, 469-476.

Hasegawa, Y. and Nakagawa, S. (2011). Revisiting the function of nuclear scaffold/matrix binding proteins in X chromosome inactivation. RNA Biology *8* (5), 735-9.

Helbig, R., and Fackelmayer, F.O. (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, K.M., Watson, I.J., Blackhall, F.H., Rom, W.N., Pass, H.I., Ainscough, J.F., and Coverley, D. (2012). Variant Ciz1 is a circulating biomarker for early-stage lung cancer. PNAS USA *109*, E3128-3135.

Ho, Q. and Kuo, C. (2007). Vascular Endothelial Growth Factor: Biology and Therapeutic Applications. International Journal of Biochemistry and Cell Biology *39* (7-8), 1349–1357.

Holers, V.M. (2008). The spectrum of complement alternative pathway-mediated diseases. Immunological Reviews *223*, 300-316.

Jackson, D.A., and Cook, P.R. (1988). Visualization of a filamentous nucleoskeleton with a 23 nm axial repeat. The EMBO journal *7*, 3667-3677.

Jiang, J., Lau, A.C., and Csankovszki, G. (2014). Pluripotent cells will not dosage compensate. Worm *3*, e29051.

Jones, P.A. and Baylin, S.B. (2011). A decade of exploring the cancer epigenome-biological and translational implications. Nature *11*, 726-734.

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999-1011.

Kadowaki et al. (2005). Accelerated lymphangiogenesis in malignant lymphoma: possible role of VEGF-A and VEGF-C. British Journal of Haematology *130*(6), 869-77.

Koehler, D.R., and Hanawalt, P.C. (1996). Recruitment of damaged DNA to the nuclear matrix in hamster cells following ultraviolet irradiation. Nucleic Acids Research *24*, 2877-2884.

Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nature genetics *21*, 400-404.

Lee, J.T. (2005). Regulation of X-chromosome counting by Tsix and Xite sequences. Science *309*, 768-771.

Lee, J.T. et al. (2014). ATRX directs bindingof PRC2 to Xist RNA and Polycomb targets. Cell *159*, 869-883.

Lesher, A.M., Nilsson, B. and Song, W.C. (2013). Properdin in complement activation and tissue injury. Molecular Immunology *56* (3), 191-198.

Li, S.M. , Valo, Z., Wang, J., Gao, H., Bowers, C.W., Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. Plos One *7* (2), e31751.

Lund, A.H. and Van Lohuizen, M. (2004). Polycomb complexes and silencing mechanisms. Current Opininion in Cell Biology *16* (3), 239-46.

Lyon, M.F. (1971). Possible mechanisms of X chromosome inactivation. Nature New Biology *232*, 229–232.

Lyon, M. F. (1972). X-chromosome inactivation and developmental patterns in mammals. Biol. Rev. Camb. Philos. Soc. *47*, 1 -35.

Lyon, M.F. (2000). LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proceedings of the National Academy of Sciences of the United States of America *97*, 6248-6249.

Lyon, M.F. (2003). The Lyon and the LINE hypothesis. Seminars in Cell & Developmental Biology *14*, 313-318.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes & Development *11* (2), 156-166.

Margueron, R. and Reinberg, D. (2011). The Polycomb Complex PRC2 and its mark in life. Nature *469*, 343-349.

Maves, K.K. and Weiler, J.M. (1992). Detection of properdin mRNA in human peripheral blood monocytes and spleen. Journal of Laboratory and Clinical Medicine *120* (5), 762-6.

Michelitsch, M. D., and Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. PNAS *97*, 11910-5.

Mirkovitch, J., Mirault, M.E., and Laemmli, U.K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell *39*, 223-232.

Mitsui, K., Matsumoto, A., Ohtsuka, S., Ohtsubo, M., and Yoshimura, A. (1999). Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. Biochemical and Biophysical Research Communications *264*, 457-464.

Mootha et al, 2003. PGC-1-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics *34*, 267-273.

Moreira de Mello, J.C., de Araujo, E.S., Stabellini, R., Fraga, A.M., de Souza, J.E., Sumita, D.R., Camargo, A.A., and Pereira, L.V. (2010). Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. PLOS *5*, e10947.

Moss, T.J. and Wallrath, L.L. (2007). Connections between epigenetic gene silencing and human disease. Mutation Research *618* (1-2), 163-74.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693-1695.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer.

Nuytten, M. et al. (2008). The transcriptional repressor NIPP1 is an essential player in EZH2-mediated gene silencing. Oncogene *27* (10), 1449-60.

Pinter, S.F. et al (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Research *22* (10), 1864-76.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annual Review of Genetics *36*, 233-78.

Pollex, T., and Heard, E. (2012). Recent advances in X-chromosome inactivation research. Current Opinion in Cell Biology *24*, 825-832.

