Functional Analysis of Aberrant CIZ1 Forms in Cancer

Gabrielle Lea Turvey

PhD

University of York

Biology

January 2023

Abstract

Cip1-interacting zinc finger protein 1 (CIZ1) is a nuclear matrix protein that forms large sub-nuclear assemblies at the inactive X chromosome (Xi) in females, and smaller assemblies throughout the nucleus in males and females. CIZ1 is linked with maintenance of histone modifications that specify facultative heterochromatin, and is extensively mis-expressed in human cancers, including under-expression, over-expression, and mis-splicing events.

Here, I describe uncoupled expression of the CIZ1 N-terminal replication domain and C-terminal anchor domain (AD), leading to over-expression of AD amplicons in breast cancer and derived cell lines. Modelling over-expression of the AD in murine cells led to observed dispersal of endogenous CIZ1 assemblies at the Xi. Therefore, AD fragments appear to have dominant negative (DN) activity, and these DN effects coincide with H2AK119ub1 and H3K27me3 loss from Xi chromatin.

Further analysis revealed that DN effects occur in early G1, when new CIZ1 assemblies normally reform at the Xi after mitosis. A mutagenesis screen identified the matrin 3-homology domain as important for the DN effect on endogenous CIZ1 assemblies, and for stable self-interaction to form a homo-dimeric complex *in vitro*, suggesting that DN interference involves disruption of CIZ1 dimers.

Additionally, I investigated the CIZ1B variant that is prevalent in lung cancer. I show that exclusion of part of the acidic domain in CIZ1B does not influence dimerisation. However, a mass spectrometry interaction study revealed increased and decreased affinity for protein interaction partners in the C-terminus of CIZ1B compared to wild-type, most notably CIZ1B had a reduced ability to bind to linker histones and DNA damage response associated proteins.

These observations suggest that CIZ1 could be implicated in mechanisms for protection of the genome and epigenome, and that aberrant expression of CIZ1 in early-stage cancers could therefore contribute to genome-wide loosening of gene repression, and a move toward epigenetic decay.

List of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgements

It's hard to summarise four years of "thank yous" in one section but I'll give it a go. First up I'd like to thank my supervisor Dawn Coverley for offering guidance with experimental design and providing much needed encouragement when things didn't go to plan! Additionally, I'd like to thank my co-supervisor Fred Antson and my TAP member Bob White, for their input and support in this project. I also want to give a huge thank you to my funder Chris Gatenby who singlehandedly made this project a reality via the creation of the "Georgina Gatenby Scholarship", and to whom I am eternally grateful. I also need to thank Julie Tucker for her guidance early in the project when conducting my mutagenesis experiments, Maria Chechik in the Antson lab for providing continued training on the AKTA system, and Andrew Mason for conducting the bioinformatics and for his assistance with the analysis and interpretation of the results.

Next I would like to thank all members of the Coverley lab past and present, including Justin, Emma, Louisa, Grace, Lewis, Elena, Jo, Caitlin and the many others who I can't all list. Thank you all for making the lab a really supportive environment and being there to lend an ear to my science and non-science based conversations. A special shoutout goes to Justin for teaching me lab techniques and providing crucial insight into the expansive literature. I also need to highlight Louisa my lab "mum" who was so helpful with lab organisation and day to day support and provided an endless supply of fun facts. Emma my lab "auntie" who has been here from the start and helped me grow into a self-functioning scientist, and in return I helped her move house on quite possibly the hottest day I've ever experienced in the UK. Grace my lab "sister" who always kept me sane and made me laugh on both the bad and the good days, and who I will miss not spending everyday with.

Another huge thank you goes out to all my family and friends who spurred me on in those final months and forgave me for temporarily dropping off the face of the earth on them. My parents, my grandparents and my "in-laws" have all been amazingly supportive throughout in every way possible, from letting me practice my presentations on them, providing much needed supplies of groceries and just always being at the end of a phone. My mum in particular has always been my greatest ally, and I am so grateful for her unwavering support in all the years leading up to this, and no doubt beyond.

Finally, my partner Alex. My pandemic buddy, my "do you think if I did my experiment this way instead it would work better?", my favourite person. Thank you for always being there when I needed someone to lean on, and for spurring me along when I needed it the most. I couldn't have done it without you.

Declaration

I declare that this thesis is a presentation of original work and that 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.

Text and figures contained in Chapter 3 are part of a manuscript currently in production where I am the primary author.

Turvey, G.L., López de Alba, E., Sercombe, H., Williamson, L., Ainscough, J.F.X., Mason, A., Coverley, D. Dominant CIZ1 fragments interfere with nuclear CIZ1 assemblies and epigenetic stability in human cancers

My contributions to these publications have been reproduced in full.

All work generated by other authors, or instances where data has been generated as part of a collaboration with others, has been explicitly stated throughout.

1. Introduction

In mammalian cells there are many well-studied regulatory mechanisms in place to ensure homeostasis is achieved and that cells remain healthy. Their corruption can lead to development of a plethora of different pathologies, including cancer. Cancer is defined as disease caused by uncontrolled division of abnormal cells, leading to the development of tumours. If left untreated tumours can go on to metastasise to other locations in the body, and when this occurs patient outcomes become poor, and mortality is high.

Whilst all cells in an individual should have the same genome, the epigenome of individual cells is modified depending on cell type, and in response to stimuli or throughout ageing (Murrell et al., 2005). The epigenome can dynamically alter gene expression via modification of DNA, or associated proteins, and in cancer it is often corrupted (Sharma et al., 2010). Epigenome corruption can be caused by environmental exposures (Bai et al., 2018) or genetic mutations in the epigenetic machinery (Baylin and Jones, 2011), indeed three of the top ten most commonly mutated genes in cancer play roles in histone modification or histone remodelling (Mendiratta et al., 2021). This makes it of interest in cancer research, since clinal applications can be far reaching.

Cip1-interacting zinc finger protein 1 (CIZ1), has been implicated in epigenetic maintenance, and its loss in early passage embryonic fibroblasts was associated with a loss of repressive epigenetic histone modifications, leading to altered gene expression (Stewart et al., 2019). In addition to this, aberrant CIZ1 expression has been extensively implicated in tumourigenesis, these alterations are variable and include over-expression, under-expression and alternative splicing, however how these are implicated in disease remains unknown.

The main focus of my PhD work was to further profile CIZ1 expression early in tumourigenesis in breast and lung cancer, and to attempt to deconvolute the potential impact of changes, in relation to cancer initiation or development. Here, I outline concepts and literature relevant to the function of CIZ1, followed by a detailed description of the known roles of CIZ1 and its interacting partners with respect to its characterised domains and regions. Finally, details of aberrant CIZ1 expression that have been associated with pathology are documented, to highlight the breadth of altered CIZ1 expression that has been observed to date.

1.1. The cell cycle

1.1.1. Stages of the cell cycle

To facilitate the development of a multicellular organism, cells must be able to grow, divide and respond to differentiation cues. Production of two daughter cells involves progression through gap 1 (G1), when a newly divided cell increases in size, replicates its organelles in preparation for the downstream stages of the cell cycle, and receives growth promoting or growth inhibitory signals. In synthesis (S) phase DNA is replicated to create two chromatids for each chromosome, and in gap 2 (G2) it resolves problems in preparation for chromatid division and cytokinesis at mitosis (M) to create two daughter cells (Figure 1.1). The non-mitotic stages of the cell cycle (G1, S and G2) as a collective is referred to as interphase. There are at least three independent checkpoints in the cell cycle that ensure that no errors are propagated through to the next stage. These occur in late G1 to check for DNA damage prior to DNA replication; G2 to check for replication errors that occurred in S phase; and in M phase to ensure the chromosomes have aligned properly prior to segregation. If errors or incomplete assembly of underpinning structures are detected, the cell will normally either correct the errors and continue through the cycle, or initiate programmed cell death via apoptosis. Failure to accurately execute this decision can cause emergence of corrupted cell lineages.

However, most cells in the body are not actively cycling. If replication is not required, the cell can reversibly exit the cell cycle in G1, prior to the G1 checkpoint, into G0. In this state cells are referred to as quiescent cells and include adult stem cells, lymphocytes, hepatocytes and oocytes (Marescal and Cheeseman, 2020).
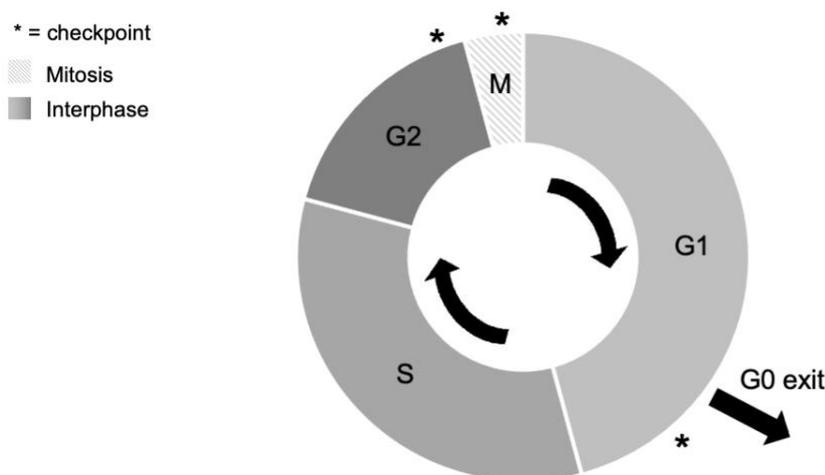
Figure 1.1. The cell cycle
Stages of the cell cycle. The cycle can be thought of as starting in gap 1 (G1), followed by synthesis (S) phase, then gap 2 (G2) and mitosis (M). The size of each segment is reflective of the approximate time a cell in culture spends in each phase, though this can vary significantly. Stars reflect the approximate timing of the checkpoints.

Mitosis is categorised into distinct phases based on the condensation state of chromosomes. Chromosomal changes were first formally documented by Walther Flemming (Flemming, 1879), which he subsequently named "mitosis", based on the Greek word for thread in reference to the chromosomes appearance. Additionally Eduard Strasburger coined some of the names of the different mitotic stages that are used today (prophase, metaphase and anaphase) (Strasburger, 1883). Today, we further breakdown the stages of mitosis into: prophase, prometaphase, metaphase, anaphase, telophase and cytokinesis, illustrated below in Figure 1.2.
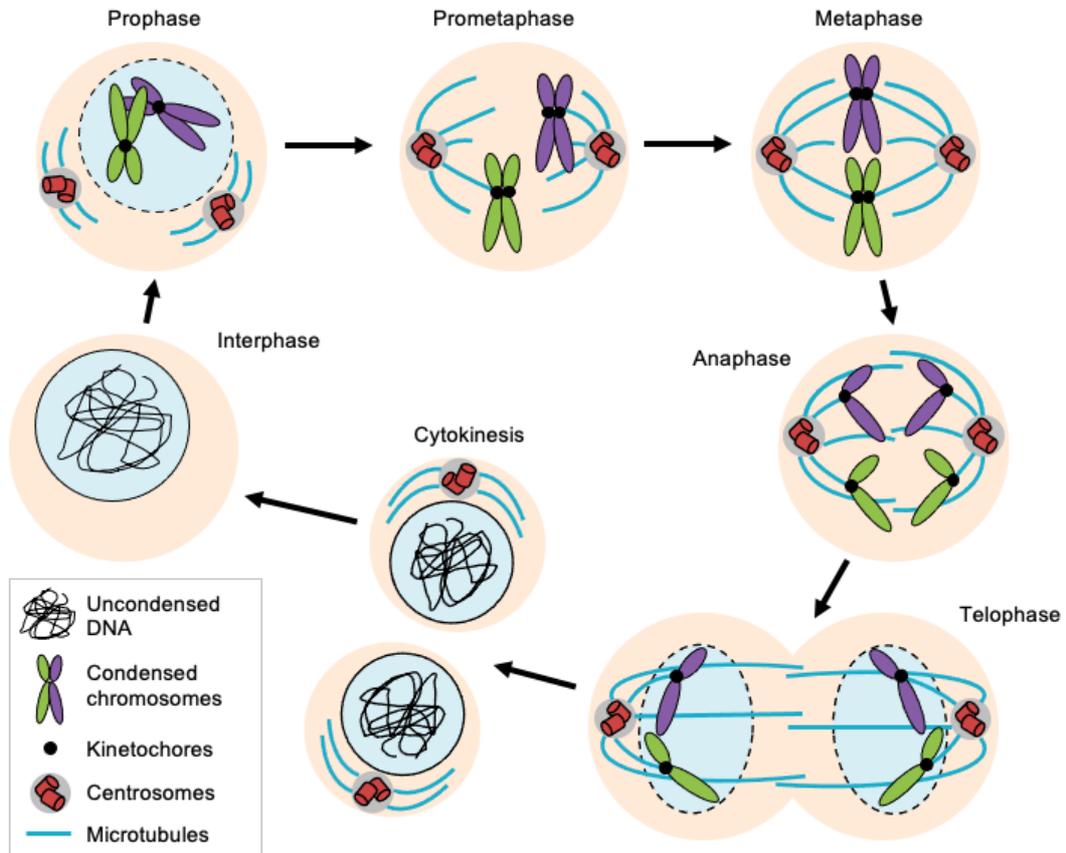
Figure 1.2. The stages of mitosis
Prophase - the chromatin begins to compact, the centrosomes begin to migrate to opposite poles, and the nuclear envelope starts to break down. Prometaphase - the nuclear envelope has fully broken down, the chromosomes are moved apart and begin progressing towards the metaphase plate, and microtubules that form the spindle fibres are also polymerising towards the middle of the cell. Metaphase - the chromosomes are aligned perpendicular to the spindle fibres, and each sister chromatid is attached via its respective kinetochore to microtubules that connect to the centrosomes at opposite poles of the cell. The mitotic checkpoint is conducted to ensure the chromosomes are aligned properly and that there are no problems with the segregation machinery. Anaphase - the sister chromatids are pulled towards the spindle poles, whilst simultaneously the spindle poles move further apart, thus facilitating full chromosome segregation. Telophase - two nuclear envelopes reform around the segregated chromosomes, and the DNA begins to decondense. Cytokinesis - the cytoplasm is divided by a contractile ring into two daughter cells, each containing a nucleus with the correct number of chromosomes.

1.1.2. Cyclins and cyclin dependent kinases in G1/S progression

Cyclin dependent kinases (CDKs), as their name suggests, are inactive until bound to their cyclin counterparts. Expression of cyclins controls the timing of activation of the CDKs whose expression is relatively consistent throughout the cell cycle (Obaya and Sedivy, 2002). Cyclin binding causes conformational changes to the CDK substrate site and removal of a blockade at the entrance (Jeffrey et al., 1995, Russo et al., 1996).

Different cyclin/CDK complexes control specific stages of the cell cycle, presented in Figure 1.3. The canonical model of G1/S progression suggests that in G1 cyclin D is produced in response to growth factors (Baldin et al., 1993), and activates CDK4/6 to target the retinoblastoma protein (pRb) (Kato et al., 1993). pRb arrests cells in G1 via suppression of the E2F transcription factor, which is responsible for (but not limited to) transcription of DNA polymerase, proliferating cell nuclear antigen (PCNA), cyclin E and cyclin A (DeGregori et al., 1995). When cells are arrested pRb is complexed with histone deacetylase (HDAC) and the hSWI/SNF nucleosome remodelling complex to maintain gene repression. Subsequently the cyclin D/CDK4/6 complex phosphorylates pRb, which leads to derepression of cyclin E expression (Zhang et al., 2000). This model is supported by the observation that cyclin E expression peaks in G1 prior to cyclin A (Koff et al., 1992). Next cyclin E/CDK2 and cyclin A/CDK2 hyperphosphorylate pRb to fully overcome pRb mediated suppression (Hinds et al., 1992). This stepwise phosphorylation is supported by the observation that there are different phosphorylation sites on pRb that are targeted by cyclin D/CDK4/6 and cyclin A/E/CDK2 (Kitagawa et al., 1996, Zarkowska and Mittnacht, 1997). Recent studies show that cyclin D/CDK4/6 is responsible for initial pRb inactivation which leads to a gradual increase in CDK2 activity, but whilst CDK2 activity is low cells can exit G1. Then when cyclin E/CDK2 activity reaches a threshold there is a switch where cells commit to progressing to the G1/S transition (Kim et al., 2022).

Here I have outlined the canonical model for cyclin/CDK timing and involvement at the G1/S transition, but there are still multiple unknowns and some experimental evidence that provides conflicting information, which is described in detail in a recent review (Rubin et al., 2020).

1.1.3. Cyclins and cyclin dependent kinases in S, G2 and M phase

Cyclin A has been shown to be required for both DNA synthesis in S phase and for entry into mitosis, by binding to CDK2 and CDK1 respectively (Pagano et al., 1992). Early in G2 cyclin A/CDK2 is implicated in the pathway that downregulates checkpoint kinase 1 (Chk1), thus allowing entry to mitosis (Oakes et al., 2014), and a cyclin A/CDK1 phosphorylation pathway is responsible for de-phosphorylation and subsequent activation of cyclin B/CDK1 which drives mitotic entry (Vigneron et al., 2018). During mitosis both cyclin A and cyclin B are targeted by the anaphase promoting complex/cyclosome (APC/C) for destruction via the ubiquitination pathway (Sudakin et al., 1995). This occurs in a stepwise manner, with destruction of cyclin A in prometaphase (den Elzen and Pines, 2001), then cyclin B in metaphase (Clute and Pines, 1999), to allow progression and exit out of mitosis.

1.1.4. CIZ1 and p21 involvement in cell cycle regulation

Initiation of programmed cell death in response to checkpoint activation in G1 is carried out by many proteins, but a key one to highlight here is p21, which is induced by p53. p53 has been dubbed the "guardian of the genome" and is a key tumour suppressor (Lane, 1992), and mutations in *TP53* (the gene encoding p53) are the most common mutation in tumours (Kandoth et al., 2013). If DNA damage is detected, p53 upregulates expression of p21, and early work showed that p21 inhibits all CDK2 complexes (Harper et al., 1993). Follow up work confirmed that p21 effectively inhibited CDK2, CDK4, and CDK6, and less effectively inhibited CDK1 (Harper et al., 1995). Notably, in CDK2 knockout (KO) mice, CDK1 was prematurely translocated to the nucleus, and able to compensate to facilitate G1/S progression. Under these circumstances, upon DNA damage induced by gamma irradiation, p21 effectively arrested cells via inhibition of CDK1 (Satyanarayana et al., 2008). This suggests that p21 inhibition is effective within particular phases of the cell cycle (G1 and S), rather than being selective against particular cyclin:CDK complexes. This p21 mediated inhibition allows the cell to prohibit cell cycle progression, so that repair or apoptosis can take place. Conversely to p53, Cip1-interacting zinc finger protein 1 (CIZ1) has been suggested to be modulating the repressive role p21 plays in cell cycle progression, by interacting with p21 and facilitating CIZ1/p21 complex sequestering in the cytoplasm (Mitsui et al., 1999). Thus, upstream regulators of p21, including p53 and CIZ1, are crucial to facilitate correct control over the cell cycle. Therefore, in instances where p53 or CIZ1 are corrupted, cells could progress through the cell cycle with DNA damage, or be inappropriately prevented from progressing.
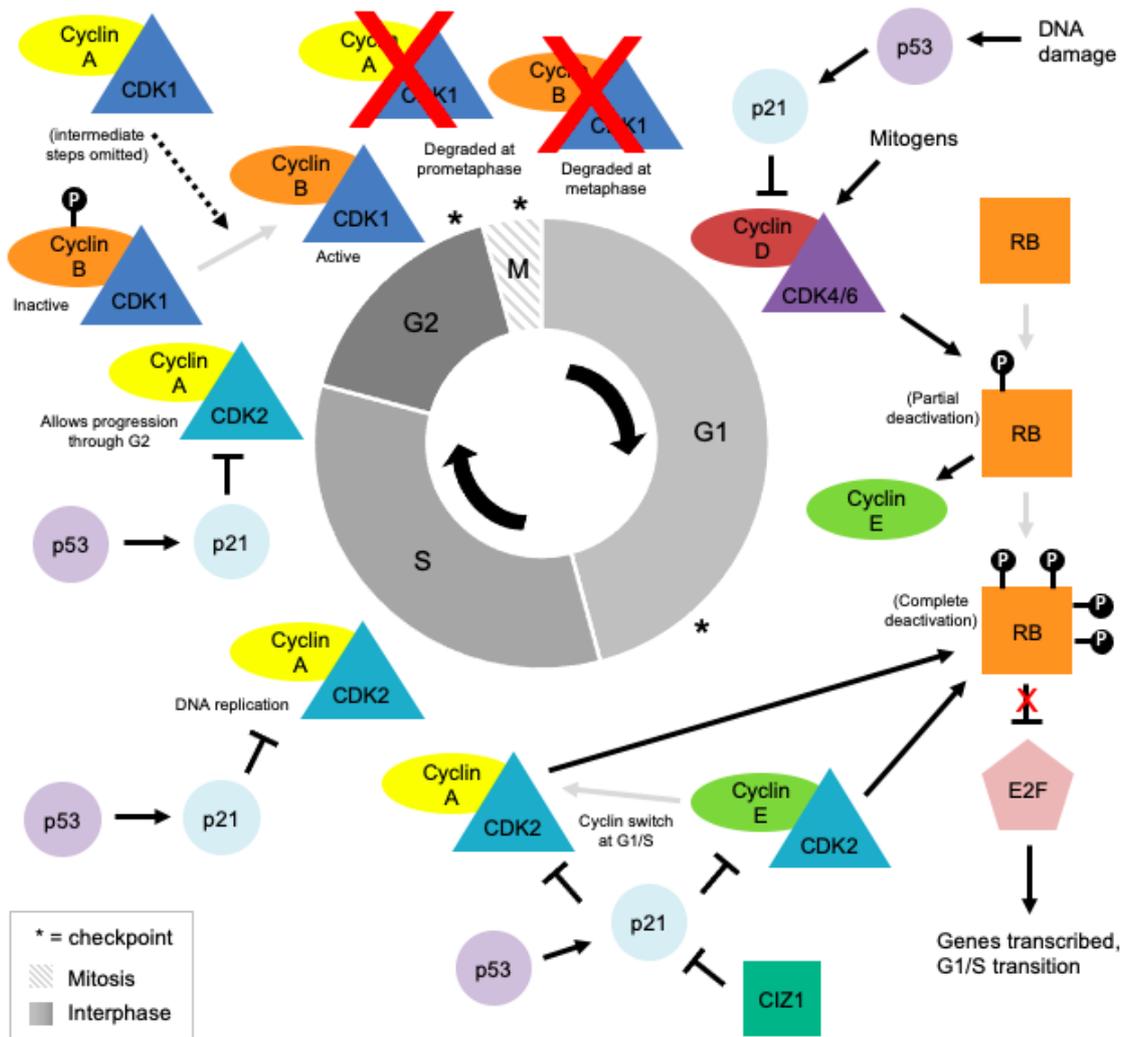
Figure 1.3. The timing of cyclins and CDKs in the cell cycle

The canonical model of the cyclin/CDK complexes implicated in the cell cycle and their approximate timing for involvement. Black sharp arrows indicate stimulation of a pathway, whereas black blunt arrows indicate inhibition of a pathway. A dashed arrow represents stimulation in a pathway where the intermediate steps have been omitted for simplicity. Grey arrows reflect stepwise events occurring in a pathway. "P" labels represent phosphorylation events.

1.1.5. Re-entry to cell cycle after quiescence

Not only are there requirements for cyclin/CDK complexes in cycling cells, cells that have exited the cycle and then re-enter upon stimulation can also be harnessed to reveal specific cyclin/CDK requirements upon entry from quiescence. Analysis in a cell free system revealed that during re-entry to the cell cycle from G0, the cyclin E/CDK2 complex supports assembly of pre-replication complexes, then with increasing cyclin A/CDK2 concentrations the assembly phase is terminated, and at higher concentrations DNA synthesis initiated. Thus by ordering events along a concentration gradient, cyclin A/CDK2 helps to prevent re-assembly and ensures that DNA is only replicated once in each cycle (Coverley et al., 2002). In these *in vitro* studies, when the cyclins were not applied at the correct concentration in the right order, G1/S transition was limited.

These observations are supported by the result that loss of cyclin E in mice during development was not detrimental, but that murine cells were unable to re-enter the cell cycle from G0 in the absence of cyclin E (Geng et al., 2003). Contrastingly, loss of cyclin D during mouse development was lethal, but cyclin D deficient murine fibroblasts can re-enter the cell cycle, albeit with a greater requirement for mitogen stimulation (Kozar et al., 2004). This differing requirement for cyclin presence in continuous compared to non-continuous cell cycling is still not fully understood. This points to a key area of research, when considering that post development most cells are not actively cycling, and instead are exiting and entering the cell cycle in response to stimuli (Viatour, 2012). Thus, in the development of adult disease such as cancer, variations to the canonical models must be considered, and any corruption evaluated in the correct context.

1.2. CIZ1 structure and normal function

CIZ1 was first cloned in 1999 (Mitsui et al., 1999) as an interaction partner of p21. As described above, it was suggested that these interactions could be modulating the repressive role p21 plays in cell cycle progression, by facilitating CIZ1/p21 complex sequestering in the cytoplasm (Mitsui et al., 1999). Subsequent work showed that, in addition to this, the protein contains two distinct regions that perform independent functions and facilitate its involvement in several different cellular processes. The N-terminal replication domain (RD) has been implicated CIZ1's role in DNA replication (Coverley et al., 2005), whereas the C-terminal anchor domain (AD) is required to anchor CIZ1 to the nuclear matrix (NM) (Ainscough et al., 2007). The diagram below (Figure 1.4) shows the domains contained in CIZ1, and features that facilitate its involvement in key processes.
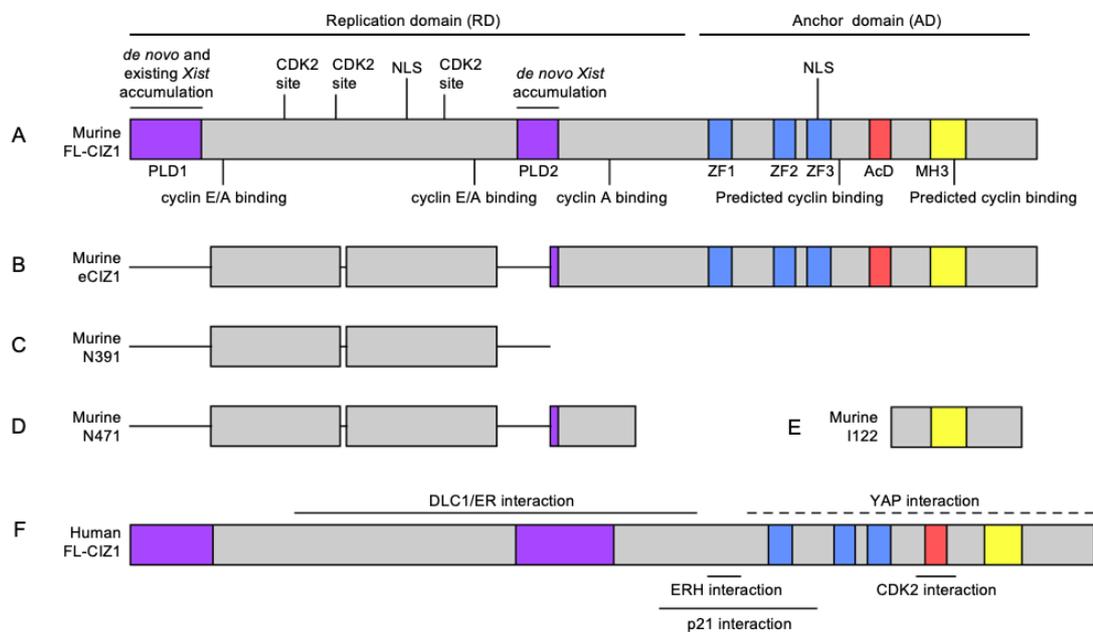
Figure 1.4. CIZ1 domains, regulatory and interaction sites, and fragments used in analysis

A. Full length murine CIZ1 (UniProt Q8EVH2). Purple: prion like domain 1 and 2 (PLD1) and (PLD2) as documented in (Sofi et al., 2022). Blue: C2H2 zinc finger 1-3 (ZF 1-3). (ZnF_C2H2 SM00355, ZF_C2H2 sd00020 and ZF_C2H2 sd00020 respectively). Red: the acidic domain (AcD). Yellow: the matrin 3-homology (MH3) domain (ZnF_U1 smart00451). Nuclear localisation sequences (NLS), CDK2 phosphorylation sites and cyclin binding sites are annotated and discussed in the main text.

B. Embryonic CIZ1 (eCIZ1) lacking regions in exon 2, 6 and 8. Almost all of PLD1 and PLD2 is excluded by alternative splicing, while all C-terminal domains remain intact.

C. and D. Truncated N terminal fragments created from eCIZ1 used to identify CIZ1 cyclin binding sites (Copeland et al., 2010).

E. C-terminal fragment used to measure CIZ1 anchorage to the nuclear matrix (Ainscough et al., 2007).

F. Full length human CIZ1 (UniProt Q9ULV3), used in (Higgins et al., 2012). Domains are the same as those described in A, with the exception of a larger PLD2 region.

1.2.1. CIZ1 replication domain

As stated above the N-terminal replication domain (RD) is associated with CIZ1's role in DNA replication (Coverley et al., 2005). The additional domains and motifs contained within the RD include: An N-terminal nuclear localisation sequence, cyclin binding motifs and prion like domains, and they are described in full below.

1.2.1.1. NLS contained in the replication domain

The main subcellular location of CIZ1 is the nucleus (Coverley et al., 2005). This is facilitated by a conserved classical K-K/R-X-K/R nuclear localisation sequence (NLS) (Chelsky et al., 1989) in the N terminus of CIZ1 (Figure 1.4A). There is also a bi-partite NLS (K/R-K/R-X_{10}-K/R in ≥3/5 of the following residues) (Robbins et al., 1991) located within the third zinc finger of CIZ1 (Figure 1.4A). Mutation of the N-terminal NLS in murine CIZ1 (Figure 1.4A) led to highly perturbed nuclear targeting in multiple cell lines, but contrastingly mutation of the C-terminal NLS appeared to have no effect in the context of full length CIZ1 (Sofi et al., 2022). However, the retained ability for CIZ1 nuclear import in ~30% of cells, despite the removal of the N-terminal NLS, suggests that the bi-partite NLS or some other region may encode some function to facilitate nuclear targeting.

1.2.1.2. Cyclin binding motifs and CDK2 phosphorylation sites

Embryonic CIZ1 (eCIZ1) (Figure 1.4B), an alternatively spliced version of CIZ1 derived from day 11 murine embryos lacks exon 2, part of exon 6 and part of exon 8 (Coverley et al., 2005). However, it retains the 5 K/R-X-L cyclin binding motifs, three of which are contained in the RD (Copeland et al., 2010) (annotated on FL-CIZ1 in Figure 1.4A). Comparison of two RD forms of CIZ1, N391 (Figure 1.4C) which contains two of the RD cyclin binding sites, and N471 (Figure 1.4D) which contains all three of the RD cyclin binding sites, revealed cyclin interaction differences. Cyclin E directly interacted with both forms, whereas cyclin A only interacted with the larger N471, indicating that the third cyclin binding motif is responsible for cyclin A interactions and that cyclin E interacts with one or both of the upstream sites. Wherever cyclin interactions were observed this also coincided with CDK2 recovery, suggesting that CDK2 does not interact with CIZ1 RD directly, but instead interacts via the cyclins (Copeland et al., 2010).

This led to the suggestion that CIZ1 helps coordinate cyclin delivery and function by sequentially recruiting them to the same site in the correct order. This suggestion is supported by the observation that in isolated nuclei if CIZ1 is depleted, entry to S phase is restrained (Coverley et al., 2005). Additional analysis identified that three sites in the RD of CIZ1 were phosphorylated by CDK2 during S and G2 phase, and this phosphorylation led to the loss of cyclin A/CIZ1 interactions. Therefore CIZ1 is suggested to be acting as a kinase sensor by promoting initiation of DNA replication at low kinase levels, since when un-phosphorylated it is able to deliver cyclin A/CDK2 complexes, but when kinase levels rise and it is phosphorylated it no longer interacts with cyclin A/CDK2 (Copeland et al., 2015). While these analyses describe a regulatory pathway, they do not address what CIZ1 actually does during replication.

1.2.1.3. CIZ1 prion-like domains

Prion like domains (PLDs) are low complexity sequences found in RNA binding proteins that are named as such due to their similarities to prion domains (Ross et al., 2005). These PLDs contain polyglutamine repeats that have been extensively implicated in neurodegenerative diseases and are linked with protein misfolding and aggregation (Scherzinger et al., 1997, Warrick et al., 1998, Rubinsztein et al., 1999). Murine and human CIZ1 both contain two PLDs (PLD1 and PLD2) in their RD. These are in the same location in both species, but in human CIZ1 PLD2 is longer and subject to more complex alternative splicing. CIZ1's PLD domains have been implicated in its RNA-dependent recruitment to the inactive X chromosome (Sofi et al., 2022) and will be described in detail in section 1.5.7.

In human cells, a region containing PLD2 has been shown to interact with dynein light chain 1 (DLC1) (den Hollander and Kumar, 2006) (Figure 1.4F). Here it was documented that DLC1 is overexpressed in breast cancers, and that downregulation resulted in reduced cell cycle progression in breast cancer cell lines. The same region has also been implicated in oestrogen receptor (ER) binding (den Hollander et al., 2006), and it was shown that CIZ1 promoted recruitment of the ER complex to target gene chromatin, leading to oestrogen hypersensitivity. Follow up experimentation is required to understand what role the PLD might play in this.

1.2.2. CIZ1 anchor domain

Since the anchor domain (AD) of CIZ1 is required for its immobilisation within insoluble non-chromatin sub-nuclear structures (Ainscough et al., 2007), it can be thought of as part of the nuclear matrix. The exact region or regions within the AD that are responsible for these interactions remains unknown. Additionally, the AD has been implicated in interactions with yes-associated protein (YAP), which regulates the Hippo signalling pathway, where CIZ1 overexpression was associated with higher transcriptional activity of YAP in hepatocellular carcinoma (Lei et al., 2016). This region is illustrated with dashed lines in Figure 1.4F since the authors did not provide interaction boundaries. The domains and motifs contained within the AD include: the triple set of zinc fingers, the acidic domain and the matrin 3-homology domain. These are described in detail below.

1.2.2.1. CIZ1 zinc fingers

Zinc fingers (ZF) are a protein motif that contain a tetrahedrally coordinated zinc ion for stability of their secondary structure (Klug, 2010). CIZ1 contains three C2H2 ZF motifs within the AD, and these are conserved with a sequence identity of 86-96% in murine and mouse CIZ1. ZF containing proteins have a diversity of functions, including but not limited to, DNA recognition, RNA packaging, gene expression regulation, protein folding and assembly, and lipid binding (Laity et al., 2001). The C2H2 ZF is well characterised, and historically has been implicated mostly in interactions with DNA (Berg, 1988), however more recent reviews discuss their involvement in protein:protein interactions (Brayer and Segal, 2008) and interactions with RNA (Hall, 2005). Indeed, the triple ZF consensus that is contained in CIZ1 has been implicated in all three interactions, and whilst a triplet of ZFs can only interact with one ligand at a time, in the context of a multimeric protein it is suggested that this could allow multiple interactions at once and act as a bridge between two ligands (Iuchi, 2001). This suggestion aligns with CIZ1 when considering the multivalent interactions described in Sofi et al., and could be one of the driving interactions for CIZ1 NM anchorage.

Early work that first identified CIZ1 as a novel interactor with p21/Cip1, noted that it interacts via an area that contains ZF1 (Figure 1.4F). Additionally, CIZ1 was shown to interact with enhancer of rudimentary homolog (ERH) (Lukasik et al., 2008), a highly conserved protein in eukaryotes that is not well understood but has been implicated in cell cycle and transcriptional regulation (Weng and Luo, 2013). Recent analysis confirmed a 31 residue region upstream of ZF1 that is responsible for CIZ1/ERH interactions, and that CIZ1/ERH forms a 2:2 hetero-tetramer *in vitro* (Wang et al., 2022). Finally CIZ1 has been shown to interact with the DNA sequence ARYSR(0-2)YYAC (Warder and Keherly, 2003), this interaction could be mediated by the ZFs but has not been formally tested.

1.2.2.2. CIZ1 acidic domain

Downstream of the triplet of ZFs there is a dense region of negatively charged aspartates and glutamates, (21/24 of the amino acids present in both murine and human CIZ1), referred to as the acidic domain (AcD). The region containing the AcD between the three ZFs and the MH3 domain is not predicted to contain secondary structure based on modelling using the AlphaFold consortium (Jumper et al., 2021), and is predicted to be highly disordered compared to the highly ordered ZFs and MH3 domain, as shown below in (Figure 1.5).
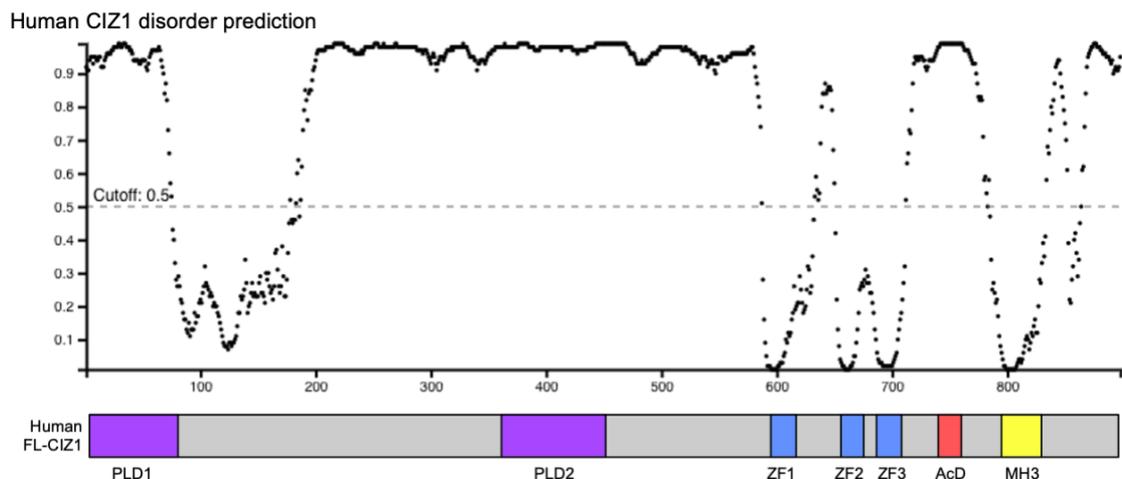
Figure 1.5. Disorder prediction of human CIZ1
Predicted areas of disorder in human full length CIZ1 using DISOPRED software (Jones and Cozzetto, 2015). This is aligned with a diagram of CIZ1 to facilitate interpretation of the disorder profile in the individual domains of CIZ1.

The AcD resembles a transactivation domain (TAD), more specifically the "acid blobs" or "negative noodles" that are unstructured regions with an excess of negatively charged amino acids (Sigler, 1988). TADs provide a binding site for other proteins required for transcriptional regulation, possibly suggesting that CIZ1 could be responsible for bridging interactions *in vivo*.

The region that contains the AcD has also been implicated in direct interactions with CDK2 (den Hollander and Kumar, 2006). As described above, a region containing PLD2 was also implicated in DLC1 interactions in the same study, and it was hypothesised that the multiprotein complex comprised of CDK2, CIZ1 and DLC1 promote cell cycle progression by reducing the nuclear levels of the cell cycle inhibitor p21.

Finally, a cancer specific splice variant of CIZ1, referred to as CIZ1B, involves the deletion of 8 amino acids at the boundary of the AcD, but how this is implicated in tumourigenesis is not known. Thus, overall the AcD of CIZ1 remains unclear and is the subject of my investigation in Chapter 5.

1.2.2.3. CIZ1 matrin 3-homology domain

The matrin 3-homology (MH3) domain in CIZ1 is named due to its homology with a region of matrin 3 (Belgrader et al., 1991, Mitsui et al., 1999). Matrin 3 was first identified as part of a novel set of nuclear matrin proteins (Nakayasu and Berezney, 1991). The function of matrin 3 remains poorly understood, but has been implicated in interactions with DNA, RNA and proteins, and mutations in matrin 3 are associated with development of familial ALS (Malik and Barmada, 2021). Given the lack of understanding of matrin 3 function, we can't easily draw clues about the function of the MH3 homology domain in CIZ1. It is a U1-like zinc finger (SMART entry SM00451), and inspection of proteins listed on the InterPro database (Paysan-Lafosse et al., 2022) that also contain this domain returned 30 reviewed proteins in *Homo sapiens*. Gene ontology analysis (Ashburner et al., 2000) of these proteins to identify MH3 domain function yields limited information. All 30 proteins have a nucleic acid binding molecular function, with 14 of these as RNA binders, which is unsurprising given that the MH3 domain is a class of ZF. Previous analysis of a small C-terminal CIZ1 fragment referred to as I122 (Figure 1.4E), that only contains the MH3 domain, identified that it was able to remain associated with the chromatin-depleted NM (Ainscough et al., 2007). Thus, despite not knowing its exact role *in vivo* and what interactions are facilitated by the MH3 domain, it appears to be a key domain responsible for CIZ1 AD function, and is investigated in Chapter 3.

1.3. Nuclear matrix

1.3.1. Early work identifying NM proteins

The existence of a nuclear matrix (NM), an apparent structural framework of proteins retained in rat liver nuclei after treatment with a high concentration of salt, deoxyribonuclease (DNase) and ribonuclease (RNase), was visualised nearly 50 years ago by electron microscopy (Berezney and Coffey, 1974). Subsequent analysis revealed that the majority of these proteins were acidic proteins (Berezney and Coffey, 1976). It was noted that when RNase treatments were excluded from the extraction, there was a fraction that was approximately 20% RNA in composition (Berezney, 1980), suggesting that RNA could be an integral part of the NM. The structural entities retained after NM extraction procedures can be categorised as: i) the outer nuclear pore complex composed of nuclear lamins and pore proteins, ii) nucleoli residue, and iii) the inner matrix that includes heterogeneous nuclear ribonucleoproteins (hnRNPs), small nuclear ribonucleoproteins (snRNPs), nuclear actin and enzymes associated with DNA and RNA metabolism (Verheijen et al., 1988). Follow up work identified that hnRNPs and B23 (also referred to as nucleophosmin (NPM)) were major components of the internal NM in HeLa cells (Mattern et al., 1996). In rat livers hnRNPs and B23 were also consistently identified in the NM, alongside lamin A, B and C, and additionally a novel set of nuclear matrin proteins (matrins 3, 4, 12 and 13 and matrins D-G) were identified (Nakayasu and Berezney, 1991).

Advances in high throughput screening techniques allows for more in depth identification of NM proteins. One study utilised mass spectrometry to identify 333 NM proteins in HeLa cells, and showed enrichment for disordered proteins, alongside many other previously documented NM proteins (Ishii et al., 2008). In this study 250mM ammonium sulphate was used for the extraction process. Another study that compared 25mM lithium 3,5-diiodosalicylate (LIS, a mild anionic protein solubilising detergent), 250mM ammonium sulphate and 2M sodium chloride (NaCl) extraction in mouse lymphocytes, identified 320, 346 and 344 NM proteins respectively, with 272 shared across all three methodologies (Engelke et al., 2014). Further work comparing 2M NaCl and LIS extracted cells revealed that whilst LIS could be useful for solubilising NM proteins for downstream analysis, the cells lacked morphologically distinct residual nucleoli and their internal structure was depleted (Smith et al., 1987). Lack of preservation of a recognisable structure, compounded with highly variable techniques across labs, has led to much disagreement in the field. Nonetheless results obtained from Ishii et al. and Engelke et al. suggest there are approximately 300 NM proteins present in most cells, though these are not necessarily the same across cell types.

1.3.2. NM protein roles and retention via S/MARs

Early work identified that actively transcribed genes are associated with the NM, suggesting that the NM is the site of transcription (Ciejek et al., 1983). Indeed it has been shown that NM proteins provide the structure required for DNA replication, RNA synthesis and processing, nuclear transport and steroid hormone action (He et al., 1995). NM proteins attach to specific regions of DNA referred to as scaffold attachment regions (SARs) (Mirkovitch et al., 1984) and matrix attachment regions (MARs) (Cockerill and Garrard, 1986). These are sometimes referred to collectively as S/MARs. Whilst the differences in extraction technique produces a different pool of NM proteins, the differences are not substantial. However, comparison of high salt and LIS extractions techniques reveals different attachment regions, thus routinely today MARs correspond to high-salt extractions and SARs correspond to LIS extractions (Dobson et al., 2017). Only half of all the attachment regions are shared between MARs and SARs. When comparing the two, MAR enrichment is biased towards intergenic regions and when contained within a gene is associated with silenced genes, however SARs exhibit a more even distribution across the chromosome and when upstream of a gene they are associated with transcript presence (Linnemann et al., 2009). Despite their potential differences in roles, which remains largely unknown, both are characterised by AT rich regions and topoisomerase II DNA binding sites (Boulikas, 1993).

1.3.3. CIZ1 as a component of the NM

CIZ1 has been identified as a NM protein based on its resistance to high-salt extraction and DNase digestion of chromatin (technique described in (Wilson et al., 2016)). Based on data that shows a cyclin-dependent role in initiation of DNA replication (Copeland et al., 2010), it has been suggested to form a molecular link between the DNA replication machinery and the sub-nuclear structures that organize their function (Ainscough et al., 2007). Importantly, the fraction of CIZ1 that accumulates at the inactive X chromosome (see section 1.5) is retained upon application of detergent, high salt and DNase, but is lost upon application of RNase during most of the cell cycle. This suggests that most of the time CIZ1 is part of an RNA-dependent nuclear matrix at Xi (Ridings-Figueroa et al., 2017).

1.4. Epigenetics

1.4.1. Early work and definitions

The term "epigenetics" was coined by Conrad Waddington, where he described how cells can differentiate into alternative lineages during development leading to different cell fates (Waddington, 1957). Subsequent early work in this field focused more on the discussion that epigenetic components were inheritable, and could persist throughout multiple rounds of cell division, however there was still much confusion in the field regarding the absolute definition of what epigenetic regulation included (Nanney, 1958). Key studies have shown that changes in gene expression, that are caused by epigenetic changes due to environmental exposures, can be inherited. An example of this is the observation that the children of women who were in the early stages of gestation during the Dutch famine, have a higher likelihood of developing various health issues such as coronary heart disease or COPD (Roseboom et al., 2006). Additionally, the maternal grandchildren are more likely to have "poorer health" such as cancer and autoimmune diseases, presumably due to the *in utero* undernutrition their mother received (Painter et al., 2008). This highlights that the epigenetic effects of environmental exposures can be long lasting. However more recently the definition of epigenetics has been relaxed to include non-inheritable changes, and is defined as: "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). This allows inclusion of more transient changes that alter gene expression, such as certain histone modifications.

1.4.2. Histone post translational modifications

There are many different histone post translational modifications (PTMs), that form a "histone code" that can be read by other proteins (Strahl and Allis, 2000). These PTMs can be associated with gene activation or gene repression, and can be modified by "writers" or "erasers" and interpreted by "reader" proteins (Hyun et al., 2017). A recent comprehensive review of known histone modifications, details 82 different histone modifications in humans with known functions and modifying enzymes, with yet more of unknown function (Zhao and Garcia, 2015). The PTMs mentioned throughout this body of work have been discussed further below, or have been introduced in the individual chapter introductions.

1.4.3. Polycomb group proteins gene regulation

Polycomb group proteins (PcGs) were discovered decades ago as being responsible for silencing of *Hox* genes in certain parts of the body of *Drosophila melanogaster* during development (Lewis, 1978). Homologs of PcGs are found in all metazoans, and there is a high degree of conservation between *Drosophila melanogaster* and *Homo sapiens* (Wang et al., 2015). Thus unsurprisingly, PcGs are also key regulators of developmental genes in human (Lee et al., 2006) and mice (Boyer et al., 2006) embryonic stem cells. In mammals there are two different PcG complexes, referred to as polycomb repressive complex 1 and 2 (PRC1 and PRC2). PRC1 is responsible for laying down ubiquitin (a small 8.6 kDa protein that is detected in animals, bacteria and yeast (Goldstein et al., 1975)) at lysine 119 on histone H2A (H2AK119ub1) (Wang et al., 2004), and PRC2 is responsible for laying down trimethylation (three CH_3 groups) at lysine 27 on histone H3 (H3K27me3) (Cao et al., 2002). Both H2AK119ub1 and H3K27me3 PTMs are associated with gene repression (Cao and Zhang, 2004, Wang et al., 2004). There are variants of PRC1 (Gao et al., 2012) and PRC2 (Guo et al., 2021) identified in humans, that can be associated with canonical and non-canonical recruitment, and varying capabilities of compacting chromatin (Simon and Kingston, 2009, Gao et al., 2012). Where this is relevant to X-inactivation, this is described in section 1.5.5, that details PcG involvement in this process, where both PRC1 and PRC2 are directly recruited and are key for correct gene silencing.

1.4.4. CIZ1 involvement in epigenetic maintenance

As mentioned in the preface, CIZ1 has been implicated in epigenetic maintenance, where loss in primary embryonic fibroblasts (PEFs) was associated with a loss of repressive epigenetic histone modifications H3K27me3 and H2AK119ub1 at the inactive X chromosome (Stewart et al., 2019). Though, because CIZ1-null cells show genome wide de-regulation of polycomb regulated genes, it was suggested that it may perform similar functions across the genome (Stewart et al., 2019). This makes it of interest to study when considering the consequences of altered epigenetic events that are observed in cancers (Jones and Baylin, 2002).

1.5. X-inactivation

1.5.1. The function of X-inactivation

Since female cells contain two X chromosomes, but males only contain one, the process of X-inactivation to epigenetically silence one of the X chromosomes is a mechanism of dosage compensation in female mammals (Ohno et al., 1959). This leads to the development of a visible and dense heterochromatin structure in female cells that was coined the Barr body (Barr and Bertram, 1949). In murine development, the extraembryonic tissue that goes on to form the placenta and other supportive tissue during foetal development, undergoes imprinted X-inactivation, whereby the paternal X chromosome is preferentially silenced (Takagi and Sasaki, 1975). Contrastingly, in the inner cell mass the paternal X chromosome is reactivated, and during differentiation into epiblasts random X-inactivation occurs, and this is completed by approximately day 6 of gestation (Monk and Harper, 1979). Once the paternal or maternal X-chromosome has been inactivated (termed the Xi), this is maintained in subsequent cell divisions, leading to a mosaic pattern of X-inactivation in adult organisms (Lyon, 1961).

1.5.2. The role of *Xist* in early stages of X inactivation

The long non-coding RNA (lncRNA) X-inactive specific transcript (*Xist*) was discovered in 1991 in both humans (Brown et al., 1991) and mice (Brockdorff et al., 1991). Early in the stages of X-inactivation *Xist* is transcribed from the X-inactivation centre (XIC), and accumulates in *cis* (Clemson et al., 1996) via interactions with YY1 (Jeon and Lee, 2011), and then sequentially spreads to sites across the X-chromosome based on their 3D distance from the XIC (Engreitz et al., 2013).

Xist is required for X-inactivation (Penny et al., 1996), and female mice that inherited a mutated paternal *Xist* gene were severely growth-retarded and died early in embryogenesis (Marahrens et al., 1997). Contrastingly male mice were healthy, and female mice that inherited a defective maternal *Xist* gene, were also normal, but their paternal X-chromosome was always inactive. This paternal specific lethality is due to the failure to conduct the imprinted X-inactivation in the extraembryonic tissue as described above. Creation of female mice that only lacked *Xist* in the epiblast, to avoid the lethality associated with loss of the imprinted Xi in the extraembryonic tissue, revealed that loss of *Xist* in this instance was not lethal (Yang et al., 2016). However, the female mice were smaller at birth, displayed organ abnormalities, and did not survive post weaning. Thus, highlighting the importance of *Xist* in homeostasis.

1.5.3. *Xist* structure

Xist contains 6 separate regions of tandem repeats referred to as repeat A-F. The most 5' repeat, repeat A, is required for *Xist* mediated silencing of the Xi, whereas downstream elements were implicated in *Xist* targeting to the Xi (Wutz et al., 2002). It is now known that in murine *Xist* repeats B and C are required for recruitment of polycomb complexes, repeat E is implicated in *Xist* targeting and repeat D and F remain less understood (reviewed in (Raposo et al., 2021)). Raposo et al. also highlighted that most of the information we have today has been generated using mouse models, and that there are differences that have been observed between human and mouse X-inactivation that the scientific community are only now beginning to investigate further. One example is the recent discovery that in human cells, in addition to repeat A, repeat F and the terminal 3' end of the *Xist* lncRNA is also required for silencing (Dixon-McDougall and Brown, 2022). This highlights that whilst mouse models have many merits, the data they yield should be validated in human cells wherever possible.

1.5.4. *Xist* binding partners

Early work identified scaffold attachment factor A (SAF-A) as a crucial interacting partner for *Xist* function. SAF-A interacts with *Xist* via its RGG RNA binding domain (Helbig and Fackelmayer, 2003), and when SAF-A is lost *Xist* is dispersed across the nucleoplasm (Hasegawa et al., 2010). Proteomic studies to attempt to identify the full interactome of *Xist* have only been conducted more recently (Chu et al., 2015, McHugh et al., 2015, Minajigi et al., 2015), with varying interacting partners identified depending on the methodology utilised (Moindrot and Brockdorff, 2016). Key *Xist* interactors include cohesins, condensins, and chromatin remodellers, that synergistically repress transcription from the Xi (Minajigi et al., 2015); the SMRT and HDAC associated repressor protein SHARP (also referred to as SPEN) that interacts with SMRT to activate HDAC3 (McHugh et al., 2015); and the lamin B receptor (LBR) that recruits the Xi to the nuclear lamina via *Xist* interactions, enabling spreading of *Xist* to actively transcribed genes across the X-chromosome (Chen et al., 2016).

1.5.5. Epigenetic modifications at the Xi

Early studies showed that the trimethylation at lysine 9 on histone H3 (H3K9me3) mark is an early feature at the Xi (Heard et al., 2001, Mermoud et al., 2002), and that whilst both H3K9me3 and H3K27me3 enrichment are a feature of the Xi, they rely on different methyltransferases so play complementary and nonredundant roles (Rougeulle et al., 2004). As previously described above, the H3K27me3 PTM is laid down by PRC2, which was shown to be recruited to the Xi via *Xist* (Plath et al., 2003). Additionally *Xist* has been implicated in Xi nucleolar association during S phase for maintenance of transcriptional silencing, where *Xist* loss led to erosion of H3K27me3 accumulation at the Xi (Zhang et al., 2007). However, it has been noted that when using a mutant form of *Xist* that coats the Xi but does not silence it, PRC2 was still recruited, suggesting that H3K27me3 alone is not sufficient for gene repression (Plath et al., 2003). More recent studies show that the PRC2 cofactor Jarid2 is implicated in PRC2 Xi targeting, by acting as an intermediate between PRC2 and *Xist*, and Jarid2 loss prevents efficient PRC2 and H3K27me3 enrichment to *Xist*-coated chromatin (da Rocha et al., 2014).

In addition to PRC2, there is also recruitment of PRC1, to lay down the H2AK119ub1 mark, and it has been shown that double deletion of the RING1A and RING1B subunits of PRC1 led to loss of ubiquitination at the Xi (de Napoles et al., 2004). The canonical pathway of PRC recruitment suggests that PRC2 lays down H3K27me3 first, and then PRC1 binds to H3K27me3 to lay down H2AK119ub1 (Cao et al., 2002, Bernstein et al., 2006). However, in Eed deficient mice (one of the subunits of PRC2), PRC1 was still recruited via *Xist* interactions with RING1B, despite the lack of H3K27me3 (Schoeftner et al., 2006). Additionally, in PRC2 deficient mice, a complex of the RING1 and YY1-binding protein (RYBP) and PRC1 is recruited to the Xi (Tavares et al., 2012). Finally, it was shown that the noncanonical polycomb group RING finger 3/5 (PCGF3/5)-PRC1 complex, recruits both PRC1 and PRC2 in response to *Xist* expression, and that PCGF3/5 removal was lethal in female mice (Almeida et al., 2017). This recruitment is aided by hnRNPK, which recruits PCGF3/5 to *Xist* to facilitate subsequent Xi targeting (Pintacuda et al., 2017). In addition to this, histone deacetylation by HDAC3 has been suggested to be one of the earliest events in X-inactivation, and partial deacetylation is required for H2AK119ub1 spreading (Zylicz et al., 2019). Thus, the current model suggested in mice is that non-canonical PRC1 is responsible for recruitment of PRC2 and subsequently canonical PRC1, and together these epigenetic modifications are implicated in early X-inactivation.

1.5.6. Maintenance of X-inactivation

There is conflicting evidence on the requirement for *Xist* during the maintenance of X-inactivation in somatic cells. One study noted that the XIC (including the *Xist* gene locus) was not strictly required for maintenance of transcriptional inactivity, as X-linked genes measured remained transcriptionally silent (Brown and Willard, 1994). Another study, however, noted that whilst loss of *Xist* did not lead to X-linked gene reactivation, it did lead to a fall in macroH2A enrichment at the Xi (Csankovszki et al., 1999). Follow up experimentation from the same researchers observed that loss of *Xist* led to varied re-activation of the transgenes measured, and that *Xist* acts synergistically with DNA methylation and hypoacetylated histones to maintain gene repression (Csankovszki et al., 2001). The authors acknowledge that the transgenes might not be behaving in the same way as endogenous genes, which could provide a rationale for these variable results. A recent review (Loda et al., 2022), highlights that *Xist* requirement for maintenance of the Xi is variable across cell types, with some being more prone to Xi reactivation upon *Xist* loss, thus exact role of *Xist* in the maintenance of X-inactivation is still unknown.

Since there are many different epigenetic features at the Xi, including (but not limited to): *Xist* expression, DNA CpG methylation, reduction of active histone PTMs such as H3K4me3 and histone acetylation, elevation of repressive histone PTMs such as H3K27me3 and H3K9me3, and an enrichment in H1, macroH2A and heterochromatin protein 1 (HP1) (Chadwick and Willard, 2003), this suggests that there are many redundant processes that are implemented to ensure maintenance of Xi gene repression. Indeed, gene silencing at the Xi is incredibly stable, and chromosome-wide reactivation of genes on the Xi has not been achieved unless the cell is reprogrammed to an early developmental stage, thus suggesting that all these epigenetic modifications work in tandem to maintain the gene repression at the Xi (Wutz, 2011).

1.5.7. CIZ1 involvement in X-inactivation

CIZ1 was also identified in one of the proteomic screens for *Xist* interactors, alongside 80 other proteins (Chu et al., 2015). CIZ1 is enriched at the Xi in female mammalian cells and is recruited by *Xist* via interactions with repeat E in exon 7. When CIZ1 is lost, *Xist* is dispersed across the nucleoplasm rather than targeted to the Xi (Ridings-Figueroa et al., 2017, Sunwoo et al., 2017). Follow up work confirmed that the PLD regions of CIZ1 interact with RNA and can support stable and direct interaction with *Xist*, with preference for its repeat E motifs (Sofi et al., 2022). Both PLD1 and PLD2 were required for efficient formation of *de novo* CIZ1-Xi assemblies in CIZ1 null PEFs.

It has been suggested that *Xist*:protein complexes phase separate to form condensates at the Xi (Pandya-Jones et al., 2020), and the *Xist* E repeat region is required to seed these interactions. Condensates are membrane-less compartments generated through liquid-liquid phase separation (LLPS), and examples include nucleoli, stress granules and nuclear speckles (Strom and Brangwynne, 2019). The literature generally describes LLPS condensates as spherical entities, similar to those generated in oil and water demixing. However in our studies, while purified CIZ1 fragments self-assemble into higher order complexes in a manner dependent on its PLD domains, they form an irregular fibrillar network influenced by inclusion of RNA (Sofi et al., 2022).

1.5.8. Consequences of loss of CIZ1 at the Xi

Murine CIZ1 knockout (KO) embryos showed evidence of de-regulation of a subset of X-linked genes, but there were no overt defects in embryogenesis. However, all the females developed lymphoproliferative disorders with enlarged primary and secondary lymphoid tissues compared to the males (Ridings-Figueroa et al., 2017). As introduced above, follow up work confirmed that CIZ1 was an essential component in the maintenance of specific epigenetic marks at the Xi (H3K27me3 and H2AK119ub1), as normal differentiated cells pass though the cell cycle (Stewart et al., 2019). It was also observed that introduction of ectopic CIZ1 that lacked PLD1 or PLD2, so was unable to localise to the Xi, failed to form H3K27me3 and H2AK119ub1 enriched chromatin at the Xi (Sofi et al., 2022).

Additionally, CIZ1 was found to facilitate movement of the Xi towards the nucleolus during S phase, as this inward movement was significantly impeded in CIZ1 null cells (Stewart et al., 2019). This suggests that mechanistically CIZ1 appears to support translocation of replicating chromatin from one location in the nucleus to another. Thus, CIZ1 appears to be part of an epigenetic fidelity mechanism that couples DNA replication to chromatin maintenance, a process that could underpin all of the disease observations associated with CIZ1. However, at this stage it is not known what regions or domains of CIZ1 are responsible for this.