Raaphorst, F.M. (2005). Deregulated expression of polycomb-group oncogenes in human malignant lymphomas and epithelial tumors. Human Molecular Genetics *14*, R93-100.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rahman, F., Ainscough, J.F., Copeland, N., and Coverley, D. (2007). Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. Human Mutation *28*, 993-1004.

Roberts et al, (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology, *12* (3).

Robinson, S.I., Small, D., Idzerda, R., McKnight, G.S., and Vogelstein, B. (1983). The association of transcriptionally active genes with the nuclear matrix of the chicken oviduct. Nucleic Acids Research *11*, 5113-5130.

Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U., and Richter, A. (1992). Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. The EMBO Journal *11*, 3431-3440.

Sado, T. and Brockdorff, N. (2012). Advances in understanding chromosome silencing by the long non-coding RNA Xist. Philos Trans R Soc Lond B Biol Sci *5,* 368(1609).

Sato et al. (2013). PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. Scientific Reports *3*, 1911.

Schmittgen, T.D., Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nature Protocols *3*,1101–1108.

Schwartz, Y.B., and Pirrotta, V. (2014). Ruled by ubiquitylation: a new order for polycomb recruitment. Cell Reports *8*, 321-325.

Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., Zhu, X., Jiao, B., Hall, L.L., Green, M.R.*,* et al*.* (2010). Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature *467*, 977-981.

Sirchia, S.M., Ramoscelli, L., Grati, F.R., Barbera, F., Coradini, D., Rossella, F., Porta, G., Lesma, E., Ruggeri, A., Radice, P*.,* et al. (2005). Loss of the inactive X chromosome and replication of the active X in BRCA1-defective and wild-type breast cancer cells. Cancer Research *65*, 2139-2146.

Starmer, J., and Magnuson, T. (2009). A new model for random X chromosome inactivation. Development *136*, 1-10.

Stavropoulos, N., Rowntree, R.K., and Lee, J.T. (2005). Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. Molecular and Cellular Biology *25*, 2757-2769.

Subramanian et al, (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS, *102* (43), 15545–15550.

Tattermusch, A. and Brockdorff. N. (2011). A scaffold for X chromosome inactivation. Human Genetics *130*, 247-253.

Tian, D., Sun, S and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143* (3), 390-403.

Trapnell et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, *28* (5), 511-515.

Trapnell et al. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology, *31* (1).

Van Zante, A. et al. (2003). Lymphocyte–HEV Interactions in Lymph Nodes of a Sulfotransferase-deficient Mouse. JEM *198* (9), 1289-1300.

Viña, J. and Lloret, A. (2010). Why women have more Alzheimer's disease than men: gender and mitochondrial toxicity of amyloid-beta peptide. Journal of Alzheimer's Disease *20*, S527-33.

Wang, G.G., Konze, K.D. and Tao, J. (2015). Polycomb genes, miRNA, and their deregulation in B-cell malignancies. Blood *125* (8), 1217-25.

Warder, D. E. and Keherly, M. J. (2003). Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0-2)YYAC. Journal of Biomedical Science *10*, 406-417.

Warnes, G.R. et al. (2016). gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. https://CRAN.R-project.org/package=gplots.

Weinberg, R.A. and Hanahan, D. (2000). The Hallmarks of Cancer. Cell *100* (1), 57-70.

Werner et al. (2013). A critical role for the cholesterol-associated proteolipids PLP and M6B in myelination of the central nervous system. Glia *61* (4), 568-586.

Wiederschain, D. et al. (2007). Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. Molecular Cell Biology *27* (13), 4968-79.

Wu, J., Lei, L., Gu, D., Liu, H., and Wang, S. (2015). CIZ1 is upregulated in hepatocellular carcinoma and promotes the growth and migration of the cancer cells. Tumour biology *37*(4), 4735-4742.

Xiao, J. (2012). Mutations in CIZ1 cause adult-onset primary cervical dystonia. Annals of Neurology *71* (4), 458–469.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Research *20*, 614-622.

Yang et al (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biology *12*, 1-17.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. Cell *152*, 727-742.

Yoo et al, (2012). EZH2 methyltransferase and H3K27 methylation in breast cancer. International Journal of Biological Sciences *8*, 59-65.

Zeitlin, S., Parent, A., Silverstein, S., and Efstratiadis, A. (1987). Pre-mRNA splicing and the nuclear matrix. Molecular and Cellular Biology *7*, 111-120.

Zhang, L.F., Huynh, K.D., and 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.

Zhang, W., Huang, R.S., Duan, S., Dolan, M.E. (2009). Gene set enrichment analyses revealed differences in gene expression patterns between males and females. In Silico Biology *9* (3), 55-63.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. and Lee, T.J. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322* (5902), 750-6.