Since we can measure CIZ1 behaviour at the Xi upon response to stimuli, this is frequently used by the Coverley lab as a model for CIZ1 behaviour at other locations in the genome, and this line of experimentation was utilised throughout my studies to test the effects of overexpression of mutant forms of CIZ1.

1.6. CIZ1 in disease

1.6.1. CIZ1 alternative splicing events

More than twenty splice variants of CIZ1 have been identified (Rahman et al., 2010), but their possible links with human pathology remain unclear in most cases. Some are normal splice variants seen in development such as eCIZ1 (Coverley et al., 2005), but others have only been seen in disease states. A cancer associated CIZ1 variant lacking exon 4 maintained the ability to replicate DNA, but was not able to form the correct replication foci normally observed in the nucleus, and when co-expressed with full length CIZ1 it had a dominant negative (DN) effect on endogenous foci (Rahman et al., 2007).

The CIZF variant is produced by an alternative splicing event leading to the loss of exon 8-12, with a subsequent inclusion of a small alternative reading frame (ARF) at the C-terminus (Swarts et al., 2018). This produces a CIZ1 polypeptide that contains the first 275 amino acids of CIZ1 including PLD1, the N-terminal NLS and some cyclin interaction motifs, but lacking all downstream domains. CIZ1F is overexpressed in early-stage breast and colon cancers and is a potential biomarker for stratification.

Finally, the CIZ1B splice variant involves an in-frame exclusion of 24 nucleotides leading to loss of eight amino acids in the acidic domain of CIZ1 (Higgins et al., 2012). Since CIZ1B depletion via RNAi led to reduced tumour proliferation, it has been identified as a possible cancer driver. A fragment of the CIZ1B variant can be measured in the plasma of lung cancer patients and is currently under development for use in the clinic as a biomarker for early-stage lung cancer diagnosis (Coverley et al., 2017). A diagram of the structure of these CIZ1 splice variants is shown below (Figure 1.6).
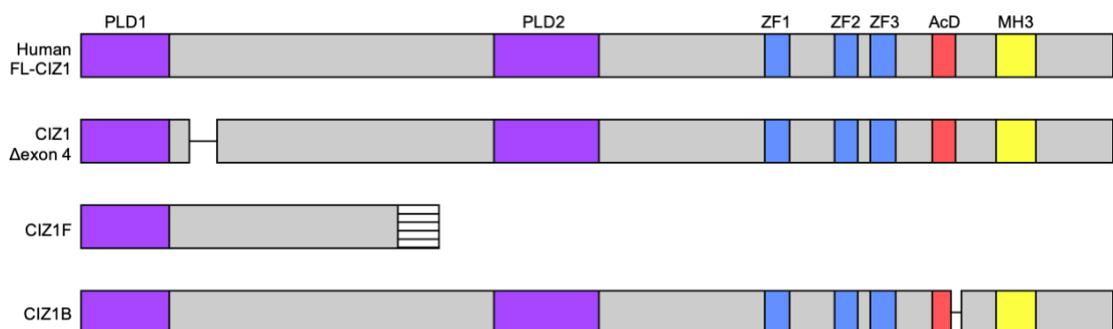
Figure 1.6. CIZ1 splice variants associated with cancer
Full length human CIZ1 (UniProt Q9ULV3). Purple: prion like domain 1 and 2 (PLD1) and (PLD2) as documented in (Sofi et al., 2022). Blue: C2H2 zinc finger 1-3 (ZF 1-3). (ZnF_C2H2 SM00355, ZF_C2H2 sd00020 and ZF_C2H2 sd00020 respectively). Red: the acidic domain (AcD). Yellow: matrin 3-homology (MH3) domain (ZnF_U1 smart00451). The striped region in CIZ1F illustrates the ARF created by the alternative splicing event. CIZ1 Δexon 4, CIZ1F and CIZ1B have been associated with Ewing tumour, breast and colon cancer and lung cancer respectively.

1.6.2. CIZ1 over-expression and under-expression

Aberrant expression of CIZ1 has been implicated in several other cancers, as compiled below in Table 1.1. Most commonly this is described as CIZ1 overexpression, though often the detection tools employed do not interrogate alternative splicing events, and for some papers cannot be traced. Reducing the overexpression of CIZ1 via siRNA knockdown has been shown to reduce cell proliferation (Yin et al., 2013, Wu et al., 2016). However, CIZ1 has also been identified as a tumour suppressor, where two studies showed that CIZ1 KO results in leukaemias and lymphomas in mice (Nishibe et al., 2013, Ridings-Figueroa et al., 2017). Additionally CIZ1 KO has been associated with motor abnormalities in younger mice (Xiao et al., 2016), and mouse embryonic fibroblasts (MEFs) from older CIZ1 KO mice displayed heightened sensitivity to the effects of gamma-irradiation displaying persistent DNA breaks, aberrant cell-cycle progression, and apoptosis (Khan et al., 2018). Overall, this collectively suggests that any shift from the normal expression of CIZ1 is associated with negative effects, though complete loss of CIZ1 is not lethal.

More recent data from our lab has shown that the C-terminal part of CIZ1 is over-represented in most common solid tumours. This is discussed alongside my data in Chapter 3, which explores the functional implications of misbalanced CIZ1 expression.

Table 1.1. Aberrant CIZ1 expression and pathology

Pathology	CIZ1 Variation	Experimental Model	Implication
Alzheimer's disease	Alternative splicing (AS)	Human tissues and cell lines and transfections in murine cells	Inability to form nuclear foci and associate with the nuclear matrix (Dahmcke et al., 2008)
Brain tumours	Highly expressed	Human tissues and murine xenografts	CIZ1 potentially implicated in tumorigenic phenotype (Warder and Keherly, 2003)
Breast cancer	Over-expression	Transfections in human cell lines and xenografts in murine cells	Overexpression promotes growth-rate and anchorage-independence, and leads to oestrogen sensitivity (den Hollander et al., 2006)
Cervical dystonia	Mutation driven alterative splicing	Human sample screening and transfections in murine cells	Altered the nuclear localisation (Xiao et al., 2012)
Colorectal cancer	Highly expressed	Human cell lines; Human matched normal and cancer samples	siRNA silencing suppressed RKO cell proliferation (Yin et al., 2013). Positive CIZ1 expression leads to lower five-year disease-free survival and increased recurrence (Wang et al., 2014)
Ewing Sarcoma	Mutation driven AS	Human cell lines and transfections in murine cells	Inappropriate localisation of CIZ1 (Rahman et al., 2007)
Gallbladder cancer	Highly expressed	Human matched normal and cancer samples and human cell lines	CIZ1 overexpression promoted growth and migration, while knocking down inhibited growth, migration and tumorigenesis (Zhang et al., 2015)
Haemangioma	Highly expressed	Human normal and haemangioma samples and a human cell line	Increased the proliferation and migration (subsequently reduced by shRNA transfection) (Wang et al., 2019)
Hepato-cellular carcinoma	Highly expressed	Human matched normal and cancer samples and human cell lines	Forced expression of CIZ1 promoted the growth and migration. RNAi knockdown inhibited the growth, migration and metastasis (Wu et al., 2016)
Leukaemia	CIZ1 null	Murine cells and mice	Increased sensitivity to hydroxyurea-mediated replication stress (Nishibe et al., 2013)
Lung cancer	AS, highly expressed	Human matched normal and cancer samples, human and murine cell lines; Human matched normal and cancer samples	RNAi depletion of CIZ1B variant restrained growth of tumour cells (Higgins et al., 2012). Specific over-expression in LSCC tissues could indicate contribution to the growth and angiogenesis of LSCC (Zhou et al., 2018)
Neuro-degeneration	CIZ1 null	Murine cells and mice	Deficits in motor and cognitive functioning, overt DNA damage, NF-kB upregulation, oxidative stress, vascular dysfunction, inflammation and cell death (Khan et al., 2018)
Prostate cancer	Highly expressed	Human tissue samples, human cell lines and xenografts in murine cells	Reduced cell proliferation and colony formation, induced cell cycle arrest in G1 and inhibited tumour formation in nude mice after CIZ1 was silenced (Liu et al., 2015a)
Solid tumours	Alternative splicing	Human cell lines and human tissue samples	Does not accumulate at the Xi (Swarts et al., 2018)
Uveal Melanoma	Under expressed, highly expressed & AS	Human bioinformatics information from TCGA	Lower relative expression correlates to higher risk subtypes and decreased survival (Kucherlapati, 2018)

1.7. Project Aims

The two main focuses of my thesis were to further characterise the structure and function of two cancer associated CIZ1 variants, to attempt to decipher what their roles are in tumourigenesis. The hypothesis was that these forms could be exerting DN effects, as observed with other CIZ1 forms, but this had not been tested. Individual objectives of the project included:

- Measuring the expression of *CIZ1* 3' transcripts relative to 5' levels, to quantify the overexpression occurring in the cancer genome atlas (TCGA) samples of breast tumours compared to normal samples (Chapter 3)
- Identify if there are cancer specific CIZ1 C-terminal fragments in breast cancer cell lines compared to normal breast tissue (Chapter 3)
- Profiling the downstream consequences of CIZ1 C-terminal fragment expression using a transient transfection and lentiviral transduction system in a murine experimental model (Chapter 3 and 4)
- Conducting biochemical analysis of the native state of C-terminal fragments and potential changes upon individual domain deletion (Chapter 3)
- Monitoring CIZ1 localisation in different stages of the cell cycle to identify mechanisms behind CIZ1 regulation that could be dysregulated in disease (Chapter 4)
- Investigating any potential changes in structure, interaction partners or *in vivo* function of CIZ1 and CIZ1B (Chapter 5)

2. Materials and Methods

2.1. Cells

2.1.1. Generation of mouse cell lines

All mouse cells were generated by Dr Justin Ainscough. All mouse primary embryonic fibroblasts (PEFs) strains were derived from day 13 or 14 embryos from C57BL/6 mice. Cells were cultured up to a maximum of passage 4, after passage 4 these cells are referred to as mouse embryonic fibroblasts (MEFs) and were not used here. CIZ1 null mice were generated from C57BL/6 ES clone IST13830B6 (TIGM) harbouring a neomycin resistance gene trap inserted downstream of exon 1. Confirmation of CIZ1 status (CIZ1$^{+/+}$, CIZ1$^{+/-}$, CIZ1$^{-/-}$) and sex of murine cells for selection of appropriate cells for use in experimentation was conducted by Dr Justin Ainscough and Dr Emma Stewart via qPCR, immunofluorescence and immunoblot.

2.1.2. Cell culture and maintenance

Cells were cultured in media (Table 2.1) at 37°C with 5% CO_2, passaging wherever necessary with trypsin (Gibco, Cat no. 15400-054) to maintain cells in a rapidly cycling state.

Table 2.1. Cell culture media for cell lines used

Cell Line	Media	Supplements
D3T3	HG DMEM (Gibco, Cat no. 31966-021)	10% (v/v) FBS (PAA, Cat no. A15-151), 1% (v/v) Pen/Strep/Glutamine (Gibco, Cat no. 10378-016)
PEFs	HG DMEM (Gibco, Cat no. 31966-021)	10% (v/v) FBS (PAA, Cat no. A15-151), 1% (v/v) Pen/Strep/Glutamine (Gibco, Cat no. 10378-016)
HMEC	MEBM™ Basal Medium (Lonza, Cat no. CC-3150)	MEGM™ SingleQuots™ Supplements (Lonza, Cat no. CC-4136)

2.1.3. Transient transfection

Cells were seeded on glass coverslips (SLS, Cat no. MIC3306) at approximately 30% confluency one day prior to transfection to produce a population at ~60% confluency at time of transfection. Complexes were prepared and incubated for 30 minutes prior to application on cells. The composition of the complex per coverslip was: 50 μL Opti-MEM® Medium (Gibco, Cat no. 31985-062), 1.5 μL X2 Transfection Reagent (Mirus, Cat no. MIR 6003) and 200-500ng DNA plasmid (200ng in the C-terminal CIZ1 fragment experiments and 500ng in the full length CIZ1 experiments). This was applied to cells dropwise and left to proceed for 24 hours prior to fixation and imaging. Details of plasmids used in experimentation are listed Table 2.4 in section 2.4.1.

2.1.4. Drugs and inhibitors

Drugs and inhibitors used in experimentation are listed below in Table 2.2. Thymidine was applied for 24 hours, and nocodazole was applied for 16-24 hours. Cells arrested in M phase were isolated by mitotic shake off and replated for analysis post release. Cells held in S phase were released by washing twice with PBS then replacing with fresh media. Two different concentrations of barasertib were applied to cells for 4 hours and cells were subsequently collected for immunofluorescence imaging. PR-619 was applied to cells during the transient transfection window.

Table 2.2. Drugs used in cell experimentation

Name	Role/target	Concentration	Catalogue reference
Thymidine	Hold in S phase	2.5 mM	Sigma, T1895
Nocodazole	Hold in M phase	50 ng/mL	Sigma, M1404
Barasertib	AURKB inhibitor	0.1 and 1 µM	Selleck Chemicals, AZD1152-HQPA
PR-619	DUB inhibitor	5 µM	Bio-Techne, 4482/10

2.1.5. Flow cytometry

Cells were trypsinised then neutralised in media and counted with a haemocytometer. Cells were then centrifuged at 145 xg for 5 minutes and resuspended in PBS to obtain a concentration of 500,000 cells/mL. 500 µL of cells was mixed with 55 µL 10x FACS mix (10 mg/mL propidium iodide, 20% Triton X-100, 10x PBS). Analysis was performed by Karen Hogg in the Imaging and Cytometry department of the Bioscience Technology Facility at the University of York. DNA content was measured by flow cytometry, CytoFLEX (Beckman Coulter) using the DNA binding dye propidium iodide (excitation 561nm, emission 610/20). Fluorescence PI signal was used to exclude doublets and aggregates using the height versus area parameters. The event rate was 30 µL per minute: a minimum of 5,000 single cells per sample were recorded for analysis using cell cycle algorithm software FCS Express V7 (Dotmatics) (Dean and Jett, 1974).

2.1.6. Lentivirus transduction

Protocol provided by Dr Emma Stewart in the Coverley lab. 8×10^5 HEK cells were seeded per well in a 6 well plate prior to transfection. Plasmids (1µg transfer vector containing the CIZ1 entity, 0.75µg packaging plasmid, 0.25µg envelope plasmid) were diluted in 100µl optiMEM (Gibco, Cat no. 31985-062). 20µl of PolyFect transfection reagent (Qiagen, Cat no. 301105) was then added to the DNA solution and mixed by pipetting up and down 5 times, or by vortexing for 10 sec. This was incubated for 5-10 min at room temperature to allow complex formation. While the complex formation took place, growth medium was gently aspirated from the dish and fresh cell growth medium added. 0.6 mL of cell growth medium was added to the reaction tube containing the transfection complexes, gently mixed by pipetting up and down twice, and the total volume immediately transferred to the cells in the plate. The cell culture dish was gently swirled to ensure uniform distribution of the complexes. Complexes were incubated with cells overnight, then the media was replaced with 3 mL fresh growth medium (supplemented with addition of 1M HEPES at pH 7.2-7.5 (Cat no. 15630-056) to a final concentration of 20mM). At 48 hours post transfection the viral supernatant was harvested for transduction of cells (can also replace media and harvest at 72 hours if needed). Alongside the initial transduction of the HEK cells, PEFs to be used in experimentation were plated appropriately, for example cells to be transduced for microscopy analysis were seeded onto glass coverslips prior to transduction. The viral supernatant was filtered to remove HEK debris by using a low-protein binding filter (0.45µm, Sarstedt, Cat no. 83.1826). Viral supernatant was placed on the cells to be transduced supplemented with final 4 µg/mL polybrene (Sigma, TR-1003-G). This was incubated overnight, then fresh media (without addition of HEPES) was added to the target cells. Transduction occurred after approximately 48 hours, at this point cells could be harvested or kept for long term culture as required.

2.2. Microscopy

2.2.1 Immunofluorescence

Cells received an initial PBS wash to remove excess media. Cells were then permeabilised in CSK-D (CSK buffer containing 0.05% Triton X-100) for 1 minute, and fixed for 15 minutes in 4% paraformaldehyde. After fixation, cells were rinsed twice with PBS, then incubated at room temperature for at least 30 minutes in BSA antibody buffer. Coverslips were then incubated for 1 hour at 37°C in a humidified chamber with BSA antibody buffer containing primary antibody (all primary antibodies used are listed in Table 2.14 in section 2.7). After incubation cells were washed three times in BSA antibody buffer, and remounted in the humidified chamber for an incubation with secondary antibody for 1 hour. Finally, cells were washed three times in BSA antibody buffer and mounted in VECTASHIELD® Antifade Mounting Medium with DAPI (Vector Laboratories, Cat no. H-1200). Buffer compositions are shown below in Table 2.3.

Table 2.3. Buffers used in immunofluorescence

Buffer	Composition
Cytoskeletal (CSK) buffer	10 mM 1,4-piperazinediethanesulfonic acid (PIPES)/KOH, pH 6.8 (Sigma, Cat no. P3768) 100 mM sodium chloride (NaCl) (Sigma, Cat no.71380-M) 300 mM sucrose (Sigma, Cat no. 84097) 1 mM ethylene-bis(oxyethylenenitrilo)tetraacetic acid (EGTA) 1 mM magnesium chloride ($MgCl_2$) (Sigma, Cat no. M2393)
BSA antibody buffer	1x PBS (Gibco, Cat no. 14190144) 1x Detergent mix (0.02% SDS + 0.1% Triton X-100) 10 mg/mL BSA (Sigma, Cat no. A3294)
1x PBS (Dulbecco's PBS -$CaCl_2$ -$MgCl_2$) (Gibco, Cat no. 14190144), pH 7.0-7.3	137 mM NaCl 8.1 mM Na_2HPO_4 2.7 mM KCl 1.5 mM KH_2PO_4

2.2.2. Imaging and image enhancement

Fluorescence images were captured using a Zeiss Axiovert 200M fitted with a 63X/1.40 Plan-Apochromat objective and Zeiss filter sets 2, 10, 15 (G365 FT395 LP420, BP450-490 FT510 BP515-565, BP546/12 FT580 LP590), using Axiovision image acquisition software (SE64 release 4.9.1). If images were to be used for fluorescence intensity analysis all parameters were kept consistent in the imaging process. Photo enhancement for visualisation was performed in FIJI after intensity measurements were conducted.

2.2.3. Phenotype Scoring in Dispersal Assay

The status of endogenous CIZ1 was monitored after CIZ1 C-terminal fragment transfection or transduction to measure CIZ1 dominant negative effects. Cells were inspected by eye directly on samples, and avoidance of bias was achieved by independent and blinded analysis. In the two-tier scoring system of ectopic DNA effect, cells were grouped into two categories, either having a CIZ1 Xi patch or no CIZ1 Xi patch. This was further optimised to a three-tier scoring system with inclusion of an intermediate category (dispersed/diffuse/reduced). An untreated or negative control (empty vector) and positive control WT-C181 was used, respectively, to determine the untreated and treated CIZ1 Xi frequency baseline in different cell batches. Notably some cells in the population are observed to contain two CIZ1-Xi accumulations, indicative of a X chromosome duplication. Where this occurred, cells were scored as though they contained one Xi, since in all cases both Xi phenotypes contained in one nuclei were the same. These genetic alterations are a common feature of cell lines and indicates the requirement for additional experimentation on primary cell lines wherever possible to strengthen results.

2.2.4. Fluorescence intensity measurements FIJI

Open tiff image

Image > Colour > Split channels

Select the DAPI channel

Image > Adjust > Threshold > Otsu. This should create an outline around the outer limit of the DAPI stain, adjustments can be made if needed. > Apply

Process > Binary > Convert to mask

Analyse > Analyse particles > change size to 0.01-Infinity, ensure "Display results" and "add to Manager" are selected > OK

Select red image Image > Overlay > Show overlay. This will overlay the mask from the blue channel. Select the nucleus to measure > Analyse > Measure

Results will appear in results window with intensity measurements. Keeping the green channel open alongside this allows identification of untransfected and transfected cells.

2.3 Statistical analysis and data visualisation

A variable number of biological replicates, technical replicates and independent counts (N value), was conducted in experimentation, allowing generation of error bars depicting ± SEM. The breakdown of the N value in each experiment is provided in the figure legend. The number of cells that were measured varied, and this is stated individually in each experiment (n value). Wherever possible two independent PEF lines were used in experimentation. Statistical analysis was carried out in SPSS (IBM Corp. Released 2021. IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp) or Microsoft Excel, using a two-sample unpaired t-test, or one-way ANOVA followed by an appropriate post-hoc test. The statistical test used for each experiment is stated in the figure legend alongside the individual significance values. Graphs were generated using Microsoft Excel, for illustrative purposes asterisks indicate statistical significance (*=$p<0.05$, **=$p<0.01$ and ***=$p<0.001$). Where data is represented in a box and whisker plot, the box represents the middle 50% of the data (the boundaries are quartile 1 (Q1) and quartile 3 (Q3), which are the cut off of the lowest 25% and the highest 25% of the data respectively). The line across the middle of the box represents the medium, and the X in the box corresponds to the mean. The value obtained when subtracting Q1 from Q3 is referred to as the interquartile range (IQR). The whisker above the box is Q3 + (1.5 x IQR), and the whisker below the plot is Q1 – (1.5 x IQR). Outliers are shown as individual data points, where all outliers have not been illustrated by excel due to size constraints, the number of outliers is stated in the figure legend. Data points that fall within the box and whiskers are not shown.

2.4. DNA and cloning

2.4.1. Plasmids

All plasmids used in experimentation are listed below in Table 2.4., with vector maps shown in Figure 2.1. detailing the restriction enzymes used to insert the CIZ1 fragment into the plasmid. Information on the restriction enzymes used cannot be provided for GST-H720 as it was produced by and purchased from Fujirebio Diagnostics AB. The plasmids used in experimentation were cloned into the following bacterial growth stains: GFP transient transfections – DH5α (Invitrogen, Cat no. 18265017), recombinant protein – BL21 (Agilent Technologies, Cat no. 230250) and lentivirus transductions – Stbl3 (Invitrogen, Cat no. C737303).

Table 2.4. Plasmids used in experimentation

Experimental series	Plasmid description	Vector backbone	Identifier
GFP transient transfections	GFP full length (FL) CIZ1 series (Chapter 5)	pEGFP-C3 (Higgins et al., 2012)	Addgene plasmid #6082-1
GFP transient transfections	GFP-181 CIZ1 series (Chapter 3)	pEGFP-C2	Addgene plasmid #6083-1
Recombinant protein	GST-275/181 CIZ1 series (Chapter 3)	pGEX-6P-3 (Sofi et al., 2022)	Cytiva, Cat no. 28-9546-51
Recombinant protein	GST-H720 CIZ1 series (Chapter 5)	pGEX-6P-1	Cytiva, Cat no. 28-9546-48
Lentivirus transductions	Packaging plasmid	psPAX2	Addgene plasmid #12260
Lentivirus transductions	Envelope plasmid	pMD2.G	Addgene plasmid #12259
Lentivirus transductions	ZsGreen C181 CIZ1	pLVX-EF1α-IRES-ZsGreen1	Takara, Cat no. 631982

GFP transient transfection plasmids
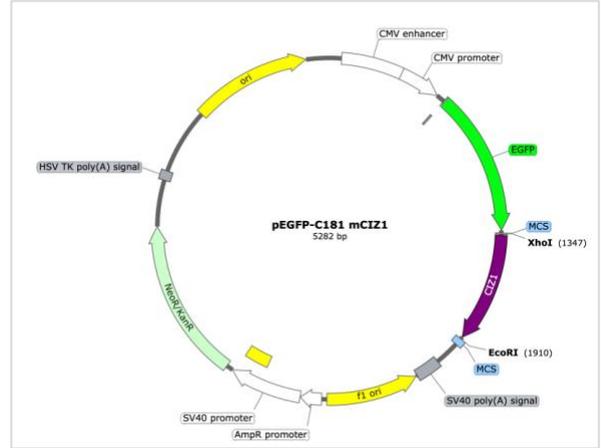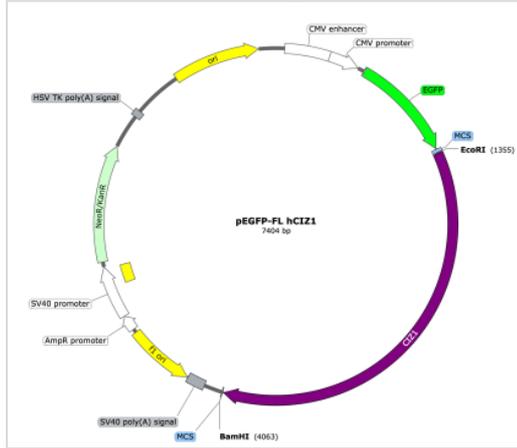
GST tagged recombinant protein plasmids
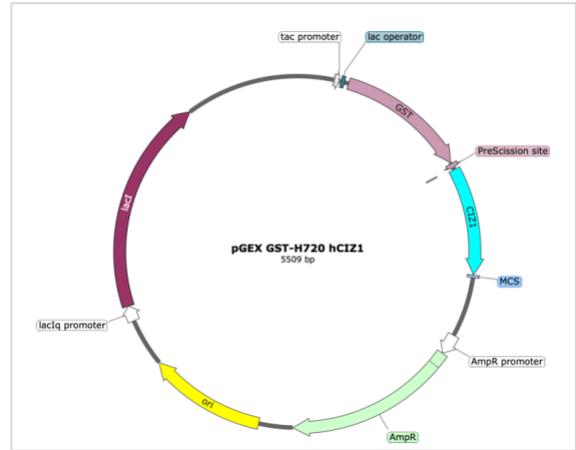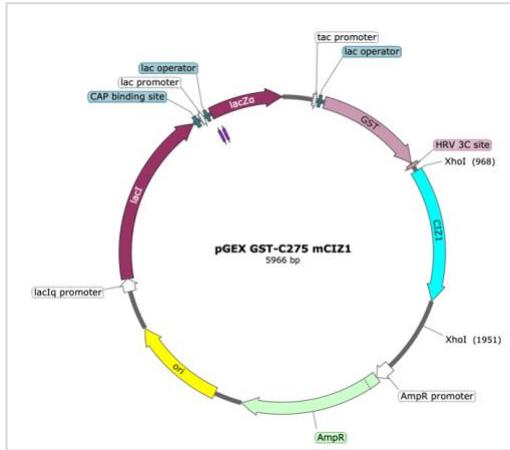
Lentivirus transduction plasmid
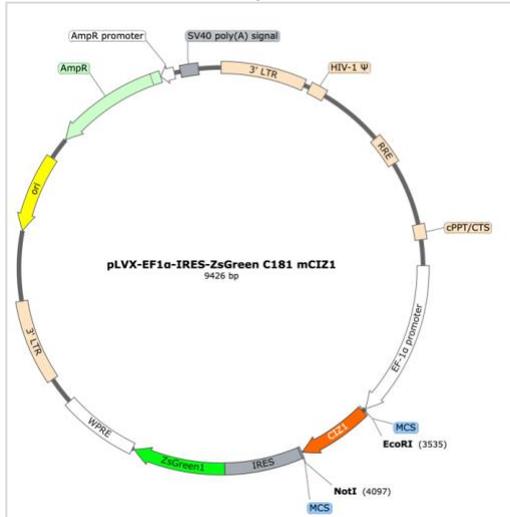
Figure 2.1. Vector maps of CIZ1 plasmids used in experimentation

2.4.2. Site directed mutagenesis

Mutagenic primers that contain additions, substitutions or deletions of regions of CIZ1 were created for use in a PCR mutagenesis using a protocol provided by Dr Julie Tucker at the University of York. The primers used are listed in Table 2.5, where FP corresponds to forward primer and RP corresponds to reverse primer. The PCR reaction was created to a total reaction volume of 10 μL, as shown in Table 2.6, and ran for 18 amplification cyles under conditions listed in Table 2.7. All primers were purchased at 0.025μmole (Merck) and 1 mM stock solutions prepared in TE buffer (10 mM Tris, 1 mM EDTA) and stored at -20°C. The I122 construct was a pre-existing plasmid created in the Coverley lab by Dr Justin Ainscough via restriction digest and religation using EcoRI and BamHI (Ainscough et al., 2007).

Table 2.5. Primers used in mutagenesis

Template	Primer sequence, * reflects boundary of complementarity to regions upstream and downstream of mutation	Plasmid created
GFP-C181	FP:GGAGAGATTGA*GGTGAAGCCGAGAGAAACATCC RP:GGATGTTTCTCTCGGCTTCACC*TCAATCTCTCC	GFP-C181 Δ706-713 (CIZ1B)
GFP-C181	FP:GGATTTCCTGGTGCCAGTGATG*AAAGCCAAGAACCCAAGC RP:GCTTGGGTTCTTGGCTTT*CATCACTGGCACCAGGAAATCC	GFP-C181 Δ746-779 (ΔMH3)
GFP-C181	FP:CCTGACTGCACTGTTC*TGATAGAAGCTTCGAATTCTGC RP:GCAGAATTCGAAGCTTCTATCA*GAACAGTGCAGTCAGG	GFP-C181 Δ809-845 (Δ37)
GFP-C181	FP:GCCCTCCTCCTACC*AGCCACCAGCCCAGCC RP:GGCTGGGCTGGTGGCT*GGTAGGAGGAGGGC	GFP-C181 Δ796-811 (ΔNALTAF)
GST-C275	FP:ATCCCCGAATTCCCGGGTCGAC*AAGGAGACAGGCAGCCC RP:GGGCTGCCTGTCTCCTT*GTCGACCCGGGAATTCGGGGAT	GST-C181
GST-C181	FP:GCTTTGAGAGTGGTCAA*TTCTGCAAGCAGGTGAAGC RP:GCTTCACCTGCTTGCAGAA*TTGACCACTCTCAAAGC	GST-C181 Δ689-709 (ΔAcD)
GST-C181	FP:GGATTTCCTGGTGCCAGTGATG*AAAGCCAAGAACCCAAGC RP:GCTTGGGTTCTTGGCTTT*CATCACTGGCACCAGGAAATCC	GST- C181 Δ746-779 (ΔMH3)
GST-C181	FP:CCTGACTGCACTGTTC*TGATAGAGGGAGC RP:GCTCCCTCTATCA*GAACAGTGCAGTCAGG	GST-C181 Δ809-845 (Δ37)
GST-C181 Δ746-779 (ΔMH3)	FP:GCTTTGAGAGTGGTCAA*TTCTGCAAGCAGGTGAAGC RP:GCTTCACCTGCTTGCAGAA*TTGACCACTCTCAAAGC	GST-C181 Δ689-709, 746-779 (ΔAcD,MH3)
GST-C181 Δ746-779 (ΔMH3)	FP:CCTGACTGCACTGTTC*TGATAGAGGGAGC RP:GCTCCCTCTATCA*GAACAGTGCAGTCAGG	GST-C181 Δ746-779, 809-845 (ΔMH3,37)

Table 2.6. PCR reaction

Component	Volume (µL)	Concentration/Amount
Template (5 ng/µL)	0.8	4ng
Primer pair mix (Sigma) (15 ng/µL)	1	10-20ng
2x CloneAmp HiFi PCR premix (Takara, Cat no. 639298)	5	1x
ddH$_2$O	3.2	-

Table 2.7. PCR cycling conditions

Stage	Temperature (°C)	Time (s)
Initial denaturation	98	30
Amplification cycle (1)	98	10
Amplification cycle (2)	55	30
Amplification cycle (3)	72	15 secs per kb
Final extension	72	10

2.4.3. Gel electrophoresis

Confirmation of mutagenic PCR reactions were carried out on a 1% (w/v) agarose in 1x TBE (89 mM tris, 89 mM boric acid, 2 mM EDTA) gel containing 1x SYBR™ Safe DNA gel stain (Invitrogen™, Cat no. S33102) using 5 µL of the reaction with 1 µL of loading dye (New England Biolabs, Cat no. B7025S). Agarose gel was electrophoresed at 70V until resolution of bands required was reached. Unedited methylated DNA template remaining was digested for 1 hour at 37°C using 2 units of DpnI restriction enzyme (New England Biolabs, Cat no. R0176S) in CutSmart® Buffer (New England Biolabs, Cat no. B7204S) as per manufacturer's recommendation. Gels were visualised using a transilluminator in the PXi gel imaging system (Syngene).

2.4.4. Bacterial transformation

1 µL of plasmid PCR product was incubated with 25 µL of DH5α competent cells (Invitrogen, Cat no. 18265017) for 30 minutes on ice, then underwent heat shock at 42°C for 45 seconds, and finally incubated on ice for 2 minutes. This was allowed to recover for 1-2 hours, depending on the antibiotic used, in 225 µL SOC broth (Invitrogen™, Cat no. 15544034) at 37°C in an orbital shaker at 200 rpm. Bacteria were then plated onto antibiotic selective LB agar at varying concentrations for optimum colony production and incubated overnight at 37°C.

2.4.5. DNA extraction and sequencing

Individual colonies were selected and grown in 5 mL of LB supplemented with selective antibiotics overnight. The following morning plasmid DNA was extracted using QIAprep Spin Miniprep Kit (QIAGEN, Cat no. 27104), as per manufacturer's instructions. Glycerol stocks of bacterial cultures in LB supplemented with 15% (v/v) glycerol were also made at this point if required. Plasmid DNA sequencing was undertaken to confirm presence of mutagenesis in house or using Eurofins TubeSeq Service, using standard sequencing primers from Eurofins or using the designed CIZ1 sequencing primer with the sequence AGACAGGCAGCCCAGATGAGG.

2.5. Protein analysis

2.5.1. Protein expression and purification

Varying murine and human CIZ1 constructs with an in-frame N-terminal tag glutathione S-transferase (GST) in pGEX-6P expression plasmids were expressed in BL21-CodonPlus-RP *E. coli* using lactose-driven auto-induction. A starter culture of bacteria expressing protein of interest was incubated in LB broth (Merck, Cat no. 1102850500) containing selective antibiotic overnight at 37°C. Autoinduction media was inoculated with starter culture until the OD_{600} was between 0.1-0.15 and incubated at 20°C for 24 hours. This produced a bacterial pellet after centrifugation (4392 xg, 4°C, 15 minutes) which was resuspended in PBS buffer that also contained 1 in 100 EZBlock™ protease inhibitor cocktail (BioVision, Cat no. K272) and 1 mM PMSF. Bacterial cells were lysed on ice via sonication for 5 cycles (30 sec on, 30 sec off) at 60% amplitude using 6 mm probe (microtip MS 73, Bandelin SONOPULS). Lysates were clarified by centrifugation for (24652 xg, 4°C, 30 minutes) in a Heraeus™ Multifuge™ X1 centrifuge with a F15-6x100y fixed angle rotor. All chromatography was performed on an ÄKTA™ chromatography system. Lysates were loaded onto a 5 mL glutathione sepharose column (Cytiva, Cat no. 28401747) at 0.5 mL/min then washed extensively with 10 column volumes of PBS buffer followed by 10 column volumes of cleavage buffer, both at 1 mL/min. This was incubated at 4°C overnight with 2 units of PreScission protease (GE Healthcare, Cat no. 27-0843-01) in cleavage buffer. Cleaved protein was eluted in fresh cleavage buffer and concentrated to 0.5 mL using a centrifugal concentrator (Sartorius, Cat no. VS0601). Protein concentration and quality was determined by NanoDrop® ND-1000 spectrophotometer (Labtech, version V3.2.1). Purified protein was supplemented to 5% (v/v) glycerol and snap frozen in liquid nitrogen and stored at -80°C. Glutathione Sepharose column was regenerated for future use using 5 column volumes of elution buffer, then stored in 20% ethanol. Buffer compositions are shown below in Table 2.8.

Table 2.8. Buffers used in protein expression

Buffer	Composition
PBS buffer (Sigma, Cat no. P3813), pH 7.4	137 mM NaCl 8.1 mM Na_2HPO_4 2.7 mM KCl 1.5 mM KH_2PO_4
Cleavage buffer	50 mM Tris-HCl (Sigma, Cat no. T6666), pH 7.0 150 mM NaCl (Sigma, Cat no.71380-M) 1 mM EDTA (Invitrogen™, Cat no. AM9261) 1 mM Dithiothreitol (Sigma, Cat no. 43819)
Elution buffer, pH 8	50 mM Tris-HCl (Sigma, Cat no. T6666) 10 mM L-glutathione reduced (Sigma, Cat no. G4251)

2.5.2. Western blot analysis

Samples were denatured in the appropriate volume of 4x loading buffer at 95°C for 5 minutes, vortexed, heated for an additional 5 minutes, and centrifuged prior to loading. Samples were electrophoresed in a 4–15% Mini-PROTEAN® TGX™ precast acrylamide gel (Bio-Rad, Cat no. 4561085) submerged in running buffer at 40V for 15 minutes, then 90V for approximately 1.5 hours, until desired resolution has been achieved. The gel was transferred to a nitrocellulose membrane via a dry transfer method using the iBlot® Gel Transfer Device (Transfer stacks from Invitrogen™, Cat no. IB301002) or a semi-dry transfer method in transfer buffer cell (0.1µM NC, Cytiva Amersham™ Protran™, Cat no. 10600000). The membrane was then blocked for 30 minutes in a blocking solution of 5% bovine serum albumin (BSA) (Sigma, Cat no. A3294) or 10% milk (Marvel dried skimmed milk) in 1x PBS containing 0.1% TWEEN®20 (PBST). Primary antibody was applied overnight in blocking solution at 4°C, and the following day underwent three 10 minute washes in blocking solution. The blot was probed with secondary antibody for one hour in blocking solution at room temperature, then underwent three 10 minute washes in PBST. Proteins were visualised using Pierce™ ECL Western Blotting Substrate (Thermo Scientific™, Cat no. 32109) and processed using the PXi gel imaging system (Syngene). Buffer compositions are shown below in Table 2.9.

Table 2.9. Buffers used in western blotting

Buffer	Composition
Loading buffer (final composition)	75 mM Tris, pH 6.8 (Sigma, Cat no. T1503) 4% (w/v) SDS (Sigma, Cat no. L3771) 15% (v/v) glycerol 250mM β-mercaptoethanol (Sigma, M6250)
Running buffer, pH 8.3	25 mM Tris (Sigma, Cat no. T1503) 192 mM Glycine (Sigma, Cat no. G8898) 0.1% (w/v) SDS (Sigma, Cat no. L3771)
Transfer buffer	300 mM Tris (Sigma, Cat no. T1503) 10 mM CAPS (Sigma, Cat no. C2632) 0.02% (w/v) SDS (Sigma, Cat no. L3771) 10% (v/v) Methanol (Sigma, Cat no. 32213-M)
PBST	1x PBS buffer (Sigma, Cat no. P3813) 0.1% TWEEN® 20 (Sigma, Cat no. P1379)

2.5.3. Size exclusion chromatography (SEC)

Experiments were conducted at room temperature on an ÄKTApurifier™ system and the solvent was 0.2µm filtered before use. Samples were 0.45µm filtered prior to injection (Corning ®, Cat no. CLS8163). The column was equilibrated with 2 column volumes of running buffer before use. Blank buffer injections were used where appropriate to check for carry-over between sample runs. A UV_{280} detector was used to detect protein presence. Samples were collected on the Frac-950 fraction collection system. Data was analysed using the UNICORN™ control software. Running conditions are shown below in Table 2.10.

Table 2.10. Running conditions for SEC

Column	Superdex 200 increase 10/300 GL (Cat no. 28990944)
Flow rate	0.5 mL/min
Injection volume	500 µL
Running buffer	150 mM NaCl, 50 mM Tris, 1 mM DTT, 0.2 µm filtered, pH 7

2.5.4. Size exclusion chromatography multi-angle laser light scattering (SEC-MALLs)

Analysis was performed by Dr Andrew Leech in the Molecular Interactions department of the Bioscience Technology Facility at the University of York. Experiments were conducted at room temperature on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Solvent was 0.2µm filtered before use and a further 0.1µm filter was present in the flow path. The column was equilibrated with at least 2 column volumes of solvent before use and flow was continued at the working flow rate until baselines for UV, light scattering and refractive index detectors were all stable. Shimadzu LC Solutions software was used to control the HPLC and Astra V software for the HELEOS-II and rEX detectors. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data was analysed using the Astra V software. Molecular weights were estimated using the Zimm fit method with degree 1. A value of 0.19 was used for protein refractive index increment (dn/dc). Running conditions are shown below in Table 2.11.

Table 2.11. Running conditions for SEC-MALLS

Column	Superdex 200 10/300 GL (Cat no. 17517501)
Flow rate	0.5 mL/min
Injection volume	100 µL
Running buffer	150 mM NaCl, 50 mM Tris, 1 mM EDTA, 1 mM DTT, 0.2 µm filtered, pH 7

2.6. Interaction studies

2.6.1. Mass spectrometry sample preparation

HeLa cells were scrape harvested in isotonic buffer supplemented with 1mM PMSF and Dounce homogenised. Cytosolic proteins were isolated via centrifugation (1520 xg, 4°C, 5 minutes) and discarded. The remaining pellet was resuspended in an intermediate salt extraction buffer and centrifuged (1520 xg, 4°C, 5 minutes). Intermediate salt extracted supernatant was isolated and the remaining pellet was suspended in a high salt extraction buffer and centrifuged (18620 xg, 4°C, 5 minutes). High salt extracted supernatant was isolated and the pellet discarded. The intermediate and high salt supernatants were pooled and diluted with the appropriate volume of hypotonic buffer to obtain 135 mM final NaCl concentration HeLa protein pool. Alongside this, two different recombinant protein forms of CIZ1 (H720 and a mutant form referred to as H720B, described in detail in Chapter 5) and a negative control of GST tag only, were separately expressed and purified using glutathione sepharose beads (incubated at 7 rpm, 4°C, 1 hour on a Grant Instruments™ 360° Vertical Multi-function Rotator). Since glutathione beads have a high binding capacity, precautions were taken to saturate the binding sites to reduce non-specific binding that could occur at later stages. Each recombinant protein-glutathione slurry was extensively washed in hypotonic buffer then incubated with the isolated HeLa protein (7 rpm, 4°C, 2 hours). The slurry then underwent follow-up washes in isotonic buffer supplemented with 0.05% Triton X-100 and then isotonic buffer alone. Samples were separated into four technical replicates for analysis. Buffer compositions are shown below in Table 2.12.

Table 2.12. Buffers used to generate samples for mass spectrometry analysis

Buffer	Composition
Isotonic buffer	10 mM HEPES, pH 7.8 0.5 mM $MgCl_2$ 5 mM K-acetate 135 mM NaCl 1 mM DTT
Intermediate salt extraction buffer	50 mM HEPES, pH 7.8 5 mM $MgCl_2$ 5 mM K-acetate 400 mM NaCl 1 mM DTT
High salt extraction buffer	50 mM HEPES, pH 7.8 5 mM $MgCl_2$ 5 mM K-acetate 800 mM NaCl 1 mM DTT
Hypotonic buffer	50 mM HEPES, pH 7.8 5 mM $MgCl_2$ 5 mM K-acetate 1 mM DTT

2.6.2. Mass spectrometry machine set up

Analysis was performed by Dr Adam Dowle in the Metabolomics & Proteomics department of the Bioscience Technology Facility at the University of York. Proteins were on-bead digested with the addition of Promega sequencing grade modified trypsin and incubation overnight at 37°C. Resulting peptides were extracted and desalted using C18 ZipTip before drying and resuspending in aqueous 0.1% TFA for LC-MS. LC-MS/MS was performed with elution from a 50cm C18 EasyNano PepMap column over 1 hour driven by a Waters mClass UPLC onto an Orbitrap Fusion Tribrid mass spectrometer operated in DDA TopSpeed mode with a 1 second cycle time. MS1 spectra were acquired in the Orbitrap mass analyser at 120K resolution and MS2 spectra were acquired in parallel in the linear ion trap following HCD fragmentation. Resulting LC-MS chromatograms in Thermo .raw format were imported into Progenesis QI for peak picking and alignments. A concatenated MS2 peak list in .mgf format was exported and searched using the Mascot search program against the human subset of the SwissProt proteome appended with common proteomic contaminants. Matched peptides were filtered using the Percolator algorithm to achieve a 1% peptide spectral match false discovery rate, as assessed empirically against a reverse database search. Peptide identifications were imported onto the Progenesis QI-aligned LC-MS chromatograms and matched between acquisitions. Identified MS1 peak areas were integrated and compared for relative peptide quantification of non-conflicting peptide sequences. Relative protein quantification was inferred from underlying peptide values. Final accepted protein quantifications were filtered to require a minimum of two quantified peptides. A multi-way anova was applied to facilitate three way comparison between the pGEX, H720 and H720B samples, to identify proteins of equal abundance in all samples. P-values were multiple-test-corrected to q-values using the Hochberg and Benjamini approach. To correct for differences in total protein in samples between H720 and H720B raw precursor ion areas were normalised based on total ion signal and pairwise analysis was conducted. Further thresholding was conducted, and binding partners were accepted if; $q<0.005$, two unique peptides were identified, and there were at least two spectral counts in one sample.

2.6.3. Electrophoretic mobility shift assay (EMSA)

Protocol provided by Dr Sajad Sofi, a former member of the Coverley lab. RNA probes labelled with digoxigenin (DIG) were provided for use. A master mix of the reaction components was created (final concentration: 2.5 mM $MgCl_2$, 0.1% IGEPAL® CA-630 (FLUKA, Cat no. 56741), 0.1 mg/mL yeast tRNA (Ambion®, Cat no. AM7119), 1U RNaseOUT™ (Invitrogen, Cat no. 10777019)). Mastermix, the desired protein concentration (0-5000nM) and the appropriate volume of RNAse free water was incubated in a 0.2 mL nuclease free thin walled PCR tubes with flat cap (Axygen, Cat no.14222262) at 30°C for 20 minutes. RNA was incubated at 80°C for 3 minutes and cooled on ice for 2 minutes. 1 µL RNA (0.8ng final concentration) was added to each tube, mixed, and incubated at 30°C for 20 minutes for complex formation. 1.3 µL of loading dye was gently added and the sample was electrophoresed in a 0.8% (w/v) agarose in 1x TBE (89 mM tris, 89 mM boric acid, 2 mM EDTA) gel for approximately 1 hour a 4°C. The gel was transferred onto a 0.45µm nylon membrane (GE Healthcare, Cat no. RPN303B) by placing the membrane on top of the gel, followed by blotting papers (rolling after each paper to avoid air bubbles) and a glass plate on top, for 45 minutes. The membrane was placed on blotting paper soaked with 2x saline citrate buffer (SSC) buffer for 10 minutes and crosslinked with UV light using GS Gene linker UV Chamber (Bio-Rad, 125 mJoules for 80 seconds). Blot was prepared for imaging using a DIG wash and Block Buffer set (Roche, Cat no 11585762001). The membrane was rinsed in wash buffer for a five minutes, blocked in 25 mL 1x blocking solution for 30 minutes, and finally incubated with 1:10,000 anti-Digoxigenin Ab (Roche, Cat no. 11093274910) in 13 mL blocking solution for 30 minutes. The membrane was washed twice, for 15 minutes each time, in 25 mL wash buffer, then incubated for 5 minutes in detection buffer. CPSD was diluted to 0.25 mM final (Roche, Cat no 11655884001) in detection buffer and added to the membrane, this was then covered with clingfilm with any air bubbles removed and incubated for 10 minutes at 37°C. Blot was imaged using the PXi gel imaging system (Syngene). Buffer compositions are shown below in Table 2.13.

Table 2.13. Buffers used in EMSA

Buffer	Composition
2x SSC buffer	20x buffer (SLS, Cat no. NAT1224) made to 2x using ddH_2O, final concentration 300 mM NaCl, 30 mM sodium citrate (pH 7)
Wash buffer	10x wash buffer made to 1x using ddH_2O
Blocking solution	10x blocking solution and 10x maleic acid buffer both made to 1x using ddH_2O
Detection buffer	10x detection buffer made to 1x using ddH_2O, final concentration 100 mM Tris-HCl, pH 9.5, 100 mM NaCl

2.7. Antibodies

All antibodies used in experimentation are listed below in Table 2.14 and Table 2.15. Commercial antibodies are validated by the supplier, and we have internally evaluated each antibody by comparing immunofluorescence images or bands generated in western blot analysis, to those presented by the supplier. Secondary antibodies have been checked for non-specific interactions using secondary only controls. In-house anti-CIZ1 antibodies have been applied to CIZ1 KO cells to confirm absence of cross reactivity.

Table 2.14. Antibodies used in immunofluorescence experiments

Antibody	Species Isotype	Working Dilution	Manufacturer	Catalogue Reference
CIZ1 N term – 1793/1794	Rabbit	1:1,000	Coverley (Coverley et al., 2005)	N/A
CIZ1 C term – mAb87 (also called hC221a)	Mouse	1:20	Coverley (Swarts et al., 2018)	N/A
Ubiquityl-Histone H2A (Lys119)	Rabbit	1:25,000	New England Biolabs	8240
Tri-Methyl-Histone H3 (Lys27)	Rabbit	1:1,000	Cell Signalling Technology	C36B11
Scaffold attachment factor A (SAF-A)	Mouse	1:50	Abcam	Ab10297
Anti-rabbit IgG (H+L), Alexafluor 568 linked	Goat	1:1,000	Life Technologies	A11011
Anti-mouse IgG (H+L), Alexafluor 488 linked	Goat	1:1,000	Life Technologies	A11001

Table 2.15. Antibodies used in western blot analysis

Antibody	Species Isotype	Working Dilution	Manufacturer	Catalogue Reference
CIZ1 C term – Rabbit Ex16	Rabbit	1:2,000	Biorbyt	orb329770
CIZ1 C term – Rabbit Ex17	Rabbit	1:1,000	Novus Biologicals	NB100-74624
H3	Rabbit	1:5000	Abcam	ab1791
Actin	Mouse	1:2000	Abcam	ab11003
Anti-Rabbit IgG, Horseradish Peroxidase conjugated	Mouse	1:10,000	Jackson Laboratories	211-032-171
Anti-Mouse IgG, Horseradish Peroxidase conjugated	Goat	1:10,000	Jackson Laboratories	115-035-174

2.8. Bioinformatics

2.8.1. Measuring 3' transcript overexpression in breast cancer

Analysis was performed by Dr Andrew Mason at the University of York. Aligned RNA sequencing data for 1095 primary breast cancer samples from The Cancer Genome Atlas (TCGA) were accessed under dbGaP project 25297. Adjacent normal RNA sequencing data was available for 113 donors. Secondary tumour samples were excluded from analysis. Data were downloaded using the Genomic Data Commons command line client v1.5.0. FASTQ files were regenerated from sample BAM files using samtools v1.10 (Li et al., 2009) to exclude secondary and supplementary alignments, and then BEDTools v2.27.1 bamToFastq (Quinlan and Hall, 2010). No additional quality control steps were performed on the extracted read files. Reads were aligned to the GRCh38 Gencode primary assembly and to individual *CIZ1* transcripts (from Gencode v38 and (Veiga et al., 2022)) using HISAT2 v2.2.0 (Kim et al., 2019). Reads were also pseudoaligned to the Gencode v38 full annotation transcriptome file with kallisto v0.46.0 (Bray et al., 2016), quantified and aggregated to gene-level transcripts per million (TPM) expression values using tximport v1.24.0 (Soneson et al., 2015). The same expression analysis pipeline was also completed on publicly-available RNA sequencing data from six breast cancer cell lines (BT-474: SRR8616195; MCF7: SRR8615758; MDA-MB-231: SRR8615767; SK-BR-3: SRR8615677) and two breast epithelium transformed cell lines (MCF-10A: SRR12877369). *CIZ1* transcript read coverage was normalised to the canonical (ENST00000372938.10) exon 7 coverage, and then stratified by tumour stage, correlated with overall *CIZ1* expression, and compared with adjacent normal tissue where possible. Similar analysis of ESR1 and TP53 was completed to rule out broader 3' coverage biases. *CIZ1* alignments from both transcript and full genome mappings were inspected manually for novel, well-supported splice junctions in IGV Desktop for Windows v2.8.2 (Thorvaldsdottir et al., 2013).

Top 10 stage II patients used to check for 3' bias were as follows: TCGA-E9-A54Y, TCGA-LL-A6FR, TCGA-E9-A3X8, TCGA-AQ-A54O, TCGA-AQ-A54N, TCGA-WT-AB41, TCGA-A2-A3XV, TCGA-LL-A5YL, TCGA-GM-A2DB, TCGA-AO-A03N.

2.8.2. Measuring CIZ1B transcript

Analysis was performed by Dr Andrew Mason at the University of York. RNA-seq data from 433 lung cancer patients available from TCGA was accessed and analysed to measure expression levels of *CIZ1* exons, and normalised relative to each individual canonical *CIZ1* exon 7 as described above. Quantification of *CIZ1B* transcript was also carried out using the same 433 lung cancer patients and 1078 breast cancer patients. Reads were aligned to the *CIZ1* canonical transcript (ENST00000372938.10) using HISAT2 v2.2.0 (Kim et al., 2019), and those mapping to the 3' end of exon 14 were extracted. Reads with gapped alignments which supported the *CIZ1B*-consistent truncation of exon 14 were counted compared to the overall coverage. Alongside analysis conducted by Dr Mason, I quantified the *CIZ1B* transcript relative to normal exon 14/15 junction in normal tissues with assistance from Dr Mason to download and extract the relevant *CIZ1* transcript data from the Genotype-Tissue Expression (GTEx) project portal (GTEx, 2013). RNA-seq data from 459 and 578 normal breast and lung tissue samples respectively were analysed, and the number of *CIZ1B* reads relative to the normal 14/15 junction was calculated to match the analysis conducted in the TCGA dataset.

2.9. Ethics

Primary human mammary epithelial cells were acquired with informed donor consent by Breast Cancer Now Tissue bank and accessed under local approval from the University of York Department of Biology Research Ethics Committee.

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

3. Investigating the over-expression of CIZ1 transcripts in cancer and modelling them in murine cells and biochemically

3.1. Introduction

3.1.1. The Xi is highly corrupted in breast cancer

Inactivation of the X chromosome is a mechanism of dosage compensation in female mammals (Ohno et al., 1959). A detailed description of the events in early development leading to X-inactivation, and CIZ1 involvement in these processes is provided in Chapter 1. There is extensive literature documenting corruption of the inactive X chromosome (Xi) in human tumours, this can include genetic mutations (Jager et al., 2013, Cheng et al., 2015), but is also highly associated with epigenetic modifications. Disappearance of the Barr body (the original name for the Xi) in breast tumours has been known for decades and is considered a hallmark of cancer (Barr and Moore, 1957). Indeed one early study identified 67% of breast cancer patients exhibited Barr Body disruption (Borah et al., 1980), highlighting that it is not an uncommon phenotype that is observed in patients.

Early literature suggested that loss of the Xi in cancer is caused by duplication of the active X chromosome (Xa) and entire chromosome loss of the Xi (Sirchia et al., 2005, Richardson et al., 2006). However more recent studies suggest that there is re-activation of the Xi via epigenetic changes, such as CpG hypomethylation (Moen et al., 2015, Winham et al., 2019), or loss of H3K27me3 and X-inactive specific transcript (*Xist*) at the Xi (Chaligné et al., 2015). It has also been suggested that both mechanisms leading to Xi loss could be occurring (Benoît et al., 2007, Kang et al., 2015a), and that both subtypes of X-chromosome malfunction probably exist in patient populations. It should be noted that Sirchia et al. (2005) and Richardson et al. (2006) conducted their analysis in cancer cell lines and in basal-like and high grade tumours respectively. Due to their time in culture or their advanced position in the cancer timeline, they could therefore have sustained genetic changes downstream of earlier events, including epigenetic changes. Indeed it has been frequently observed that epigenetic changes likely precede and go on to cause genetic changes in cancer (Brower, 2011). Here, we hypothesise that early drivers of Xi loss in cancer are mediated by epigenetic changes linked to CIZ1, rather than genetic changes.

3.1.2. *BRCA1* and *Xist* in breast cancer

The DNA damage response (DDR) is a mechanism evolved by cells to detect DNA damage, notify the cell of this damage, and initiate repair pathways, or if repair is not possible induce apoptosis (Jackson and Bartek, 2009). Breast cancer type 1 susceptibility protein (BRCA1) is phosphorylated in response to double stranded DNA breaks, and thus plays a role in the DDR (Cortez et al., 1999). This makes BRCA1 an important tumour suppressor, and when mutated it has been implicated in cancer development via genetic and epigenetic dysregulation (Downs and Wang, 2015). As discussed in the introduction, *Xist* is a long non-coding RNA that is required for X-inactivation (Penny et al., 1996). Thus, if there is a functional link between BRCA1 and *Xist*, this could provide a mechanism for Xi loss in breast cancer in patients that have dysfunctional BRCA1. Initially it was suggested that loss of BRCA1 could directly lead to loss of *Xist* localisation at the Xi, and when BRCA1 was introduced to BRCA1 deficient cells *Xist* staining appeared at the Xi (Ganesan et al., 2002). However multiple research groups have shown that *Xist* coats the Xi regardless of BRCA1 status (Vincent-Salomon et al., 2007, Xiao et al., 2007, Kang et al., 2015b). A more conservative view that has been suggested, is that BRCA1 involvement is broad, and is associated with a wide failure to maintain heterochromatin effectively in cancer (Pageau et al., 2007). Individuals with *BRCA1* mutations do have a higher propensity for skewed X-inactivation (inactivation is not random and one X chromosome is selected for inactivation over the other), compared to patients without *BRCA1* mutations. However, *BRCA1* mutated patients that presented with skewed X-inactivation had later disease onset than un-skewed patients, suggesting it to be beneficial in this subset of patients (Lose et al., 2008). Considering this, clearly the relationship between BRCA1 and Xi status is still greatly misunderstood. However, regardless of the mechanism by which Xi is corrupted in breast cancer patients, the outcome is loss of dosage compensation and overexpression of X-linked genes relative to normal levels, which could include activation of oncogenes (Spatz et al., 2004).

3.1.3. CIZ1 localisation at the Xi

The Coverley lab has previously shown that CIZ1 is able to form large aggregates at the Xi, and co-localises with *Xist* and histone post translational modifications (PTMs) such as H3K27me3 (Ridings-Figueroa et al., 2017). In addition, CIZ1 null cells have been shown to form dispersed rather than discrete *Xist* clouds (Ridings-Figueroa et al., 2017, Sunwoo et al., 2017); and also display chromosome-wide loss of the repressive chromatin modifications H2AK119ub1 and H3K27me3, leading to inappropriate gene regulation (Stewart et al., 2019). This highlights the importance of CIZ1 expression, and suggests that it is part of a complex set of interactions involving, RNA:protein, protein:protein and protein:DNA interactions, though the full extent of these interactions remain unknown. Since Xi corruption is a hallmark of cancer, Ernesto Lopez, a former Masters student in the Coverley lab, investigated the status of CIZ1 at the Xi in a variety of breast normal and cancer cell lines. Utilising two different CIZ1 antibodies with epitopes at different ends of the protein allowed for identification of expression changes at the domain level (Figure 3.1A). This showed that in primary human epithelial cells and a normal breast cell line, that CIZ1 was localised to the Xi as expected, and that N and C-terminal domain expression was largely balanced (Figure 3.1B). However, in most of the breast cancer lines CIZ1 was not localised to the Xi, and additionally in some cases expression of the N and C-terminus was uncoupled (Figure 3.1C).
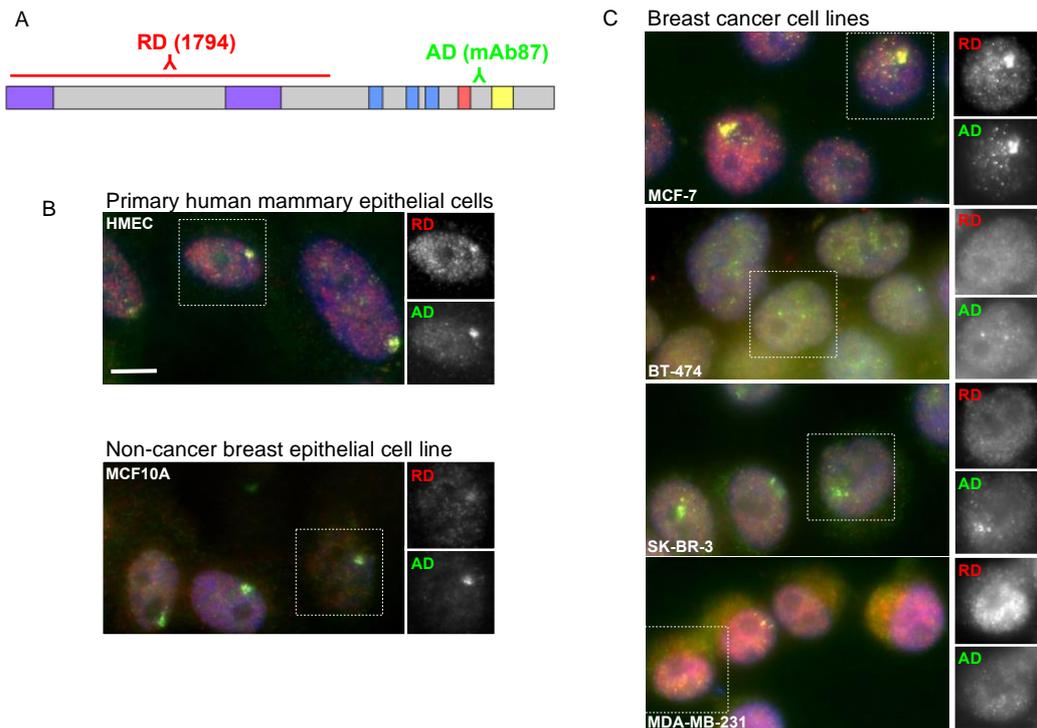
Figure 3.1. CIZ1 status at the Xi in normal and cancer breast cells

A. Diagram of CIZ1 identifying the location of the epitopes of the antibodies used to measure CIZ1 protein in breast-derived cells. Since the antibodies are from different species, they can be applied together to measure the CIZ1 replication domain (RD) and anchor domain (AD) status in the same cell at the same time. Diagram produced using IBS software (Liu et al., 2015b).

B. Immunofluorescence images of normal breast cells used in analysis, including primary human mammary epithelial cells (HMEC) and a commonly used breast normal cell line (MCF10A). Images taken by Ernesto Lopez and presented by Dawn Coverley. Coloured images are a merge of blue DNA, red CIZ1 replication domain (RD), green CIZ1 anchor domain (AD). Grayscale images are individual red and green channels of CIZ1 RD and CIZ1 AD respectively. Images from B and C are to the same scale, scale bar is 10µm.

C. Four breast cancer cell lines. Images taken by Ernesto Lopez and presented by Dawn Coverley.

3.1.4. CIZ1 corruption in cancer

An extensive description of CIZ1 corruption in cancer has been provided in the introduction, but in summary this is seen across multiple cancer types and can include under-expression, over-expression and aberrant alternative splicing. Pre-existing data produced by Heather Sercombe, a former member of the Coverley lab, identified over-expression of 3' transcripts of CIZ1 in multiple cancer types. It was hypothesised that these transcripts might lead to over-expression of translated small C-terminal fragments, implicated in a dominant negative (DN) mechanism in cancer. This is supported by our most recently published data, which showed that both the N and C-terminus of CIZ1 must be present in the same polypeptide to localise CIZ1 to the Xi (Sofi et al., 2022). Thus, overexpression of the C-terminus would be unable to localise to the Xi, but could in theory interact with endogenous CIZ1 complexes, or its binding partners, and lead to corruption as seen in cancer cell lines (Figure 3.1C). Consolidation of this data, and new investigation into the mechanisms involved, is the focus of this chapter.

3.2. Aims

- Investigate 3' overexpression of CIZ1 amplicons in human tumours by analysing The Cancer Genome Atlas (TCGA) data and relate to putative protein C-terminal overexpression. This objective was explored with support and close collaboration of Bioinformatician Dr Andrew Mason, who is an approved and registered user of TCGA and Viking

- Model CIZ1 C-terminus overexpression in cells, via transient transfection of GFP tagged CIZ1 constructs

- Explore domain requirements by creation of mutants of CIZ1 C-terminal fragments and look for differences in endogenous CIZ1 behaviour

- Create, express and purify C-terminal CIZ1 and associated mutants in bacteria and measure differences in their biochemical properties

3.3. Experimental Design

All data described here involving the use of human samples and cell lines and cancer datasets focuses on breast cancer, but may also be applicable to a range of other cancer tissue types (for which we have much less extensive evidence). We chose to focus on breast cancer for two reasons i) the vast majority of breast cancer occurs in women, so this allows us to use the Xi as a model to monitor CIZ1 corruption in cancer, and as described above this is a long documented phenotype observed in cancer; ii) normal control breast tissue is easier to obtain than for other tissue types, and further strengthens our analysis since we do not have to rely only on "normal" control cell lines.

3.3.1. Identifying 3' transcript overexpression in TCGA dataset

Patient RNA data for 1095 primary breast cancer samples, 113 adjacent normal samples and several breast normal and cancer cell lines (MCF10A and MCF7, BT-474, SK-BR-3 and MDA-MB-231 respectively), were downloaded and aligned to *CIZ1* (Dr Mason). To quantify individual exon expression and identify if 3' overexpression was observed, each transcript read coverage was internally normalised to the canonical (ENST00000372938.10) exon 7 coverage. This is in line with quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis in Figure 3.2A. This exon was utilised for calibration due to its central location and because it has not been implicated in any cancer associated alternative splicing, unlike other *CIZ1* exons (Swarts et al., 2018).

3.3.2. C-terminal fragment modelling

C-terminal overexpression was modelled in a mouse cell line (D3T3) and primary embryonic fibroblasts (PEFs), to investigate if any changes were seen in endogenous CIZ1-Xi status as a consequence. The constructs used corresponded to the terminal 181 and 275 amino acids of murine CIZ1 (referred to as C181 and C275 respectively). They both contained an N-terminal GFP tag, and were transiently transfected for 24 hours prior to immunofluorescence to monitor the status of endogenous CIZ1. A set of deletion constructs were created via PCR mutagenesis to generate proteins lacking domains of interest.

3.3.3. *in vitro* analysis

Mutant forms of the GFP tagged CIZ1-C181 were also created in the GST tagged CIZ1 system for biochemical analysis. These were transformed into BL21 cells to facilitate expression and downstream purification of protein products. These were subsequently analysed by size exclusion chromatography (SEC), size exclusion chromatography multiple laser light scattering (SEC-MALLs) and western blot analysis.

3.4. Results

3.4.1. Overexpression of *CIZ1* 3' transcripts in the TCGA

Prior to my studies, analysis of *CIZ1* transcript expression levels in different tissues (including normal and multiple stages of cancer development) measured by qRT-PCR was conducted by Heather Sercombe. Two amplicons were measured, one at the 5' end and the other at the 3' (Figure 3.2A), which correspond to the N-terminal replication domain (RD) (Coverley et al., 2005, Copeland et al., 2010, Copeland et al., 2015) and the C-terminal anchor domain (AD) (Ainscough et al., 2007) of CIZ1 respectively. Only the breast tissue data is visualised here to provide context for the follow up work that I conducted. Both domains of *CIZ1* were elevated compared to histologically normal tissue, as early as stage I, which is consistent with previous reports of elevation of *CIZ1* in breast (den Hollander et al., 2006). However, in addition to this observation, it was also noted that there was a marked disparity between 5' and 3' amplicons, with increased 3' expression as early as stage I when comparing exon 7 to exon 16.

Analysis of exon representation in tumour samples from TCGA patients confirmed the previous result, that *CIZ1* RD exons (exons 5 and 7) were significantly under-represented compared to AD (exons 14 and 16) at all stages (Figure 3.2B). Comparison of exon 5, 14 and 16 expression in the normal adjacent and tumour samples from TCGA, showed that exon 5 is under-expressed in the tumour samples relative to the adjacent samples, whereas exon 14 and 16 are over-expressed in the tumour samples relative to the adjacent samples (Figure 3.2C). This 3' elevation in tumour samples is evident between exon 9 and 12, and is further exaggerated at exon 13 onwards. This could be explained by the additional presence of the already known *CIZ1-F* splice variant that is present in breast cancer samples which lacks exon 8-12 (Swarts et al., 2018), which could lead to antagonistic mechanisms of under-representation of exons 8-12, competing with the over-expression from exon 10 onwards.

To evaluate the possibility of technical over representation of 3' sequences the ten patients that displayed the most extreme 3' bias of *CIZ1* (ranging from 2.8 to 5.3 Log_2FC when comparing exon 15 to exon 7), showed none or minor 3' bias in control genes oestrogen receptor alpha (*ERα, ESR1*) and tumour protein 53 (*TP53*) respectively (Figure 3.2D) relative to their own internal controls. This provides further confidence that the 3' overexpression is not related to library preparations via polyA capture RNA sequencing.
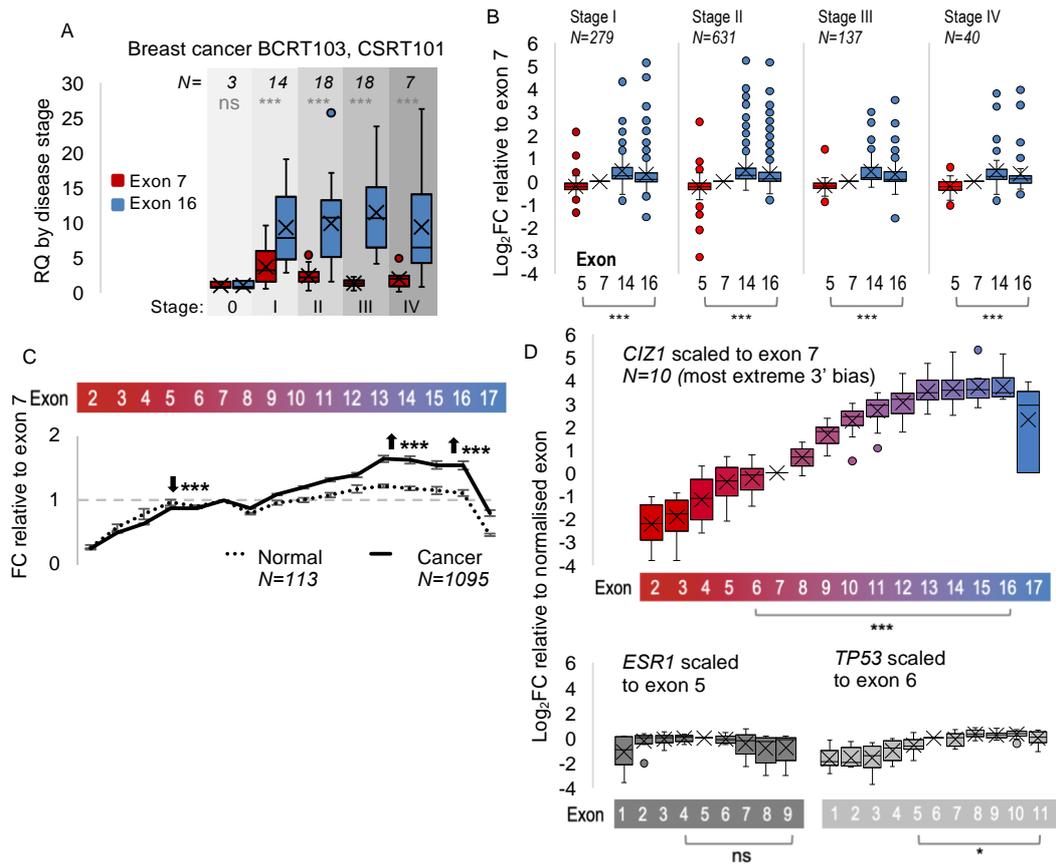
Figure 3.2. *CIZ1* RNA expression in TCGA

A. Data provided by Heather Sercombe and Dawn Coverley. *CIZ1* expression measured by qRT-PCR analysis of cDNA in a microarray (BCRT103 and CSRT101). Primers in exon 7 and 15 allow quantification of 5' and 3' *CIZ1*. 5' and 3' expression was internally normalised to normal (stage 0) to calculate relative quantity (RQ) of *CIZ1* in tumours. Comparison of exon 7 and 16 expression within stages is by t-test. ***=p<0.001. N=samples measured.

B. *CIZ1* expression in breast cancer TCGA samples across disease stages. Each patient was internally normalised to exon 7 expression to facilitate measurement of differences in fold change (FC) expression for exons 5, 7, 14 and 16. Statistical analysis of elevation of exon 16 relative to exon 5 was conducted using a Mann-Whitney U test. Stage I, p=0; stage II, p=0; stage III, p=0; stage IV, p=0.000002. N=samples in each stage. The total unrepresented number of outliers for exon 5, 14 and 16 respectively are stage I: 9, 25 and 39, stage II: 18, 54 and 86, stage III: 3, 10 and 17 and stage IV: 3, 5 and 6

C. *CIZ1* expression in all breast cancer and normal adjacent TCGA samples. Each patient was internally normalised to exon 7 expression to facilitate measurement of changes in fold change (FC) expression. Statistically significant elevation of exon 14 and 16 and reduction of exon 5 in cancer relative to exon 14,16 and 5 respectively in normal tissue was revealed by a Mann-Whitney U test. Exon 5 p=0.000007; exon 14 p=0.000004; exon 16 p=0.000175. Error bars show SEM. N=number of samples measured.

D. Comparison of exon expression of *CIZ1* and two control genes *ESR1* and *TP53* in the ten breast cancer patients that displayed the highest *CIZ1* exon 15 over expression. Transcript expression is internally normalised in each sample to the exon stated. Statistical analysis of 3' expression relative to 5' expression was conducted using a t-test. *CIZ1* p=0.00045; *ESR1* p=0.050; *TP53* p=0.037.

69

3.4.2. Cancer cell lines show *CIZ1* 3' transcript overexpression and CIZ1 protein status is corrupted

Analysis of *CIZ1* transcript coverage in the same normal and four breast cancer cell lines used in Figure 3.1 was conducted (Figure 3.3A). Cancer subtype based on information on receptor status was taken from (Subik et al., 2010). As in primary tumours, this revealed cancer specific *CIZ1* 3' overexpression starting at around exon 11 (Figure 3.3B). This was variable across the cell lines, with SK-BR-3 displaying the highest elevation, followed by MCF7 and MDA-MB-231, and very minimal elevation in the BT-474 line. However, the BT-474 line may show evidence to support the existence of the *CIZ1-F* transcript, due to the fall in expression of exon 8-12, and may therefore still harbour a cancer specific *CIZ1* transcript. Since only one cell line of each subtype was analysed it's not possible to draw conclusions on whether a particular subtype is more susceptible to a *CIZ1* 3' elevation, and this would be an interesting avenue to pursue using the TCGA data and separating patients based on breast cancer subtype.

To investigate if the 3' elevation led to expression of C-terminal protein fragments, western blot analysis was conducted (Figure 3.3C). The molecular weight of human CIZ1 is predicted to be 100kDa, and routinely members of the Coverley lab observe two different CIZ1 forms that migrate near the 130 and 100kDa marker, referred to hereafter as CIZ1 isoform 1 and 2 respectively. Since CIZ1 is routinely observed to be retarded in denaturing gel electrophoresis, (this is validated later using purified recombinant proteins), my hypothesis is that the larger entity (isoform 1) corresponds to the full-length form. Since the smaller entity (isoform 2) is consistently not identified via antibodies that recognised extreme C-terminus epitopes (Rabbit Ex17), I suggest that it lacks this part of the protein. It's elevation in cell lines compared to the normal primary patient cell lysate (human mammary epithelial cells, HMEC), suggests it is not the predominant form in normal human adult cells, but is a feature of long-term culture. Surprisingly, whilst isoform 1 is identified in all samples by the N-terminal antibody, and is relatively consistently expressed (with the exception of the metastatic cancer cell line MDA-MB-231), it is not consistently identified by the C-terminal Rabbit Ex17 antibody. In fact, detection via exon 17 suggests elevation in the three primary tumours (MCF7, BT-474, SK-BR-3) compared to the normal samples and the metastatic cancer (MDA-MB-231). Lack of agreement between N-terminal and C-terminal antibodies with respect to isoform 1, could be consistent with PTMs at the C-terminus that mask the exon 17 epitope. PTM modifications in this region, and their involvement in CIZ1 regulation, are explored further in Chapter 4 so won't be discussed further here.

Since a drop in exon 17 transcript expression is observed in both normal and cancer patients, and in the cell lines (Figure 3.2C,3.3B), this suggests that there could be limitations to using an antibody at the extreme C-terminus to measure CIZ1 C-terminal expression. Considering this, the use of different C-terminal antibodies was explored. An antibody that binds more upstream of the C-terminus (Rabbit Ex16), revealed multiple protein entities, which could include CIZ1 isoforms as well as non-specific interactions. The use of an additional C-terminal antibody, with an epitope contained in the last ZF of CIZ1, proved to be too non-specific for use (data not shown). However, shared protein forms recognised by this antibody, and the Rabbit Ex16 antibody, have been highlighted by arrows. This provides preliminary evidence for the existence of upregulated C-terminal CIZ1 fragments in cancer cell lines, but also highlights considerable complexity across the different cell lines.

The disparity between the ponceau S stained membrane (total protein), and the actin and histone antibody outputs, documents the problems that were encountered when trying to ensure consistent loading of samples from patients and cell lines. This highlights the requirement to inspect multiple loading controls to ensure correct sample loading. Whilst follow up experimentation is required, these blots provide early evidence of CIZ1 C-terminal fragments in breast cancer, and illustrates that there are multiple points of dysregulation of CIZ1 protein status in breast cancer.
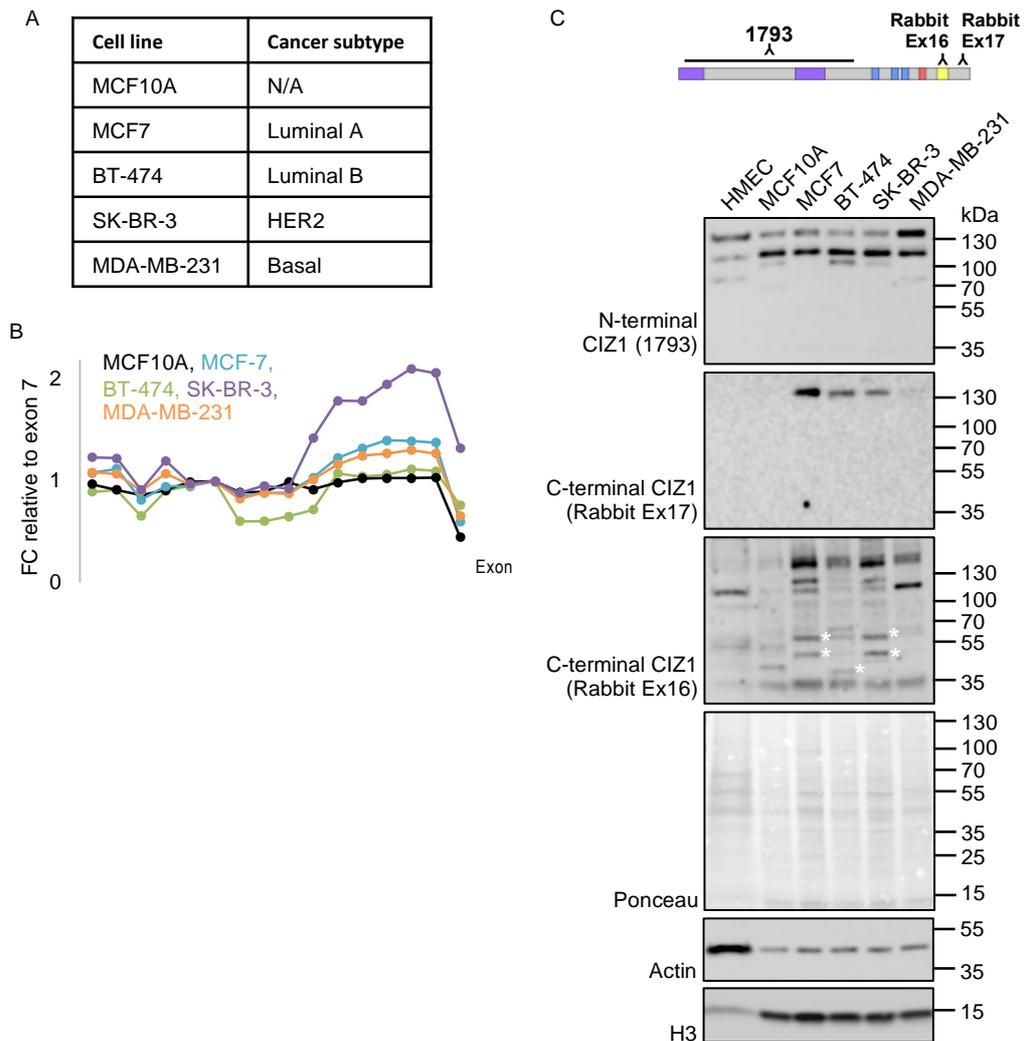
Figure 3.3. *CIZ1* RNA and CIZ1 protein expression in cell lines

A. Breast-derived cell lines used in experimentation. MCF10A are a normal tissue-derived cell line. Information taken from (Subik et al., 2010).

B. CIZ1 transcript expression in normal (MCF10A) and cancer (MCF7, BT-474, SK-BR-3, MDA-MB-231) breast cell lines. Each line is normalised internally to exon 7 expression to measure fold change (FC) across exons.

C. Western blot analysis of CIZ1 forms in primary patient (HMEC), normal (MCF10A) and cancer (MCF7, BT-474, SK-BR-3, MDA-MB-231) breast cell lines. Putative CIZ1 cancer specific forms are highlighted with arrows. Diagram details antibody epitope location on a map of CIZ1 domains.

3.4.3. Dominant Negative Affect of GFP-tagged C-terminal CIZ1 Constructs

I wanted to model the effect of the overexpression of C-terminal anchor domain (AD) containing fragments, first in a murine cell line (D3T3) to collect preliminary data and optimise the assay, and then in murine primary embryonic fibroblasts (PEFs) to confirm outcomes under near physiological conditions. GFP-tagged C-terminal constructs were transfected into female cells (D3T3 or PEF), and their effects on endogenous CIZ1 at Xi were visualised using an antibody with an epitope located in the N-terminus of CIZ1 to avoid cross reaction with the ectopic CIZ1 fragment (Figure 3.4A).

Using a lipid-based transient transfection method only a subset of cells receive vector, generating an internal control set within every population. Comparison of the untransfected (UT) population to those transfected with the GFP empty vector (vector control) showed no observable difference in endogenous CIZ1, with ~80% of cells containing a measurable CIZ1-Xi assembly (Figure 3.4B). Whilst GFP-tagged forms of full length CIZ1 localise at pre-existing Xis in WT cells (Sofi et al., 2022), the small C-terminal forms studied here are unable to. In fact, cells transfected with C275 or C181 have disrupted endogenous Xi-CIZ1. This led to a measurable drop in cells that contained a discrete CIZ1 assembly that resembles the CIZ1-Xi accumulation normally found within the nuclei of the UT and GFP population (approximately 40%). This suggests that they interfere with endogenous CIZ1 in a dominant-negative (DN) mechanism. There was no difference between the effect of C275 and C181 construct, which suggests that the zinc fingers (ZF) are not required for the dispersal mechanism. All further experimentation after this utilised the C181 fragment.

To enable subtler evaluation in the next phase of experimentation, the assay was further optimised to include the use of three scoring criteria for CIZ1 Xi status, rather than the two previously used. Nuclei that contained endogenous CIZ1 which had a dispersed or diffuse appearance across the nucleus, or where the CIZ1-Xi staining was visibly smaller or weaker compared to the expected larger and brighter accumulations at the Xi, were classified as an "intermediate" phenotype. By adding an "intermediate" category in addition to a present or absent CIZ1 marked Xi, potential ambiguity in whether to count or reject cells with these phenotypes was avoided. Example pictures of the three phenotypes observed and scored is shown in Figure 3.4C, with the green channel omitted to aid visualisation.

As seen before there was no observable difference between the UT and vector control population, while C181 caused a decrease in cells with CIZ1 marked Xis. However, with the addition of a third phenotype the ~60% of the population affected by the C181 fragment is now divided across two groups, with a measurable increase in the number of cells in the intermediate and absent category in C181 transfected population compared to the UT and vector control cells (Figure 3.4D). The reduction in CIZ1 assemblies was also confirmed by non-subjective measurement of endogenous CIZ1 fluorescence intensity using FIJI, which reported a drop in maximum nuclear CIZ1 fluorescence intensity in transfected cells compared to untransfected cells (Figure 3.4E).

During the pandemic I took a leave of absence from my studies to work in a COVID testing lab, and during this time the downstream consequences of expression of my C181 fragment was tested using the above regime by Dawn Coverley and Louisa Williamson. Using PEFs, they confirmed the effect on endogenous CIZ1, and then monitored the status of two histone repressive marks that are normally enriched at the Xi. The number of cells with the H3K27me3 modification enriched at the Xi showed no changes between UT and C181 transfected cells, which was present in approximately 80% of the population, as previously observed (Ridings-Figueroa et al., 2017). In contrast the H2AK119ub1 modification showed a significant drop in C181 transfected cells compared to UT cells, from ~70% to ~20%. We hypothesise that cells that show dispersal of the endogenous CIZ1 assemblies, correspond to cells that also show loss of H2AK119ub1. However, since these parameters were tested in independent populations, it cannot be formally concluded from this data. Upon application of the broad spectrum deubiquitinase (DUB) inhibitor PR-619, in parallel with C181 in transfected cells, the fall in number of cells with a H2AK119ub1 modification was partially abrogated (Figure 3.4F). Example pictures of the cell populations can be seen in Figure 3.4G. This suggests that exposure to DUBs may be the cause of H2AK119ub1 loss.
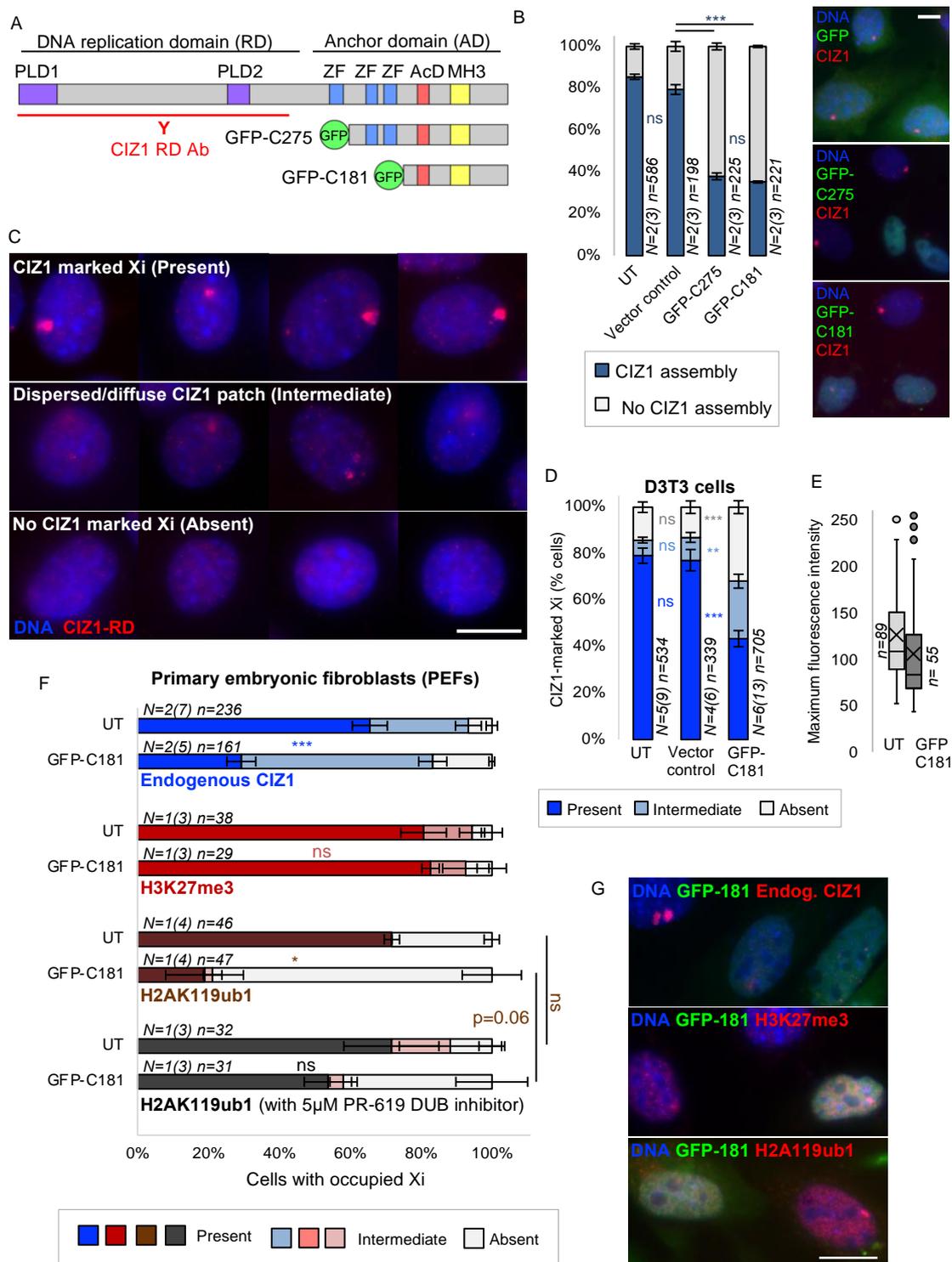
Figure 3.4. C-terminal overexpression modelling in mouse cells

A. Diagram of murine CIZ1 full-length protein and the small C-terminal fragments used to model the effects of C-terminus overexpression. The domains highlighted are: prion like domain 1 and 2 (PLD1 and PLD2) at position 1-67 and 361-399 respectively, as documented in (Sofi et al., 2022); three zinc fingers (ZF1, ZF2 and ZF3) at positions 537-561, 600-620 and 631-653 respectively (ZnF_C2H2 SM00355, ZF_C2H2 sd00020 and ZF_C2H2 sd00020); an acidic domain (AcD) containing a highly concentrated area of aspartates and glutamates at position 746-779; and a matrin-3 like domain at position 746-770 (ZnF_U1 smart00451). C275 and C181 correspond to the terminal 275 and 181

amino acids of CIZ1 respectively. Both C-terminal constructs have a GFP tag at the N-terminus for visualisation. Diagram produced using IBS (Liu et al., 2015b).

B. Left: Results of dispersal assay in D3T3 cells. Cells were grouped into two categories, either having a present CIZ1 Xi patch or an absent CIZ1 Xi patch. Statistical analysis of the proportion of cells in the present category to compare dispersal capabilities was conducted using a One-way ANOVA with a Tukey post hoc test. UT and GFP empty vector, p=0.12; GFP empty vector and C275, p=4.42×10^{-7}; GFP empty vector and C181, p=2.55×10^{-7}; C275 and C181, p=0.58. Error bars show SEM. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Right: Example pictures of status of cells used in analysis. Scale bar is 10µm.

C. Illustration of dispersal assay in D3T3 cells using a three category scoring system. Examples images were taken under identical exposure parameters, showing three categories of nuclei with either discrete aggregates of CIZ1 at Xi (red, upper), no detectable CIZ1 and Xi (red, lower), or intermediate (dispersed into multiple smaller aggregates or diminished in overall staining intensity) (red, middle). DNA is blue. Images are enhanced for reproduction, but are scored or quantified prior to manipulation. Scale bar is 10µm.

D. Comparison of transfection of C181 and GFP empty vector control in dispersal assay using the three category method. Statistical analysis of the proportion of cells in each category to compare dispersal capabilities was conducted using a One-way ANOVA with a Tukey post hoc test. UT and GFP empty vector: CIZ1 marked Xi, p=0.94; dispersed/diffuse CIZ1 patch, p=0.71; no CIZ1 marked Xi, p=0.97. GFP empty vector and C181: CIZ1 marked Xi, p=4×10^{-6}; dispersed/diffuse CIZ1 patch, p=0.0015; no CIZ1 marked Xi, p=0.00016. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars show SEM.

E. Box and whisker plot showing maximum endogenous CIZ1 fluorescence intensity per nucleus, in transfected and untransfected cells, derived from unmodified images using FIJI. n=nuclei measured. The total unrepresented number of outliers are 4 for UT and 5 for C181.

F. Graph shows the frequency of endogenous CIZ1-Xi aggregates in nuclei that are or are not expressing C181, compared to frequency of repressive marks. All data here is collected from primary embryonic fibroblasts (PEFs) at passage 2-3. Statistical analysis of the proportion of cells in the present category to compare dispersal capabilities was conducted using a t-test. Endogenous CIZ1 in UT and C181 cells, p=0.00033. H3K27me3 in UT and C181 cells, p=0.79. H2AK119ub1 in UT and C181 cells, p=0.016. H2AK119ub1 in UT and C181 cells treated with 5µM PR619, p=0.36. H2AK119ub1 in UT untreated and UT treated with 5µM PR619, p=0.99. H2AK119ub1 in C181 cells untreated and C181 cells treated with 5µM PR619, p=0.061. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars are SEM.

G. Example pictures of status of endogenous CIZ1, H3K27me3 and H2AK119ub1 in untransfected and C181-transfected cells. Scale bar is 5µm.

3.4.4. MH3 domain is implicated in the dominant negative dispersal effect of C181

To unpick the DN effect of the CIZ1 C-terminal fragment further, I focused on the functional domains in C181 and generated a set of deletion constructs (Figure 3.5A) for analysis in the cellular context. These mutants included deletion of: the matrin-3 domain (MH3); a region downstream of the MH3 that is perfectly conserved in human and murine CIZ1 (ΔNALTAF), which also harbours cancer-associated polymorphisms detected by whole exome sequencing and documented in COSMIC v90 (released 05-SEP-19) (Tate et al., 2019); the terminal 37 amino acids (Δ37), that we observe is removed during recombinant CIZ1 expression in bacteria (described further below); and I122 a pre-existing Coverley lab construct described in (Ainscough et al., 2007), that lacks most of the upstream AcD and downstream the terminal 15 amino acids.

Analysis of the number of cells positive for GFP-CIZ1 with and without the presence of detergent prior to fixation showed that they were all expressed and were primarily nuclear proteins, but had variable resistance to detergent (example pictures with average cells positive for GFP-CIZ1 can be seen in Figure 3.5B). This suggests that within a transfected population some cells have established bound/assembled forms of CIZ1, and others contained soluble forms. The two constructs that showed the highest degree of extractability were the Δ37 and the ΔMH3 constructs, with a percentage fall of 66% and 85% respectively upon addition of detergent, whereas the C181, ΔNALTAF and I122 were less affected by detergent, falling to 55%, 40% and 33% respectively (Figure 3.5C). This suggested that the MH3 domain and the terminal 37 amino acids are implicated in the anchorage of the AD of CIZ1 to insoluble components of the nucleus. Analysis of the dispersal capability of these constructs was conducted in the presence of detergent since I wanted to analyse potential effects on the bound form of endogenous CIZ1.

The majority of the CIZ1 mutants (ΔNALTAF, Δ37 and I122) showed no differences in dispersal capability compared to C181, and in all these conditions approximately 20% of the population had endogenous CIZ1 present at the Xi (Figure 3.5D). This suggests that these regions are not implicated in the dispersal mechanism. However, deletion of the MH3 domain did lead to a small but significant abrogation in the dispersal capability of the construct, leading to an increase from 20% to 40% endogenous CIZ1 present in the ΔMH3 transfected population (Figure 3.5D). FIJI analysis of maximum nuclear intensity as described above, was conducted on C181 and ΔMH3, and this corroborated the cell scoring result. It showed an increase in maximum nuclear intensity of C181 ΔMH3 compared to C181 (Figure 3.5E), though both constructs remain capable of affecting assembly intensity when compared to untransfected cells in the same population. This suggests that the MH3 domain is involved in a key interaction that facilitates the dispersal of the endogenous CIZ1.

A

665   AcD   MH3   845

GFP — C181

GYVCQICHKFYDSNSELRLSHCKSLAHFENLQKY
GFP — C181Δ746-779 (ΔMH3)

CAINARNALTALFTSS
GFP — C181Δ796-811 (ΔNALTAF)

TSSHQPSPQDTVKMPSKVKPGSPGLPPPLRRSTRLKT
GFP — C181Δ809-845 (Δ37)

PGLPPPLRRSTRLKT
GFP — CIZ1 708-830 (I122)
KETGSPDEDHFITVDAVGCFES
GQEEDEDDDEEEEEEGEIEAEE

C

Reduction in GFP positive cells(%)

C181   ΔMH3   ΔNALTAF   Δ37   I122

B

	C181	ΔMH3	ΔNALTAF	Δ37	I122
-D	33%	62%	45%	44%	60%
+D	15%	9%	27%	15%	40%

D   **D3T3 cells**

ns

*** *

CIZ1-marked Xi (% cells)

UT   C181   ΔMH3   ΔNALTAF   Δ37   I122

N=3(4) n=128
N=3(5) n=273
N=3(6) n=326
N=1(3) n=83
N=2(6) n=204
N=1(3) n=135

☐ Absent
☐ Intermediate
■ Present

DNA   Ectopic CIZ1   Endogenous CIZ1

UT   C181   ΔMH3
ΔNALTAF   Δ37   I122

E   **Primary embryonic fibroblasts (PEFs)**

Maximum fluorescence intensity

UT   C181   ΔMH3
n=200   n=94   n=80

DAPI
GFP-C181
CIZ-RD

CIZ-RD

DAPI

DAPI
GFP-ΔMH3
CIZ-RD

CIZ-RD

DAPI

78

Figure 3.5. Comparison of GFP tagged CIZ1 mutants

A. CIZ1 fragments transiently transfected into murine cells. used for biochemical analysis. Numbers indicate the boundaries respective to the amino acid numbers of the murine full length CIZ1. Domains highlighted are the acidic domain (AcD) and matrin 3-homology domain (MH3). Diagram produced using IBS software (Liu et al., 2015b).

B. Fluorescent field views of CIZ1 C-terminal mutants. –D and +D correspond to the absence or presence of detergent prior to fixation. The percentages reflect number of cells positive for GFP-CIZ1. Scale bar is 5µm

C. Percentage fall in CIZ1 construct presence upon detergent treatment

D. Left: Frequency of endogenous CIZ1-XI aggregates in nuclei that are not expressing C181 (UT) or various different C181 based constructs, expressed as three categories; present, absent or intermediate, as shown in Figure 3.4C. Statistical analysis of the proportion of cells in the present category to compare dispersal capabilities was conducted using a One-way ANOVA with a Games-Howell post hoc test. UT and C181 $p=1.9\times10^{-4}$, C181 and ΔMH3 p=0.040, C181 and ΔNALTAF p=0.93, C181 and Δ37 p=0.52, and C181 and I122 p=1. N=technical replicates, n=nuclei scored. Error bars show SEM. Analysis performed in D3T3 cells. Right: Example pictures of D3T3 cells used in analysis. Scale bar is 10µm.

E. Left: Box and whisker plot of maximum endogenous CIZ1 fluorescence intensity per nucleus, in untransfected (UT) cells compared to C181 and ΔMH3 transfected cells. n=nuclei measured in each group. Analysis performed in PEFs. The total unrepresented number of outliers in UT, C181 and ΔMH3 are 39, 0 and 14 respectively. Right: Example images of PEFs used in analysis. Scale bar is 5µm.

3.4.5. MH3 Domain is required for Dimerisation of CIZ1

To extend my *in vitro* dispersal analysis, similar C181-derived deletion mutants were generated in GST tagged vectors (Figure 3.6A), allowing for their expression and purification for use in biochemical assays. Analysis of these forms of CIZ1 in sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) shows the existence of multiple forms of CIZ1 in some cases (Figure 3.6B). During bacterial expression C181 appears to undergo a cleavage event at the extreme C-terminus, leading to the existence of a full form and a "cleaved" form observable in coomassie stained gels. The approximate cleavage location has been estimated to occur 37 amino acids upstream of the C-terminus, based on absence of representative peptides in mass spectrometry analysis (Adam Dowle at the University of York, not shown), and by loss of epitope for our C-terminal CIZ1 antibody (Figure 3.6B). To facilitate production of pure preparations, and to test if this region of CIZ1 has any functional relevance I created the Δ37 form.

All C181-derived deletion constructs were analysed via size exclusion chromatography (SEC), to observe if removal of any domains led to unanticipated changes in the elution profile. SEC allows the user to identify any changes in size or shape of the protein, since elution time should only be altered based on a proteins ability to enter the pores of the stationary phase of the column, irrespective of charge. Larger entities cannot enter through all the pores as effectively so elute earlier, whereas smaller entities can so they elute later. A shift to a later elution time would suggest a reduction in multimeric state or a more condensed shape. Allowing for the loss of 37 amino acids, the elution position of C181 and Δ37 was similar, suggesting that the terminal amino acids have no role in native CIZ1 state (Figure 3.6C). However, there was a shift in elution time in ΔMH3, suggesting either a change of conformation or a reduction in multimeric state. The creation of the double mutant ΔMH3,37 led to similar elution times as ΔMH3, suggesting that deletion of MH3 alone is sufficient for this change in native CIZ1. The traces shown in Figure 3.6C are cropped to the CIZ1 peak of interest, the full SEC traces are presented in Appendix A.

Further size exclusion chromatography coupled with multiple angle laser light scattering (SEC-MALLs) analysis of C181 and ΔMH3 generated an accurate molecular weight estimate of these entities. This confirmed that C181 exists as a dimer under native conditions, whereas ΔMH3 exists as a monomer (Figure 3.6D). This suggests that the MH3 domain is a dimerisation interface. Modelling of C181 secondary structure, using AlphaFold2 (Mirdita et al., 2022), corroborated our experiential data and predicted that a β-strand within each MH3 domain forms the basis for the dimerisation to form an anti-parallel dimer *in vitro* (Figure 3.6E). The traces shown in Figure 3.6D are cropped to the CIZ1 peak of interest, the full SEC-MALLs traces are presented in Appendix B.

To test whether the MH3 domain is required for the reported ability of murine C181 to directly bind RNA (Sofi et al., 2022), I utilised an electrophoretic mobility shift assay (EMSA). If the RNA probe being measured is able to bind to the protein, it is retarded in the gel due to the increased size, thus an observable shift upwards is reflective of a RNA:protein complex. Previous work has shown that the C-terminus of CIZ1 binds promiscuously to multiple RNA forms, including *GAPDH* RNA (Sofi et al., 2022), so this was used to compare RNA binding capabilities of C181 and ΔMH3. Analysis of the digoxigenin (DIG) labelled *GAPDH* probe showed that loss of the MH3 domain did lead to loss of RNA interactions *in vitro*. However, C181 interactions with RNA were weak, and indeed when comparing bound and unbound RNA at the highest total protein concentration there is still a large proportion of unbound RNA. This is in line with our previous results suggesting that there are CIZ1 RNA interactions in the AD, but they are weak and promiscuous compared to the RD (Sofi et al., 2022). Furthermore, at this point it is not clear whether the MH3 domain binds RNA directly or whether dimerisation itself is required for RNA binding.

Dependence on the MH3 domain for dimerisation leads us to the hypothesis that the ability to dimerise impacts on the ability of C181 to disperse endogenous CIZ1 assemblies.

(Figure contents)

A

665   AcD   MH3   845

GST   C181

EEDEDDDEEEEEEEGEIEAEEE
GST   C181Δ689-709 (ΔAcD)

GYVCQICHKFYDSNSELRLSHCKSLAHFENLQKY
GST   C181Δ746-779 (ΔMH3)

TSSHQPSPQDTVKMPSKVKPGSPGLPPPLRRSTRLKT
GST   C181Δ809-845 (Δ37)

GST   C181Δ689-709,746-779 (ΔAcD,MH3)

GST   C181Δ746-779,809-845 (ΔMH3,37)

B

kDa   Coomassie
35
25
15
10

35   Rabbit Ex17
25
15
10

35   Rabbit Ex16
25
15
10

Rabbit   Rabbit
Ex16   Ex17

C

mAU — Elution volume (mL)

C181
ΔMH3
Δ37
ΔMH3,37

kDa
35
25
Anti CIZ1 (Rabbit Ex17)

D

AU — Molecular weight (kDa) — Elution volume (mL)

C181
ΔMH3

C181   ΔMH3

20 kDa   40 kDa   16 kDa
monomer   dimer   monomer

E

Acidic domain
MH3 domain

F

nM

0   150   600   2500

GAPDH   GAPDH
C181   ΔMH3

82

Figure 3.6. Biochemical analysis of CIZ1 C-terminal fragments

A. CIZ1 fragments used for biochemical analysis. N-terminal glutathione S-transferase (GST) tag allows production of purified protein. Numbers indicate the boundaries respective to the amino acid numbers of the murine full length CIZ1. Domains highlighted are the acidic domain (AcD) and matrin 3-homology domain (MH3). Diagram produced using IBS software (Liu et al., 2015b).

B. SDS-PAGE analysis of CIZ1 C-terminal proteins. Coomassie staining allows visualisation of all protein fragments present. Western blot analysis confirms that all forms in the coomassie are CIZ1, and loss of antibody cross reactivity is based on epitope location in CIZ1 and their subsequent loss in mutants.

C. Size exclusion chromatography (SEC) analysis of CIZ1 mutants. Protein presence was detected using UV absorbance at 280nm. Elution profile is cropped to region containing the peak of interest. Western blot corresponds to C181 elution profile.

D. Size exclusion chromatography multiple laser light scattering (SEC-MALLs) analysis of CIZ1 mutants. Peaks correspond to UV absorbance at 280nm on the left hand y-axis and lines reflect light scattering measurements on the right hand y-axis

E. Alphafold2 (Mirdita et al., 2022) structural modelling of C181 and ΔMH3 illustrating the MH3 as the dimerisation interface.

F. Electrophoretic mobility shift assay (EMSA) analysis of C181 and ΔMH3. Protein was titrated (0-2500nM) with constant RNA concentration (0.8ng). A shift upwards is indicative of a RNA:protein complex that has been retarded compared to protein alone. The RNA is DIG labelled to facilitate visualisation.

3.5. Discussion

3.5.1. A CIZ1 alternative transcription start site?

At this stage we do not know the exact mechanism that is leading to production of C-terminal fragments in early-stage human cancers, and it is an avenue for further exploration. If there was an aberrant splicing event in which exon 10 onwards is spliced onto earlier exons, one might expect to observe a U-shaped profile of expression, and this is not what is observed. Furthermore, in all samples (normal and cancer) we observe reduced exon 2 and 3 expression relative to exon 7, and in cancers a small but significant reduction in exon 5. None of the above is easily explained by known alternative splicing events, so other mechanisms must be considered.

One hypothesis is that an alternative transcription start site (TSS) is preferentially utilised in tumour cells. A recent paper conducted PacBio long read sequencing on a set of breast cancer cell lines and patient samples (Veiga et al., 2022). Utilising their interactive website to monitor what forms of *CIZ1* were identified the presence of two *CIZ1* 3' only transcripts (one starting from exon 8 and another from an intronic region between exon 14 and 15), was seen, as well as the previously identified *CIZ1-F* (Swarts et al., 2018). The group used MCF7 and MDA-MB-231 cells so can be partially compared to our results, but lacks the SK-BR-3 line which had the most extreme 3' elevation in our data. It is not possible to determine the ratio of the amount of each *CIZ1* transcript from the data of Veiga et al. Upon inspection for genomic features of interest in the region of *CIZ1* corresponding to the AD, it was noted that there are 10 predicted potential alternative TSSs (Figure 3.7A) that were identified as part of the FANTOM5 project (Lizio et al., 2015). Additionally, there are changes in the epigenetic landscape in HeLa and MCF7 cells that suggest the upregulation of a promoter or enhancer in the region, based on the presence of active chromatin marks such as H3K4me1,2,3, H3K9ac and H3K27ac, and increased DNase sensitivity (Figure 3.7B). These suggest open chromatin conformation and would be consistent with alternative transcription initiation from this region. Plans to investigate this further include conducting quantitative long read sequencing (Oxford Nanopore) on additional breast cancer lines (in particular the SK-BR-3 line) to attempt to identify the exact location of CIZ1 transcript upregulation, which could then be further explored by chromatin immunoprecipitation to identify if RNA polymerase II is bound in this region in our cells of interest.

Figure 3.7. CIZ1 chromatin changes in cancer

A. Potential CIZ1 alternative transcription start sites (TSSs) in the AD predicted in the FANTOM5 project (Lizio et al., 2015).

B. The epigenetic landscape at the CIZ1 locus in human embryonic stem cells (hESC), human mammary epithelial cells (HMEC), cervical cancer cells (HeLa) and breast cancer cells (MCF7). Numbers correlate to CIZ1 exon number. Histone modifications shown are examples of active marks. Diagram generated using UCSC genome browser (Kent et al., 2002).

3.5.2. Dominant negative effects are routinely observed in cancer

DN mutants are named as such since they are able to disrupt the function of their WT counterparts (Herskowitz, 1987). Previous unpublished work in the Coverley lab has shown that the RD of CIZ1 can interfere with endogenous CIZ1 aggregates in a DN mechanism of action, leading to their dispersal. Here I showed that the AD is also able to exert DN effects. DN effects can be more deleterious than a single allele loss, since for example in the context of a DN mutant that is capable of corrupting a homodimer, only 25% would be functional (25% WT:WT, 50% WT:MUT, 25% MUT:MUT), assuming equal expression levels. This is a more extreme loss compared to a 50% fall with a single allele loss leading to haploinsufficiency (Veitia et al., 2018).

DN effects of other mutant proteins have been described, including p53, the "guardian of the genome" and a key tumour suppressor (Lane, 1992). Mutations in *TP53* (the gene encoding p53) are the most common mutation in cancer (Kandoth et al., 2013), indicating how crucial its role is in homeostasis. DN effects of mutant p53 have long been known, for example an early study identified that patients with a heterozygous p53 mutation displayed less than 25% oligonucleotide binding compared to homozygous WT patients (Srivastava et al., 1993). It was later shown that p53 mutants reduce $p53^{WT}$:$p53^{MUT}$ complexes from binding to target genes and reduce their ability to arrest the cell cycle (Willis et al., 2004), and increases cell invasion and migration (Dong et al., 2007).

DN mutants have also been identified in other genes commonly mutated in cancer such as PTEN (Papa et al., 2014) and ATM (Chenevix-Trench et al., 2002); and scaffold attachment factor A (SAF-A) (Kolpa et al., 2022), a nuclear matrix (NM) protein implicated in *Xist* anchorage at the Xi (Hasegawa et al., 2010, Kolpa et al., 2016). It has been suggested that DN therapeutic peptides that target cancer drivers could be used in the clinic as a treatment strategy (Savinov and Roth, 2021).

3.5.3. The extreme C-terminus is implicated in CIZ1-AD targeting to the nucleus

CIZ1 is known to have a nuclear localisation sequence (NLS) in the RD of the protein to facilitate nuclear import (Coverley et al., 2005), which when mutated perturbed CIZ1 nuclear import (Sofi et al., 2022). Yet since the AD alone is also able to target to the nucleus, this suggests that there are also C-terminal sequences to facilitate this import. The use of databases (Kosugi et al., 2009, Nguyen Ba et al., 2009) that predict NLS sequences did not identify any in C181. However, when utilising software that predicts subcellular location (Nakai and Horton, 1999), C181 was still predicted to localise to the nucleus with high confidence, despite the lack of a predicted NLS. This suggests that C181 nuclear import is via a mechanism that does not utilise a canonical NLS. It can be observed that the mutants of C181 display variable cytoplasmic and nuclear presence in the absence of detergent, in particular Δ37 and I122 have an increased cytoplasmic presence compared to C181, ΔMH3 and ΔNALTAF (Figure 3.5B). This suggested that the extreme C-terminus is implicated in nuclear import. Upon inspection this area is highly enriched for positively charged amino acids (7 lysine and arginine residues, 19% of the region), and lysine and arginine rich regions are implicated in nuclear import of signal transducer and activator of transcription proteins (STATs) (Melen et al., 2001). In addition to this, work conducted by other members of the Coverley lab is suggesting that the extreme C-terminus is required for C-terminal interactions with RNA (data not shown), and RNA recognition motifs (RRMs) have also been implicated in nuclear import (Cassola et al., 2010). Since Δ37 is lacking an additional 22 amino acids at the C-terminus compared to I122, which contains 3/7 of the positively charge amino acids, this could explain why it is more extractable than I122 upon detergent treatment, assuming this would lead to reduced efficiency of nuclear import.

3.5.4. MH3 dimerisation is implicated in C181 anchorage and DN effects

Nuclear import is not the only mechanism to consider, since C181 must form additional interactions in the nucleus to confer resistance to detergent extraction. The ΔMH3 construct is able to target to the nucleus efficiently, but the majority of cells no longer contain ΔMH3 upon detergent treatment (Figure 3.5B,C), suggesting that this domain is key for interactions with insoluble components of the nucleus. The disparity between the large reduction in nuclear retention, but the more conservative abrogation of ability to disperse endogenous CIZ1 from the Xi, suggests that there are multiple interactions at play. Since we have shown that the MH3 domain is required for homodimerization (Figure 3.6D), this could suggest that C181 self-interaction with full length endogenous CIZ1 could be one of the key mechanisms for C-terminal fragment retention upon detergent treatment. However, whilst ΔMH3 has a perturbed ability to disperse endogenous CIZ1, it has not been lost altogether, so there must be additional mechanisms contributing to endogenous CIZ1 dispersal.

The MH3 domain is named as such due to its homology to a region of the NM protein matrin 3 (Belgrader et al., 1991, Mitsui et al., 1999). As stated in the introduction, the function of matrin 3 remains poorly understood. Monitoring matrin 3 localisation in the nucleus identified that it was excluded from both the Xi and perinuclear and perinucleolar heterochromatin (Zeitz et al., 2009), and when matrin 3 is lost there is destabilisation of CTCF and cohesin (Cha et al., 2021). Together, this suggests it plays a role in transcriptional regulation and chromatin architecture in particular regions of the genome. Similar to CIZ1, it has been implicated as both a tumour driver where knockdown led to reduced proliferation in melanoma (Kuriyama et al., 2020), and a tumour suppressor where high levels were correlated with patient survival in breast cancer (Yang et al., 2020a). Given that Matrin 3 and CIZ1 share sequence homology in the MH3 domain one would assume that they could be implicated in similar functions *in vivo*, however too little is known at this stage to attempt to draw clear parallels, and indeed Matrin 3 exclusion from the Xi compared to CIZ1 accumulation at the Xi does not support this argument.

When trying to identify the exact role of the MH3 domain itself, biological processes gene ontology analysis (Ashburner et al., 2000) of proteins containing the MH3 U1-like zinc finger returns roles in RNA processing, which is unsurprising given the links to NM proteins and RNA processing as described in the introduction. It is also noted the MH3 domain has aromatic residues reminiscent of RNA binding domains through stacking interactions with RNA bases, however due to the negative charge it may alternatively be used as a protein interface (Muto et al., 2004). This is in line with our results that the MH3 domain facilitates dimerisation of CIZ1 and has weaker support for RNA binding as seen in the EMSA (Figure 3.6D,F).

3.5.5. Model of C181 DN effects

Considering everything discussed above, I propose that CIZ1 DN fragments interact with endogenous CIZ1 via their MH3 domain to exert their DN functions. Since C181 can interact with endogenous CIZ1, but lacks N-terminal regions required for Xi accumulation, it interferes with the efficiency of accumulation of endogenous CIZ1. This leads to loss of CIZ1 at the Xi, which we hypothesise plays a protective role, and thus an increased susceptibility to loss of epigenetic marks. Upon deletion of the MH3 domain, since it can no longer self-interact with endogenous CIZ1, the DN effects are more limited.
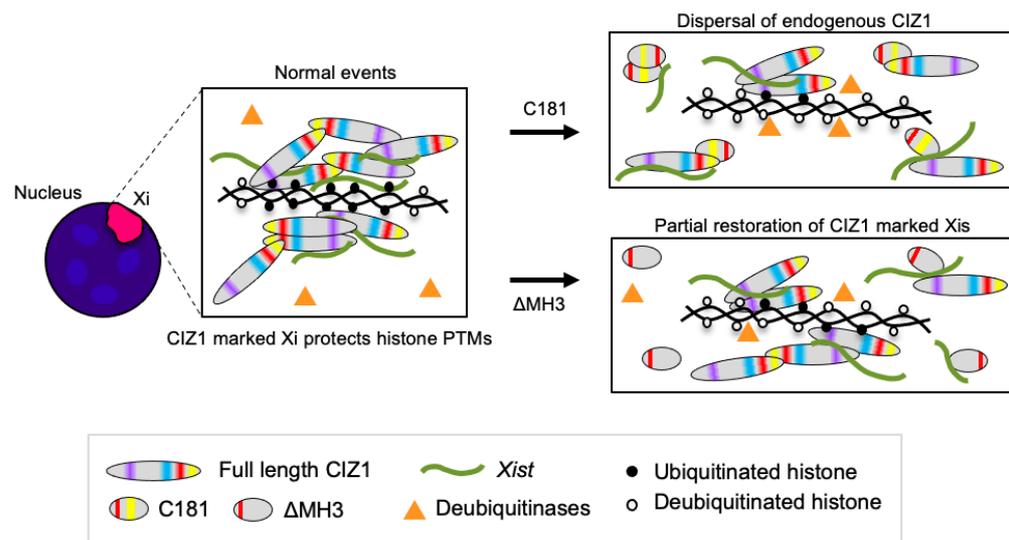
Figure 3.8. The proposed model of C181 DN effects

Considering the role of the RD of CIZ1 for Xi accumulation via prion-like domain (PLD) mediated interactions with *Xist* (Sofi et al., 2022); and CIZ1 AD to immobilise CIZ1 to the NM, possibly via self-interactions and interactions with other NM proteins and nucleic acid, conferred upon it by the MH3 domain and the extreme C-terminus respectively. This suggests that there are likely multivalent interactions that anchor CIZ1 to the NM and the Xi. This could be how the ΔMH3 mutant is still partially able to perturb endogenous CIZ1, since it still contains the extreme C-terminal domain important for RNA interactions, and this could sequester RNA (such as *Xist*) away from the Xi and lead to a weakening of the endogenous CIZ1 complex. If indeed both i) self-interaction with endogenous CIZ1 via the MH3 domain but an absence of the PLDs to enable proper Xi accumulation and ii) sequestering of RNA such as *Xist* that is required for Xi accumulation, need to be absent to completely abolish the DN features of C181, it would be predicted that a ΔMH3,37 double mutant would have no effect on endogenous CIZ1 assemblies and would be useful to test in the future.

3.5.6. Histone ubiquitination is key for genome repression, but information is conflicting

The exact mechanisms by which H2AK119 is ubiquitinated, and its role in transcriptional repression are still being investigated. As described in the introduction, there is conflicting information on the timings of H2AK119ub1 and H3K27me3 deposition in development, and if these repressive marks rely on each other for genome silencing. One study noted that a mutation in RING1B, the active subunit in polycomb repressive complex 1 (PRC1) responsible for deposition of H2AK119ub1, led to reduction in H2AK119ub1 presence. However, if RING1B was bound to chromatin there was still some maintenance of chromatin compaction, suggesting the complex itself, not just its enzymatic activity plays a crucial role (Eskeland et al., 2010). This study noted that in RING1B null cells H3K27me3 was still present, but was insufficient for gene repression. However, a different study observing the effect of the same point mutation in RING1B noted that H2AK119ub1 loss led to loss of H3K27me3 (Tamburri et al., 2020). In our study in the context of a transient transfection of DN fragments we observed that H3K27me3 was not reliant on the H2AK119ub1 mark, at least during the first cell cycle (Figure 3.4F). In addition to this there are likely differences in PRC1/2 function in development compared to adult homeostasis, as observed by the difference in RING1B target genes in induced pluripotent stem cells compared to normal and cancer cell lines (Chan et al., 2018), so this must be considered when comparing literature and attempting to deconvolute mechanisms.

Focusing on the mechanisms of H2AK119ub1 mediated repression, early work showed that there was histone "cross talk" and H2AK119ub1 repressed the presence of the active modification H3K4me3 at gene promoters (Nakagawa et al., 2008). It has also been identified that H2AK119ub1 stabilises nucleosomes and "bolts" in DNA to prevent it from peeling away from histones (Xiao et al., 2020). Additionally, it is noted that DNMT1 binds to H2AK119ub1, thus suggesting ubiquitination is involved in the maintenance of DNA methylation (Qin et al., 2015). Finally, H2A ubiquitination recruits BRCA1 to DNA damage lesions, and where this ubiquitination is lost BRCA1 recruitment was abrogated (Wu et al., 2009). This could perturb the DDR and lead to accumulation of mutations and tumourigenesis. Considering this, the ubiquitination of histones is clearly implicated in several mechanisms of genome repression and maintenance, and multiple bodies of work have highlighted the importance of CIZ1 in H2AK119ub1 accumulation at the Xi (this work and (Stewart et al., 2019, Sofi et al., 2022)).

3.5.7. Concluding remarks

This chapter presents evidence to suggest the existence of previously unidentified DN fragments of CIZ1 in breast cancer. DN fragments of CIZ1 have previously been described to alter the subnuclear location of CIZ1 (Rahman et al., 2007), and other CIZ1 splice variants could also exert their role in cancer via a DN mechanism (Higgins et al., 2012). However, those fragments of CIZ1 are generated by alternative splicing, and as discussed above I am postulating that C-terminal over expression could be the result of an internal TSS, which is novel. Since DN are a feature of disease, it is pertinent to understand their role and how they are created, since this could eventually facilitate targeting them for treatments, or alternatively their exploitation in early cancer detection. Here it is suggested that overexpression of CIZ1 DN fragments leads to epigenetic dysregulation via loss of H2AK119ub1, which could lead to downstream aberrant gene expression changes.

I have also attempted to deconvolute the roles of domains in the AD of CIZ1. Previous work demonstrated that the transient transfection of the AD of CIZ1 was retained in the nucleus after treatment with detergent, high salt and nuclease, whilst the N-terminal replication domain (RD) was removed (Ainscough et al., 2007). This highlighted the key role that the C-terminus plays in nuclear immobilisation, and since it is resistant to all these treatments implicates CIZ1 as a component of the NM (Berezney and Coffey, 1974). Here I showed that the extreme C-terminus was implicated in nuclear import, but additionally that the MH3 domain is crucial for immobilisation. This immobilisation is suggested to be via CIZ1 self-interaction, since loss of MH3 domain leads to impaired ability to homodimerise. In addition to this, previous work has shown that for CIZ1 accumulation at the Xi the low complexity PLDs in the RD must be present (Sofi et al., 2022).

Together, this shows that multiple mechanisms and interactions that allow CIZ1 to i) localise to the nucleus, ii) immobilise itself in the nucleus and iii) form large assemblies at the Xi, and raises the possibility that corruption of any of these steps might contribute to epigenetic instability in cancer.

4. Regulation of CIZ1 assembly formation and function During the Cell Cycle

4.1. Introduction

4.1.1. Kinases in the cell cycle

For cells to grow and replicate they must pass through all stages of the cell cycle correctly. This is facilitated by cyclin dependent kinases (CDKs), which are activated by mitogens or inhibited at checkpoints if errors are detected (Barnum and O'Connell, 2014). The CDK complexes implicated at different stages of the cell cycle, and CIZ1 involvement in their roles have been described in the Introduction. This chapter focuses on kinase regulation in mitosis, and whilst CDK involvement (specifically CDK1, (Malumbres, 2014)) is key, several other kinase families are also fundamental, including but not limited to, the Polo, Aurora and NIMA family (Nigg, 2001).The Aurora kinase family is made up of three members: Aurora kinase A, Aurora kinase B and Aurora kinase C. Aurora kinase A and B are associated with the earlier and later stages of mitosis respectively, whereas Aurora C is less understood and is associated with meiosis (Carmena and Earnshaw, 2003). Since the role of Aurora kinase B (AURKB) in disassembly of *Xist* from the Xi during miosis has been observed (Hall et al., 2009), it has become the focus of my experimentation, and Aurora kinase A and C will not be discussed further.

4.1.2. The role of Aurora Kinase B in mitosis

AURKB expression is upregulated in mitosis (Kimura et al., 2004), in contrast to CDK expression which is consistent throughout the cell cycle but activated by their cyclin counterparts and phosphorylation events (Obaya and Sedivy, 2002). AURKB is localised at the centromeres of chromosomes from prophase to partway through anaphase, and after this it becomes concentrated at the spindle midzone (Murata-Hori et al., 2002). During prophase to metaphase AURKB phosphorylates histone 3 serine 10 and serine 28 (H3S10 and H3S28) in condensing chromosomes (Goto et al., 2002), and this phosphorylation of H3S10 is associated with heterochromatin protein 1 (HP1) dissociation from chromosomes (Hirota et al., 2005). Loss of AURKB via dominant negative (DN) mutant overexpression or small molecule inhibition, revealed that AURKB is required for kinetochore attachment to microtubules (Murata-Hori and Wang, 2002) and to maintain checkpoint signalling during syntelic attachment (both sister chromatids interact with microtubules from the same spindle pole) (Hauf et al., 2003) respectively.

More recently it has been suggested that there is crosstalk between AURKB and CDK1 in anaphase that ensures correct chromosome separation (Afonso et al., 2019), and that

the two kinases work together to regulate kinetochore function in anaphase and subsequent disassembly in telophase (Papini et al., 2021). Indeed the heavy chain of non-muscle myosin II (NMII) has been identified as an AURKB substrate, where phosphorylation of the tail domain of NMII prevents filament formation to promote disassembly, and additionally phospho-resistant serine to alanine mutations of NMII led to the formation of long intercellular bridges indicative of aberrant cytokinesis (Babkoff et al., 2021). In the final stages of mitosis when AURKB is no longer required, it is degraded by the E3 ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C) coupled with the activator protein cdc20 homolog 1 (Cdh1) (Stewart and Fang, 2005).

4.1.3. CIZ1 removal from chromatin in mitosis

Prior to my PhD the Coverley lab monitored CIZ1 loss in mitosis. Specifically, CIZ1 accumulation was reduced in the prometaphase-metaphase stages and lost entirely by anaphase in murine cells (Ridings-Figueroa et al., 2017) (Figure 4.1). Since it was already known that *Xist* Xi accumulation is lost in mitosis (Hall et al., 2009), this is perhaps unsurprising. Indeed, more recent work has shown that when CIZ1 is absent *Xist* is dispersed across the nucleus, rather than localised to the Xi (Ridings-Figueroa et al., 2017, Sunwoo et al., 2017), and this interaction is mediated by a direct interaction between *Xist* and the prion like domains (PLDs) in CIZ1 (Sofi et al., 2022). Considering this I wanted to further investigate the changes in CIZ1 status during mitosis, and then identify if there was a link between normal CIZ1 loss in mitosis and the DN effect of C-terminal fragments of CIZ1 described in Chapter 3.
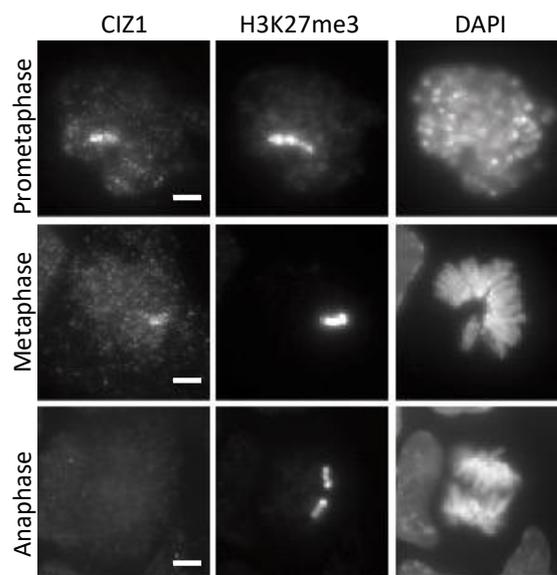
Figure 4.1. CIZ1 assembly loss in mitosis
Figure taken from (Ridings-Figueroa et al., 2017). CIZ1 and H3K27me3 status throughout mitosis was measured via immunofluorescence, identifying CIZ1 loss and H3K27me3 retention. Scale bar is 5µm.

4.1.4. Scaffold attachment factor A

The scaffold attachment factor A (SAF-A/HRNPU) is a nuclear matrix (NM) protein that possesses RNA and DNA binding capabilities (Fackelmayer et al., 1994). SAF-A has been shown to interact with active chromatin, and upon SAF-A deletion in murine hepatocytes chromatin condensation and active and inactive compartment switching was observed, suggesting it plays a role in genome organisation and chromatin architecture (Fan et al., 2018). Additionally, there is an ongoing debate on its requirement for *Xist* anchorage at the Xi (Hasegawa et al., 2010, Kolpa et al., 2016), and it has been observed that SAF-A is also removed from chromatin during mitosis in human cells (Sharp et al., 2020). All this together makes it of interest to us when considering CIZ1 function, so analysis of SAF-A during mitosis in murine cells has also been conducted in this chapter.

4.2. Aims

- Confirm CIZ1 loss in mitosis and profile CIZ1-Xi assembly reformation in G1
- Test the effects of inhibition of AURKB on the disassembly of endogenous CIZ1 and SAF-A during mitosis
- Validate the new C181 lentivirus expression system and use it to monitor endogenous CIZ1 status in different stages of the cell cycle in cells expressing CIZ1 DN fragments

4.3. Experimental Design

4.3.1. Measuring CIZ1 Xi status at different stages of the cell cycle

Cells were arrested in mitosis (M phase) and synthesis phase (S phase) by application of nocodazole and thymidine respectively. Nocodazole disrupts microtubule assembly/disassembly dynamics, thus blocking the formation of the metaphase spindles and leading to arrest in prometaphase. The thymidine perturbs DNA replication by acting as a competitive inhibitor for the deoxyribonucleotide metabolism pathway, thus cells are arrested early into S phase as the levels of deoxyribonucleotides are depleted. Thymidine was applied for 24 hours, and nocodazole was applied for 16-24 hours. Cells arrested in M phase were isolated by mitotic shake off and replated for analysis post release. Cells held in S phase were released by washing twice with PBS then replacing with fresh media. Measurement of 5-ethynyl-2'-deoxyuridine (EdU) incorporation post release from S phase facilitated confirmation of enrichment. Cell cycle arrest experiments of cells transduced with C181 was conducted 48-72 hours post transduction. To facilitate visualisation of miotic cells, cells were fixed prior to

permeabilisation, with the exception of this change to the standard protocol, all downstream processing was as described in the Materials and Methods.

4.3.2. Manipulating CIZ1/SAF-A with barasertib

Two different concentrations of barasertib was applied to cells for 4 hours and cells were subsequently collected for immunofluorescence imaging. To facilitate recovery of miotic cells, cells were fixed prior to permeabilization. Except for this change to the standard protocol, all downstream processing was as described in the Materials and Methods.

4.3.3. Lentiviral transduction system

To facilitate extended culturing of cells expressing CIZ1, and a higher efficiency of expression in the population, a lentiviral transduction system was developed by Dr Emma Stewart in the Coverley lab (described in full in the Materials and Methods Chapter). In summary, a bicistronic plasmid encoding reporter ZsGreen and C181, or a negative control plasmid encoding only ZsGreen (referred to as "Empty"), was transfected into human embryonic kidney (HEK) cells, alongside other components to facilitate lentivirus production. This lentivirus was subsequently applied to PEFs or the murine D3T3 cell line, leading to transduction and stable expression of ZsGreen alone or ZSGreen and C181. The two proteins are produced as separate entities, which facilitates the user to identify transduced cells, but does not provide information of C181 subcellular location. Samples were collected over several days to validate the system, and to measure downstream changes as a result of C181 expression (Figure 4.2).
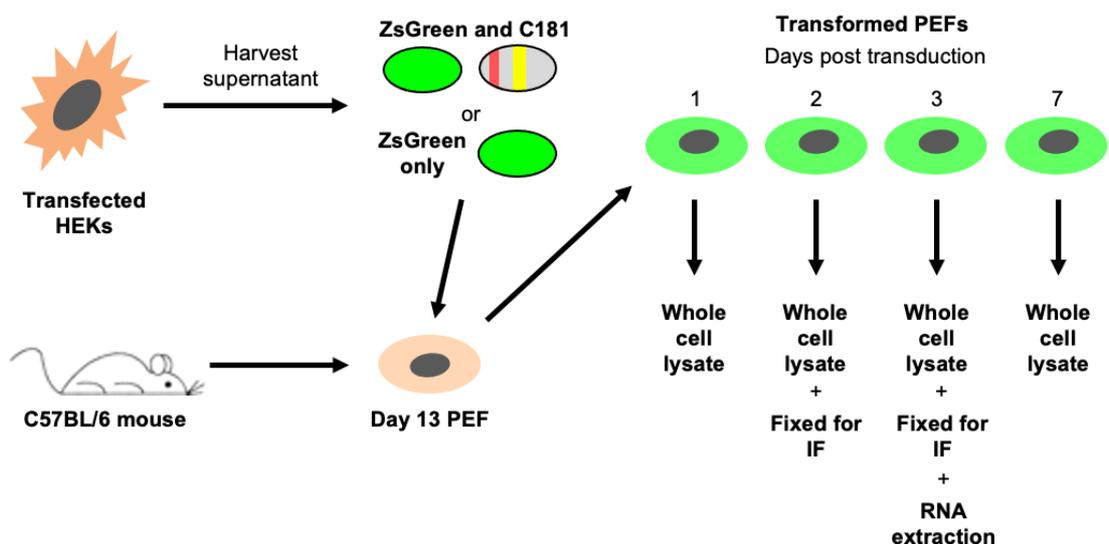
Figure 4.2. Experimental design of lentivirus transduction experimentation
C57BL/6 mice embryos were used to generate day 13 primary embryonic fibroblasts (PEFs). Human embryonic kidney cells (HEKs) were used to produce lentiviral particles for transduction of PEFs. Samples were collected over several days and included: whole cell lysates for western blot analysis, cells fixed for immunofluorescence (IF) processing, and cells collected in trizol for RNA-seq analysis.

4.4. Results

4.4.1. Assembly and disassembly of CIZ1 at Xi during mitosis

Typically 80-90% of interphase cells in a cycling population contain a clear CIZ1 assembly at Xi (Ridings-Figueroa et al., 2017). Since we know that CIZ1-Xi assemblies are lost in mitosis (Figure 4.1), we postulated that interphase cells that do not contain an observable CIZ1-Xi (10-20%) are in early G1 and yet to rebuild their CIZ1 assembly. However, this had not been formally tested. To measure this, I arrested and released a population of female D3T3 cells from M phase or S phase, and monitored endogenous CIZ1-Xi status over time. S phase enrichment was confirmed by incorporation of 5-ethynyl-2'-deoxyuridine (EdU) into newly synthesised DNA, where an increase in cells with incorporated EdU is reflective of more cells undergoing DNA replication. This identified that an average of 76% of cells were in S phase after 24 hours of arrest and 1 hour release from thymidine, in comparison to 51% in the cycling (not arrested) population (Figure 4.3A). This enrichment was not complete since only one round of thymidine was applied, and this might have been improved with a double (consecutive cell cycle) synchrony protocol. Additionally, since thymidine is slow to block cells in S phase, some cells can progress quite far through DNA replication before they arrest. Therefore, after release some cells may have already progressed to G2. Monitoring CIZ1-Xi status post-release from S phase arrest showed no changes in CIZ1-Xi assembly frequency across the time course up to 5 hours, with 86-96% of cells containing CIZ1-Xi assemblies (Figure 4.3B).

Enrichment of cells held in mitosis does not require EdU staining, and can be easily visualised via light microscopy (Figure 4.3C). Upon exit from M phase only 6% of cells contained a CIZ1-marked Xi at 1 hour post-mitosis, and this increased throughout the time course to 78% at 5 hours post-mitosis (Figure 4.3D). If additional measurements had been conducted after 5 hours this number might have increased further, however it had begun to flatten between 4 and 5 hours post-mitosis. Failure to achieve 100% could be related to the use of drugs to arrest the cell cycle, which can be highly deleterious possibly affecting the mechanisms responsible for regeneration of CIZ1-Xi assemblies. Moreover, established cell lines are highly heterogenous so a sub-population may actually lack two X chromosomes. We therefore chose to put weight on the dynamics of recovery rather than absolute values, and conclude that CIZ1-Xi assemblies are normally absent after mitosis and rebuilt in the first three hours of G1.

Interestingly maximum intensity fluorescence measurements (using FIJI as described in Chapter 3) did not identify increasing CIZ1-Xi presence (Figure 4.3E). However, minimum fluorescence intensity revealed a fall between 2 and 4 hours post-mitosis (Figure 4.3F). Example pictures of cells used in analysis can be seen in Figure 4.3G.

Since one bright pixel corresponding to a very small sub-nuclear CIZ1 assembly could present with the same maximum fluorescence intensity as a cell with a classical large accumulation of CIZ1 at the Xi, this could explain why these measures cannot be used in this experimental context. Minimum fluorescence intensity measures across the timepoints appear to be more informative. The decrease could suggest that early in the rebuilding process soluble CIZ1 forms concentrate into sub-Xi supramolecular complexes (SMACs). Then as time progresses more soluble CIZ1 appears to be recruited with the majority eventually targeted to the Xi. Thus, minimum intensity measures appear to report on soluble CIZ1 and its conversion into SMACs (Figure 4.3H).

Alongside measurements of CIZ1-Xi status, the status of the repressive mark H3K27me3 was also measured in both groups. Since this is not normally lost in mitosis (Ridings-Figueroa et al., 2017) (Figure 4.1), we hypothesised that it should be present in most cells upon release in both groups. Indeed, as expected there were no changes to H3K27me3 in either stage of the cell cycle (Figure 4.3I,J). However more cells contained H3K27me3 enrichment upon exit from S phase, compared to exit from mitosis (92% compared to 82%). This could suggest that whilst H3K27me3 is not dependent on CIZ1 presence for enrichment on Xi chromatin, the additional presence of CIZ1 stabilises its presence.

A

DNA EdU

Cycling

Non S

S

1 hour post S release

Non S

S

B

CIZ1 Marked Xi (% cells)

ns

N=1(4) n=394
N=1(2) n=198
N=1(2) n=221
N=1(2) n=179
N=1(4) n=335

Time post S phase (hours)

C

Cycling

Nocodazole treated

D

CIZ1 Marked Xi (% cells)

***

N=1(4) n=414
N=1(3) n=262
N=1(3) n=284
N=1(3) n=356
N=1(4) n=453

Time post mitosis (hours)

E

Maximum fluorescence intensity

2   3   4 hours
n=19 n=18 n=24

F

Minimum fluorescence intensity

2   3   4 hours
n=19 n=18 n=24

G

2 hours   3 hours   4 hours

DNA Endogenous CIZ1

H

Early Xi accumulation

Later Xi accumulation

CIZ1 Xi regeneration

Soluble CIZ1
CIZ1 SMACs

Soluble CIZ1

CIZ1 SMACs

Early G1   Later in G1

I

H3K27me3 Xi Enrichment (% cells)

N=1(2) n=177
N=1 n=106
N=1 n=113
N=1 n=99
N=1(2) n=182

Time post mitosis (hours)

Merge

H3K27me3

Endog. CIZ1

1 hour

H3K27me3

Endog. CIZ1

5 hours

J

H3K27me3 Xi Enrichment (% cells)

N=1(2) n=209
N=1 n=93
N=1 n=111
N=1 n=91
N=1(2) n=122

Time post S phase (hours)

Merge

H3K27me3

Endog. CIZ1

1 hour

H3K27me3

Endog. CIZ1

5 hours

98

Figure 4.3. CIZ1 status post-mitosis and S phase

A. Example immunofluorescence pictures of D3T3 cells with and without EdU incorporation in a cycling (left) and 1 hour release from S phase (right) population, during a 30 minute pulse. The proportion of cells positive for EdU incorporation indicative of presence in S phase, and cells negative for EdU indicative of non-S phase cells in each population is shown in the pie chart. S phase enrichment confirms thymidine efficacy. Scale bar is 10µm.

B. Percentage of cells containing a CIZ1 marked Xi post release from S phase. N=biological replicates and within this the total number of independent counts, n=nuclei scored. Comparison of the proportion of cells containing a CIZ1 marked Xi between 1 and 5 hours post release from S phase was conducted using a t-test, p=0.45.

C. Light microscopy pictures of cycling (left) and nocodazole treated (right) D3T3 cells. Round and refractive mitotic cell enrichment confirms nocodazole efficacy. Scale bars are 400µm.

D. Percentage of cells containing a CIZ1 marked Xi post release from M phase. N=biological replicates and within this the total number of independent counts, n=nuclei scored. Comparison of the proportion of cells containing a CIZ1 marked Xi between 1 and 5 hours post release from M phase was conducted using a t-test, p=1.15×10^{-6}

E. Maximum fluorescence intensity measures of cells 2, 3 and 4 hours post M phase release. n=nuclei analysed. The total unrepresented number of outliers are 2, 3 and 2 for 2, 3 and 4 hours respectively.

F. Minimum fluorescence intensity measures of cells 2, 3 and 4 hours post M phase release. n=nuclei analysed.

G. Example pictures of cells used for analysis in D, E and F. Time labels correspond to time post M phase release. DAPI stain used to visualise DNA (blue), CIZ1 N-term 1794 Ab used to visualise endogenous CIZ1. Scale bar is 10µm.

H. Suggested model of conversion of soluble CIZ1 to CIZ1-SMACs and Xi accumulation over time, post release from M phase.

I. Left: Percentage of cells containing H3K27me3 Xi enrichment post release from M phase. N=number of independent counts, n=nuclei scored. Right: Example pictures of cells used for analysis in D and I. Time labels correspond to time post M phase release. Blue DAPI stain used to visualise DNA, green CIZ1 C-term mAb 87 used to visualise endogenous CIZ1, and red H3K27me3 Ab used to visualise endogenous Xi H3K27me3 accumulation. Scale bar is 10µm.

J. Left: Percentage of cells containing H3K27me3 Xi enrichment post release from S phase. N=number of independent counts, n=nuclei scored. Right: Example pictures of cells used for analysis in B, J. Time labels correspond to time post S phase release. Blue DAPI stain used to visualise DNA, green CIZ1 C-term mAb 87 used to visualise endogenous CIZ1, and red H3K27me3 Ab used to visualise endogenous Xi H3K27me3 accumulation. Scale bar is 10µm.

4.4.2. AURKB is implicated in regulation of CIZ1-Xi assemblies in mitosis

During prophase, AURKB drives *Xist* dissociation from the Xi (Hall et al., 2009), and scaffold attachment factor A (SAF-A) removal from chromatin (Sharp et al., 2020). Therefore, we postulated that it could also be implicated in CIZ1-Xi disassembly at this stage of the cell cycle. I reviewed the CIZ1 amino acid sequence for the AURKB consensus phosphorylation site R/KS/T, R/KXS/T or R/KXXS/T (Kettenbach et al., 2011), and identified 23 conserved potential sites in human and murine CIZ1 (Figure 4.4A), with 8 falling in the N-terminal replication domain and 15 in the C-terminal anchor domain. They appear to cluster in groups, possibly suggesting redundancy and highlighting their likely functional relevance to both functional domains of CIZ1.

In murine cells under normal conditions, CIZ1-Xi assemblies are lost when chromosomes are aligned on the metaphase plate (Figure 4.1). I therefore measured CIZ1 retention post metaphase after exposure to barasertib, a well-characterised AURKB inhibitor that shows promise as an effective cancer therapeutic (Umene et al., 2013, Helfrich et al., 2016, Bertran-Alamillo et al., 2019). This analysis was initially conducted in D3T3 cells then repeated in PEFs, and included analysis of SAF-A status. CIZ1 Xi retention increased in both cell types with increasing concentrations of barasertib (Figure 4.4B, top panel third across). In post metaphase untreated cells, few displayed CIZ1 Xi retention as expected, however after application of barasertib CIZ1 was retained in ~80% D3T3 cells and PEFs at the higher concentration (Figure 4.4B, top panel fourth across). However extensive cell death was also observed, indicating toxicity of AURKB inhibition. As documented by Sharp et al. we also observed SAF-A loss early in mitosis (Figure 4.4B, bottom panel second across), and treatment with barasertib led to some increased retention. However, in murine cells SAF-A was still lost by prometaphase/metaphase even at the highest concentration (Figure 4.4B, bottom panel third across), whereas Sharp et al. retained SAF-A in prometaphase. Example pictures of PEFs with and without treatment of barasertib are shown in Figure 4.4C.

Further analysis into AURKB regulation of CIZ1 is currently being undertaken in the Coverley lab, including a mutation screen of serines and threonines to unphosphorylatable alanine and phosphomimetic aspartate or glutamate in the extreme C-terminus of CIZ1.

A

B

Interphase / Prophase / Prometaphase/Metaphase / Anaphase/Telophase

CIZ1 D3T3   CIZ1 PEF   SAF-A D3T3   SAF-A PEF

Barasertib concentration (µM)

C

Interphase   Prophase   Prometaphase/Metaphase   Anaphase/Telophase

Untreated

0.1 µM Barasertib
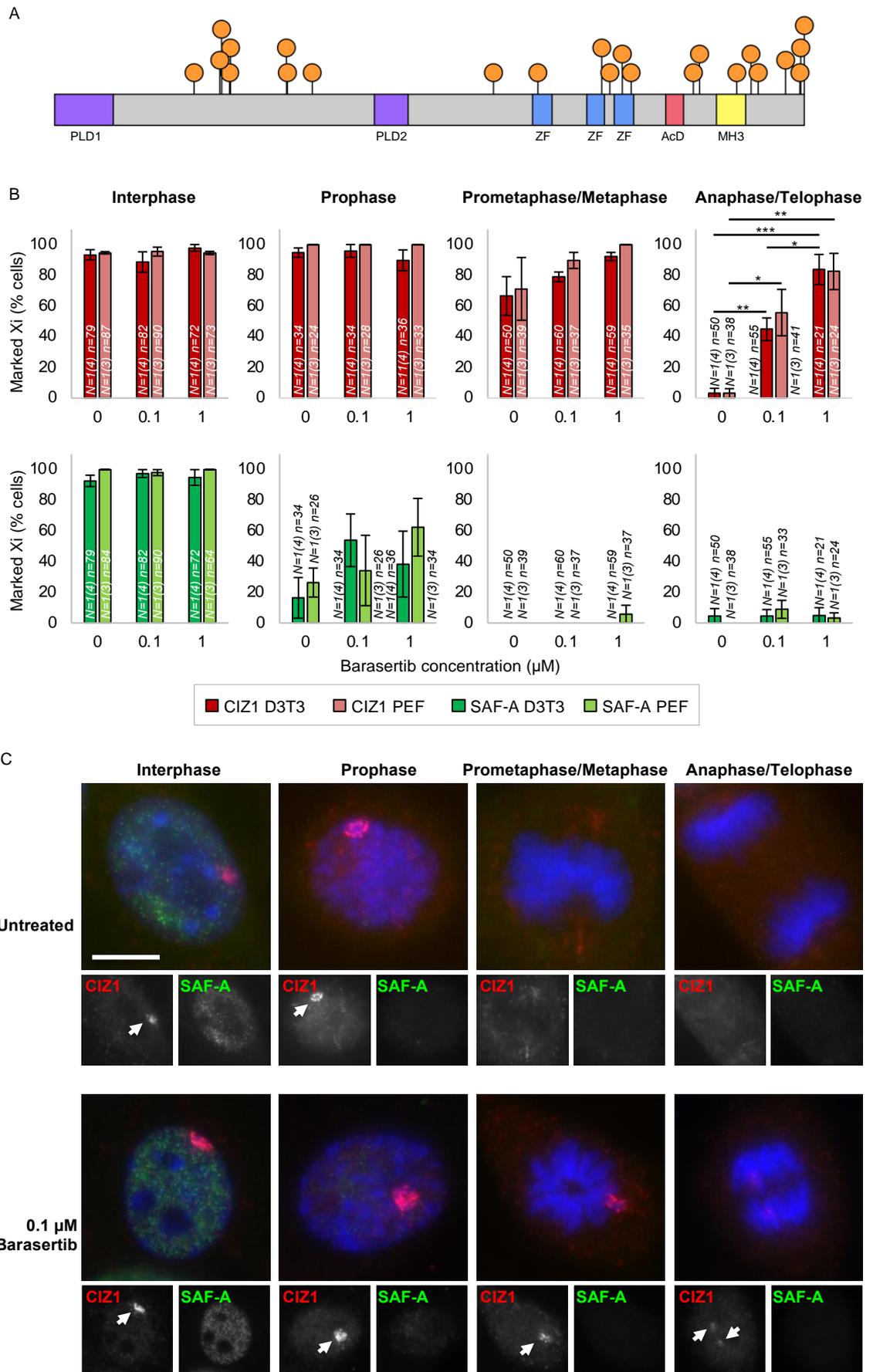
Figure 4.4. CIZ1 manipulation with AURKB inhibitor

A. Diagram of AURKB consensus sites in CIZ1. Sites labelled are sites that are conserved between human and murine CIZ1. The sites relative to murine CIZ1 (UniProt ID Q8VEH2) are: S158, T187, S189, S198, S199, S262, S263, T291, T495, S545, T617, S626, T640, S650, S720, S727, S769, S785, S793, S824, S840, S841, T845. Diagram produced using IBS software (Liu et al., 2015b).

B. Percentage of cells containing a CIZ1 marked Xi (top) or SAF-A chromatin enrichment (bottom). Comparison of CIZ1/SAF-A status at different stages of mitosis with and without the presence of barasertib allows observation of changes upon AURKB inhibition. N=biological replicates and within this the total number of independent counts, n=nuclei scored. Error bars show SEM. Statistical analysis of the proportion of cells with a CIZ1 marked Xi in the anaphase/telophase stage was conducted using a One-way ANOVA with a Tukey post hoc test within each cell type. In PEFS 0 to 0.1µM p=0.037 and 0 to 1µM p=0.0058; in D3T3 cells 0 to 0.1µM p=0.0075, 0 to 1µM p=7.3x10^{-5} and 0.1µm to 1µm p=0.011. Where states are not shown there was no statistical differences observed in the dataset, and this was omitted to reduce clutter on the graphs.

C. Example pictures of PEFs treated with barasertib. Blue DAPI stain used to visualise DNA, red CIZ1 N-term 1794 Ab used to visualise endogenous CIZ1 Xi accumulation, and green SAF-A Ab used to visualise endogenous SAF-A enriched on chromatin. All Ab information available in the Materials and Methods. Scale bar is 10µm.

4.4.3. C181 only disperses in cycling cells

Since CIZ1-Xi assembles and disassembles in a cell-cycle dependent manner, I wanted to test if C181 is limited in *when* it can exert its dominant negative (DN) function. Initially, cycling and newly contact inhibited cells were transfected with C181, and the status of endogenous CIZ1-Xi quantified at the level of frequency. This revealed that C181 only had a DN CIZ1-Xi dispersal effect in cycling cells (Figure 4.5A), with no effect on endogenous CIZ1 status in contact inhibited cells. This, and the lack of complete effect, suggests that cells are susceptible at a particular stage of the cell cycle. Example pictures of cycling and contact inhibited cells can be seen in Figure 4.5B. Measurement of cell cycle status by flow cytometry confirmed that cells were enriched in G0 when contact inhibited, or displayed G1, S and G2 peaks when cycling (Figure 4.5C).
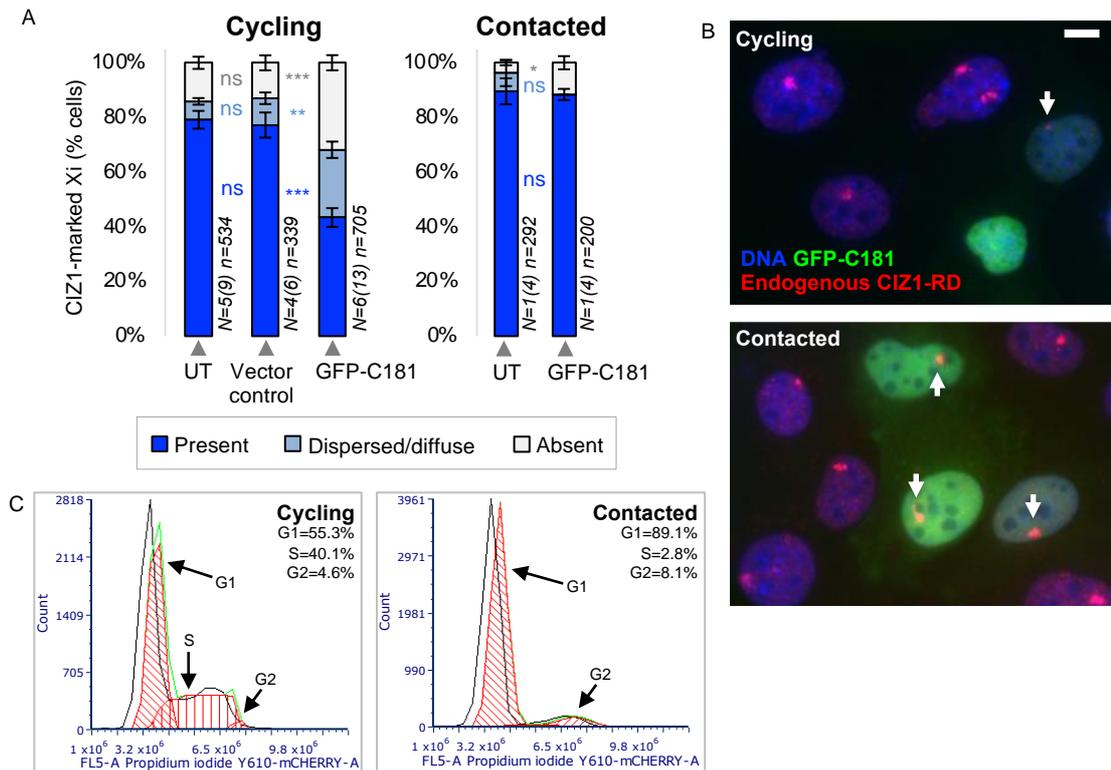
Figure 4.5. Cell cycle dependence of C181 DN effect

A. Dispersal of CIZ1-Xi in cycling and quiescent cells. Left: Data in cycling cells was previously presented in Chapter 3, Figure 3.4. Statistical analysis of the proportion of cells in each category to compare dispersal capabilities was conducted using a One-way ANOVA with a Tukey post hoc test. UT and GFP empty vector: CIZ1 marked Xi, p=0.94; dispersed/diffuse CIZ1 patch, p=0.71; no CIZ1 marked Xi, p=0.97. GFP empty vector and C181: CIZ1 marked Xi, p=4x10-6; dispersed/diffuse CIZ1 patch, p=0.0015; no CIZ1 marked Xi, p=0.00016. Right: Parallel analysis of C181 dispersal conducted in non-cycling contact inhibited cells. Statistical analysis of the proportion of cells in each category to compare dispersal capabilities was conducted using a t-test. Untreated (UT) and C181: CIZ1 marked Xi, p=0.098; dispersed/diffuse CIZ1 patch, p=0.34; no CIZ1 marked Xi, p=0.043. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars on both graphs show SEM.

B. Example pictures of cycling and contacted cells used in analysis. Blue DAPI stain used to visualise DNA, green indicates GFP tagged C181, and red CIZ1 N-term 1794 Ab used to visualise endogenous CIZ1. Scale bar is 5µm.

C. Flow cytometry analysis. Cycling cells (left) contain a G1 and G2 peak, and an S phase population between the G1 and G2 population. Contacted cells have a diminished G2 and S peak and enrichment of a G1/G0 peak indicative of a quiescent population. DNA content measured using propidium iodide.

4.4.4. C181 expression is confirmed in the lentiviral transduction system and this leads to dispersal of endogenous CIZ1 and loss of heterochromatin marks

Due to the dilution that occurs upon cell division, transient transfections cannot be used to measure long-term effects of C181. Therefore, I shifted further experimentation to a lentivirus transduction system to explore C181 DN effects in the cell cycle. The experimental workflow and sample collection process is shown above (Figure 4.2). To validate expression, live fluorescence pictures of the viral vector production cells (HEK cells) were taken at the point of viral harvest, confirming the presence of ZsGreen in the empty negative control and C181 system (Figure 4.6A). After viral application to PEFs, live imaging 48 hours later revealed ZsGreen positive cells in the population, confirming transduction of both the empty vector control, and C181 populations, and absence in the untreated (UT) population (Figure 4.6B). Analysis of whole cell lysates extracted from transduced cells showed ectopic CIZ1-C181 expression increasing up to 7 days (Figure 4.6C). It also revealed inconsistency between the histone and actin loading controls in the C181 transduced population, reminiscent of those seen in Chapter 3. Prior to sample collection, it was observed that there was increased cell death in the C181 transduced population compared to the empty control (data not shown). Since samples are initially routinely loaded with matched volumes relative to the total volume, actin appears to be the most reliable loading control in these experiments. The reduction in the total protein present in the C181 samples, reflective of cell death, was also observed in the Ponceau S stain, providing further support for the validity of the actin result. All downstream immunofluorescence analysis was conducted using PEFs or D3T3 cells at specified times post transduction. Since we know that cells are only susceptible to the DN effect of C181 whilst cycling, time points were adjusted to take into account cell density.

Once C181 transduction in PEFs was established, we set out to monitor its effect on endogenous CIZ1, and the heterochromatin repressive marks that are dependent on CIZ1, as described in Chapter 3. This experimentation was a collaboration between myself and Emma Stewart in the Coverley lab, and is part of a large body of work investigating other CIZ1 fragments. Here, only the C181-related data is presented. As seen in Chapter 3, expression of C181 caused a reduction in cells containing an observable CIZ1 marked Xi, compared to the negative control vector (Figure 4.7D), and this coincided with H2AK119ub1 loss (Figure 4.7E). Interestingly upon C181 transduction we also observed a reduction in Xi's enriched for H3K27me3 Xi (Figure 4.7F), which contrasts our observation in Chapter 3. It is notable that transient expression experiments were analysed after 24 hours and transduction experiments after 3 days, which would be consistent with delayed loss of H3K27me3 compared to H2AK119ub1. This suggests that there could be stepwise events of de-repression occurring upon CIZ1

loss, with H2AK119ub1 loss occurring more quickly, whereas H3K27me3 loss is more gradual.
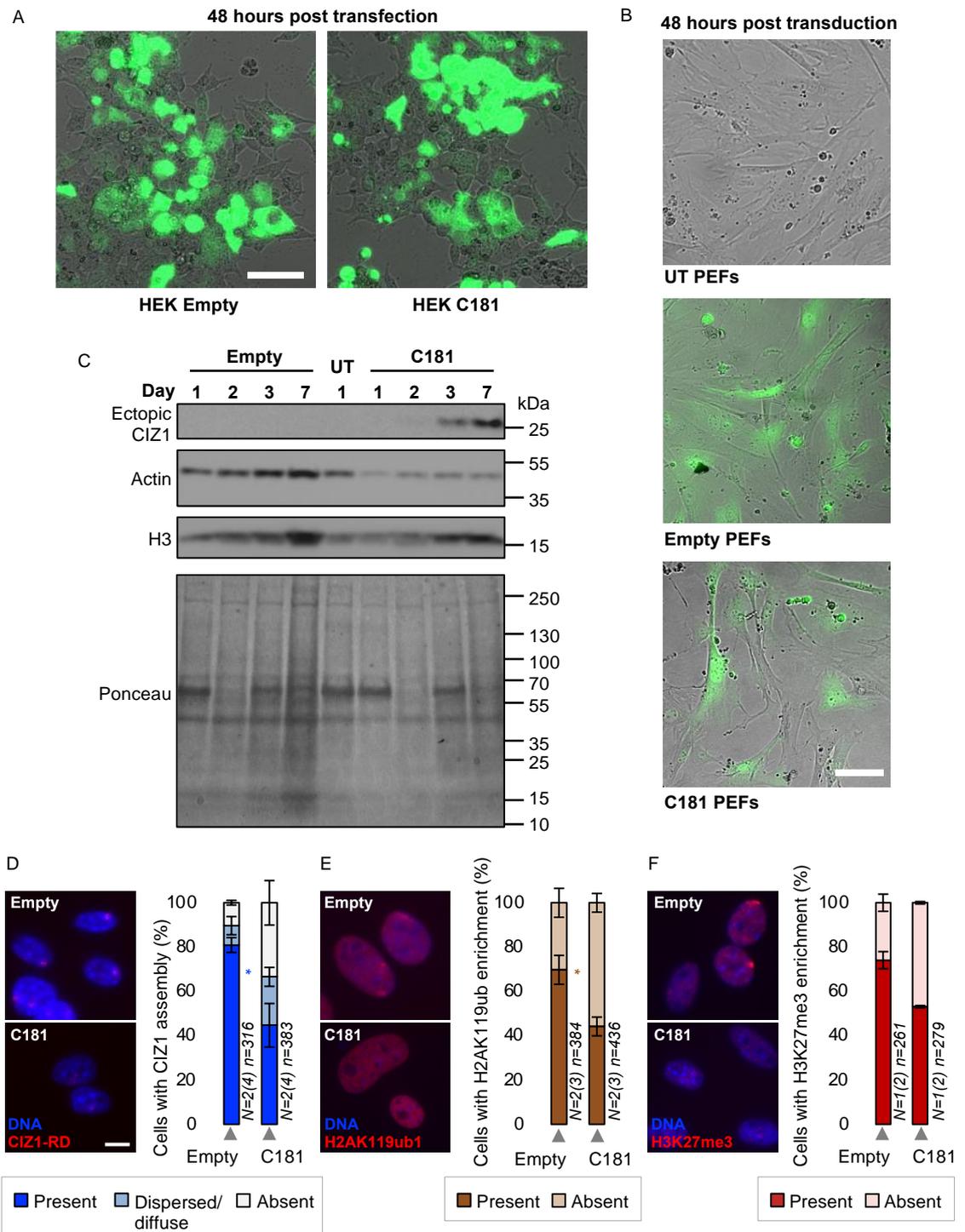
Figure 4.6. Lentivirus transduction system validation and early analysis

A. Human embryonic kidney (HEK) cells 48 hours post transfection with a plasmid containing ZsGreen only (empty) or a plasmid containing ZsGreen and C181 (C181).

B. Primary embryonic fibroblasts (PEFs) 48 hours post transduction with ZsGreen only (empty) or ZsGreen and C181 (C181). Untransduced (UT) cells also included to illustrate lack of green fluorescence in control cells. All images in A and B are to the same scale, scale bar is 50µm.

C. Western blot analysis of PEF whole cell lysates collected from day 1-7 post transduction. Ectopic CIZ1 (C181) is observed faintly from day 2 and clearly on day 3 and 7. C181 theoretical molecular weight is 20kDa but is routinely observed to at approximately 25kDa (Chapter 3 and 5). Rabbit Ex17 CIZ1 Ab used for analysis. All Ab information is documented in the Materials and Methods.

D. Frequency of endogenous CIZ1-Xi aggregates in PEFs that are transduced with the negative control (empty) or C181. Statistical analysis of the proportion of cells in the present category to measure differences in dispersal capability was conducted using a t-test, p=0.013. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars show SEM.

E. Frequency of cells containing H2AK119ub1 Xi accumulation in PEFs that are transduced with the negative control or C181. Statistical analysis of the proportion of cells in the present category to measure differences in dispersal capability was conducted using a t-test, p=0.031. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars show SEM.

F. Frequency of cells containing H3K27me3 Xi accumulation in PEFs that are transduced with the negative control or C181. N=biological replicates and within this the total number of technical replicates, n=nuclei scored. Error bars show SEM.

4.4.5. C181 delays CIZ1 Xi accumulation post mitotic exit

Since we know that CIZ1 is removed from the Xi at mitosis (Figure 4.1, Figure 4.3) and that C181 only exerts DN effects in cycling cells (Figure 4.5), a new question emerged. Can C181 disperse assemblies that are already formed, or does it simply interfere with new assembly events? We postulated that C181 delays the process of CIZ1 Xi accumulation in G1, but could be unable to exert its effects at other stages of the cell cycle, thus explaining why a subset of the population in Chapter 3 still contain observable CIZ1-Xi assemblies. To address this, I exploited the ability to enrich cell populations prior to assembly (early G1), using the cell synchrony experiment illustrated in Figure 4.3 in C181 transduced cells.

This confirmed the result observed in untransduced cells, as the majority of cells upon exit of M phase did not contain a CIZ1-Xi assembly in either the empty vector or C181 transduced population (Figure 4.7A,B). However, during the five hour time course restoration of CIZ1-Xi occurred in the control transduction population, but was supressed in the C181 transduced population (Figure 4.7A,B). In contrast, comparison of CIZ1-Xi assembly status upon release from S phase revealed that CIZ1 dispersal did not occur in the empty or C181 transduced cells (Figure 4.7C,D). This suggests that C181 cannot exert DN functions on pre-formed CIZ1-Xi assemblies, but upon disassembly in mitosis C181 can corrupt reformation of new assemblies. All data from Figure 4.7A-D is presented together in Figure 4.7E to aid simultaneous comparison of all the different data sets, and example pictures are provided in Figure 4.7F. This observation of C181 DN effects specifically post-mitosis was corroborated using FIJI maximum and minimum intensity measurement analysis, where C181 transduced cells had a reduced maximum and minimum fluorescence intensity compared to empty cells at 4 hours post release from mitosis (Figure 4.7G,H). Example pictures used in analysis are presented in Figure 4.7I.

A Empty

B C181

Exit from M phase

Cells with CIZ1 assemblies (%)

Time post M phase (hours)

n=150 *n=121* *n=162* *n=113* *n=152*

n=133 *n=137* *n=165* *n=203* *n=215*

C Empty

D C181

Exit from S phase

Time post S phase (hours)

n=68 *n=75* *n=79* *n=40* *n=110*

n=82 *n=96* *n=104* *n=116* *n=110*

- ■ Present
- ■ Dispersed or diffuse
- □ Absent

E

Cells with CIZ1 assemblies (%)

Time post S/M phase (hours)

— Empty post M   — C181 post M
···· Empty post S   ···· C181 post S

F

Exit from M phase (hours)

E

181

1   2   3   4   5

Exit from S phase (hours)

E

181

1   2   3   4   5

G 4 hours post mitosis

Maximum fluorescence intensity

Empty *n=46*   C181 *n=49*

H 4 hours post mitosis

Minimum fluorescence intensity

Empty *n=46*   C181 *n=49*

I

Empty   C181

Merge   Merge

DNA   CIZ1-RD   DNA   CIZ1-RD

108

Figure 4.7. Effects of C181 transduction in different stages of the cell cycle

A. Time course measuring frequency of endogenous CIZ1-Xi aggregates in nuclei that are transduced with the negative control post release from mitosis. n=nuclei scored.

B. Time course measuring frequency of endogenous CIZ1-Xi aggregates in nuclei that are transduced with C181 post release from mitosis. n=nuclei scored.

C. Time course measuring frequency of endogenous CIZ1-Xi aggregates in nuclei that are transduced with the negative control post release from S phase. n=nuclei scored.

D. Time course measuring frequency of endogenous CIZ1-Xi aggregates in nuclei that are transduced with C181 post release from S phase. n=nuclei scored.

E. Data from A-D combined for comparison. All data was collected from D3T3 cells.

F. Example pictures of empty (E) and C181 transduced cells analysed in A-E. Scale bar is 10µm for each set.

G. Maximum fluorescence intensity measures of D3T3 cells 4 hours post release from nocodazole in the negative control and C181 transduced population.

H. Minimum fluorescence intensity measures of D3T3 cells 4 hours post release from nocodazole in the negative control and C181 transduced population.

I. Example picture of cells used in analysis in F and G. Scale bar is 10µm.

4.5. Discussion

4.5.1. AURKB mediated sequential loss of Xi binding partners?

Results presented here suggest that CIZ1 removal from chromatin is regulated by AURKB, since upon AURKB inhibition CIZ1 was retained later in mitosis. However formal testing is required to identify if AURKB is phosphorylating CIZ1 directly, or if this is an indirect effect. This could be measured by mutation of the suspected AURKB target site to phosphomutant or phosphomimetic amino acids, which would be predicted to cause sustained CIZ1 retention or removal from chromatin respectively. This work is currently being conducted by a PhD student in the Coverley lab. It should be noted that all experimentation in this chapter has been conducted in murine cells, whereas the two papers that discuss SAF-A and *Xist* removal during mitosis used human cells (Hall et al., 2009, Sharp et al., 2020). This causes complications when seeking to compare results, since it is known that there is variance in *Xist* removal between murine and human cells, with human cells displaying *Xist* removal early in mitosis (Clemson et al., 1996), whereas *Xist* can be detected in murine cells in metaphase (Lee and Jaenisch, 1997).

My experiments directly compared CIZ1 and SAF-A in the same cells during mitosis to monitor their behaviour in murine cells. I noted that SAF-A behaviour mimicked that observed by Sharp et al., and was removed during prophase. Since CIZ1 and *Xist* are suggested to form a stable core within a SMAC (Markaki et al., 2021), the later loss of CIZ1 during metaphase is in line with previous literature of CIZ1 removal (Ridings-Figueroa et al., 2017) and *Xist* removal (Lee and Jaenisch, 1997) at this stage. Furthermore, it has been noted that loss of Xi ubiquitination occurs later in mitosis than loss of *Xist* (Hall et al., 2009), suggesting that de-ubiquitination is not an immediate consequence. Notwithstanding that H2AK119ub1 loss in mitosis in murine cells needs to be formally measured to identify if there are also murine/human differences here. This could suggest that in murine cells AURKB mediates early SAF-A removal at prophase, but *Xist*/CIZ1 removal is later. If indeed AURKB is confirmed to be directly regulating CIZ1, perhaps the CIZ1/*Xist* interaction is maintained due to the lack of predicted AURKB phosphorylation sites in CIZ1 PLDs (Figure 4.4A). This could allow CIZ1 to transiently provide a protective role for histone ubiquitination, which is then subsequently removed upon CIZ1 loss (Figure 4.8).
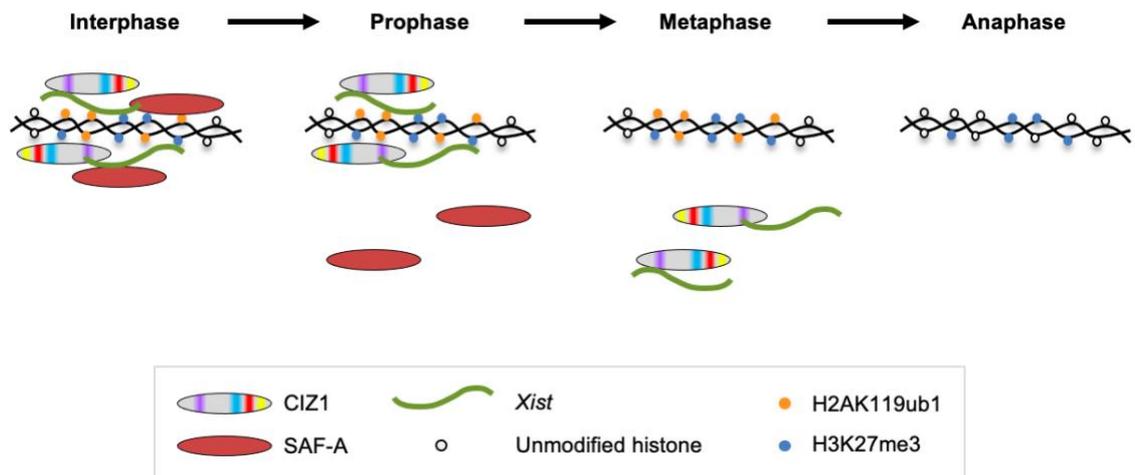
Figure 4.8. Suggested model of sequential loss of proteins and epigenetic modifications from the Xi in mitosis in murine cells

4.5.2. AURKB and PP1/2A are implicated in CIZ1-Xi dynamics

Early work showed that inhibition of protein phosphatase 1 and 2A (PP1 and PP2A) by okadaic acid led to an increase in phosphorylation of H3S10 in interphase cells (Ajiro et al., 1996), and that inhibition of PP1/PP2A leads to AURKB activation (Sugiyama et al., 2002). This suggests an interplay between the antagonistic functions of the phosphatase and kinase to regulate events in mitosis. Indeed, later it was identified that AURKB and PP1/PP2A work together to control the timing of kinetochore assembly and disassembly (Emanuele et al., 2008). The Repo-Man/PP1 complex is implicated in de-phosphorylation of H3 and reduced AURKB targeting (Qian et al., 2011), but additionally AURKB can phosphorylate Repo-Man leading to reduced histone recruitment (Qian et al., 2013). Overall, the exact mechanisms of how they counteract each other remains complex and only partially understood. Early work conducted in our lab identified that okadaic led to loss of CIZ1-Xi assemblies in interphase (Coverley, unpublished), which parallels the result observed by Hall et al, who observed *Xist* dispersal upon phosphatase inhibition. Additionally, CIZ1 has been identified as interaction partner of tartrate-resistant acid phosphatase (TRAP/ACP5) (Reithmeier et al., 2017), which could also be implicated in de-phosphorylation events that oppose kinase activity, but this requires further investigation.

4.5.3. AURKB overexpression in tumours could be implicated in CIZ1-Xi loss

AURKB is highly dysregulated in cancer, and is reported to be overexpressed in breast cancer (Zhang et al., 2020), leukaemia (Goldenson and Crispino, 2015), lung cancer (Takeshita et al., 2013), hepatocellular carcinoma (Xiao and Zhang, 2021) and gastric cancers (Wang et al., 2020). Due to the dysregulation of AURKB in cancer the AURKB inhibitor barasertib (also called AZD1152) has shown promise as an anti-cancer therapeutic in multiple myeloma (Evans et al., 2008), hepatocellular carcinoma (Aihara et al., 2010), pancreatic and colon cancer (Azzariti et al., 2011), lymphomas (Mori et al., 2011), acute promyelocytic leukaemia (Ghanizadeh-Vesali et al., 2016), and non-small cell lung cancer (Helfrich et al., 2016). One study observed barasertib application combined with ionising radiation lead to multinucleation eventually leading to cell death, indicative of inaccurate mitosis (Tao et al., 2009), likely due to loss of AURKB checkpoint activation in mitosis (Hauf et al., 2003). Since the events of CIZ1 accumulation and removal from chromatin are highly regulated by kinase and phosphatase events, the overexpression of AURKB could lead to extended loss of CIZ1-Xi assemblies. This combined with the expression of CIZ1 DN fragments such as C181, which also drive loss of CIZ1-Xi assemblies, could lead to an additive effect contributing to loss of CIZ1 epigenetic maintenance, and genome wide dysregulation in cancer.

4.5.4. C181 prevents the formation of SMACs?

As postulated in section 4.4.1, if maximum fluorescence intensity values correspond to the endogenous CIZ1 SMACs before and after Xi targeting, the fall in maximum fluorescence intensity in the presence of C181 post-mitosis (Figure 4.7F) suggests that it limits this SMAC formation. Interestingly, if the minimum fluorescence intensity values do correspond to soluble CIZ1 levels, this suggests that C181 expression also leads to a reduction in soluble CIZ1 levels (Figure 4.7G). This is surprising, since if soluble CIZ1 is reduced as CIZ1 SMACs are formed, you would expect these levels to rise in the presence of C181. Considering all this, perhaps in fact this is how C181 perturbs SMAC formation, in that it limits the build-up of soluble CIZ1 that would go on to form the SMACs (Figure 4.9). This could be as a consequence of active turnover, perhaps if the cell detects aberrant full length CIZ1/C181 complexes it attempts to degrade them. Alternatively, maybe if the cell mistakes C181 levels for endogenous CIZ1 expression it downregulates CIZ1 transcription post-mitosis. Regardless of the exact mechanism, the outcome is reduced CIZ1 accumulation at the Xi and thus a loss of the heterochromatin marks as illustrated in Figure 3.8 in Chapter 3.
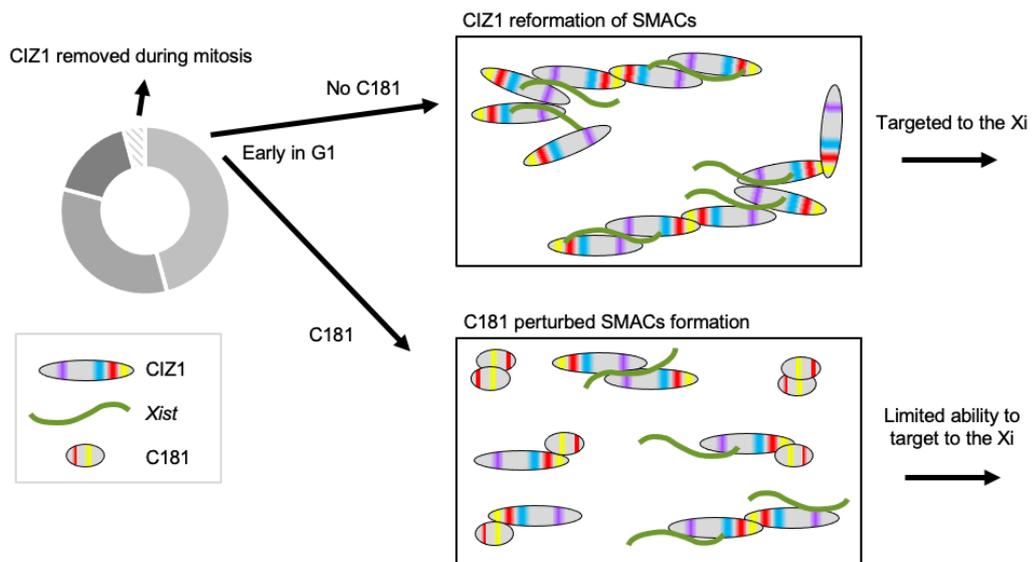
Figure 4.9. Proposed model of C181 perturbing CIZ1 SMAC formation

4.5.5. H2AK119ub1 and H3K27me3 differences in mitosis, and in response to C181 dispersal

Results presented here indicate that there are differences in H2AK119ub1 and H3K27me3 maintenance in mitosis, and in their sensitivity to perturbation when exposed to CIZ1 DN fragments. The retention of H3 methylation through mitosis is well documented (Boggs et al., 2002, Peters et al., 2002, Hall et al., 2009, Ridings-Figueroa et al., 2017), and is possibly due to the retention of PRC2 bound to chromatin in mitosis that provides a maintenance effect (Aoto et al., 2008). The reinstatement of diluted H3K27me3 levels post replication is also an important mechanism for epigenetic memory, and this has been identified as being facilitated by CDK1 and CDK2 phosphorylation events that increase PRC2 efficiency in S and G2 phase (Zeng et al., 2011). It has also been observed that upon H2K119ub1 loss there is a reduction of PRC2 bound to chromatin, and whilst this does not lead to a rapid reduction in H3K27me3 levels, they are gradually lost, consistent with replication dilution (Dobrinic et al., 2021). This suggests there is an element of reliance of H3K27me3 on H2AK119ub1, and indeed the loss of H3K27me3 at the Xi upon longer exposure to C181 DN fragments that perturb H2AK119ub1 agrees with this theory. This result is in line with our recent publication that loss of endogenous CIZ1 leads to genome wide reduction of both H2AK119Ub1 and H3K27me3 in CIZ1 null cells (Stewart et al., 2019).

4.5.6. Concluding remarks

Whilst CIZ1 C-terminal interactions are important for anchorage to the NM (Ainscough et al., 2007) and for CIZ1 self-interactions (Chapter 3), the C181 fragment alone does not measurably corrupt the existing preformed endogenous CIZ1-Xi complex. This could be due to limited accessibility, or it could be that it does not have a high enough affinity to endogenous CIZ1 or other component of the SMAC. However, when dispersed in mitosis, endogenous CIZ1 is more susceptible to interference, possibly due to accessibility of MH3 associated interactions (described in Chapter 3) that facilitate CIZ1 self-interaction. Further analysis to investigate the downstream consequences of extended CIZ1-Xi assembly and repressive heterochromatin mark loss at the Xi, will be crucial to understand how CIZ1 DN fragments might influence carcinogenesis.

5. Analysis of CIZ1B in lung cancer

5.1. Introduction

5.1.1. CIZ1 mis-splicing in cancer

As discussed in the Introduction *CIZ1* is extensively alternatively spliced, and also mis-spliced in disease contexts. Cancer associated splice variants (Rahman et al., 2010), include inappropriate exclusion of exon 4 (Rahman et al., 2007), splicing of exon 8-12 to produce *CIZ1F* (Swarts et al., 2018), and differential splicing at the exon 14/15 junction in *CIZ1B* (Higgins et al., 2012). Additionally, as shown in Chapter 3, the C-terminus of CIZ1 is overexpressed in multiple cancer types as early as stage I. The mechanisms behind the production of these alternative transcripts, and whether they are generated by shared or independent processes remains unknown. This chapter focuses on further analysis of the *CIZ1B* splice variant. Alternative splicing at the 14/15 junction in *CIZ1B* results in a 24 base pair nucleotide loss, leading to an in-frame deletion of 8 amino acids in the acidic domain (AcD) of CIZ1.

5.1.2. CIZ1 AcD and transactivation domains

Transcription factors contain distinct elements, including but not limited to, a DNA binding domain (DBD) and a transactivation domain (TAD). TADs provide a binding site for other proteins required for transcriptional regulation. Early literature describes TADs as unstructured regions with an excess of negatively charged amino acids referred to as "acid blobs" or "negative noodles" (Sigler, 1988). Later the definition of TADs was modified to include enrichment of other amino acids, such as poly glutamine (Courey et al., 1989), proline (Gerber et al., 1994) and isoleucine (Attardi and Tjian, 1993). Later work showed that there was not a set of sequence specific residues required for TADs, but rather hydrophobic residues amongst disordered regions (Warfield et al., 2014). Mutations of the TAD in Gcn4 transcription factor in yeast showed that addition of aromatic residues increased activity, whereas adding isoleucine and valine, and methionine and leucine had none or modest activating effects respectively (Staller et al., 2018). However, since Gcn4 is in the 93rd percentile for the number of hydrophobic residues it contains in its TAD, this could explain why only large bulky hydrophobic aromatics show a measurable difference in this context, and the importance of both aromatic and non-aromatic hydrophobic residues has been shown by multiple research groups (Drysdale et al., 1995, Almlof et al., 1997, Jeffery and Weinzierl, 2020, Staller et al., 2022). All this together suggests that the AcD in CIZ1, in particular the eight amino acids lost in CIZ1B (VEEELCKQ), could be defined as a TAD and behave as a scaffold for protein:protein interactions underpinning key processes. Since it is already suggested that CIZ1 binds a loose DNA consensus (Warder and Keherly, 2003), it is possible that

CIZ1 could be involved in chaperoning proteins to desired regions of the genome in response to signals. This could align with previous suggestions that it is a kinase sensor for initiation of DNA replication (Pauzaite et al., 2016), and that it facilitates movement of the Xi in S phase (Stewart et al., 2019).

5.1.3. Lung cancer has high mortality and limited diagnosis options

Lung cancer is the third most common cancer in the UK, accounting for 13% of all cancer diagnoses between 2016 and 2018 (CRUK, 2021). Since diagnosis of lung cancer is often late, with 49% of cases in 2019 in England not being diagnosed until stage IV (NHS, 2021), mortality is high, and approximately only 4% of patients diagnosed at stage IV have 5-year survival (NHS, 2022). This makes it the most common cancer death in the UK, accounting for 21% of all cancer deaths between 2017 and 2019 (CRUK, 2022).

Traditionally lung cancer mortality has been almost exclusively associated with smoking, however more recent publications are now suggesting that there is a significant proportion of cancer diagnoses that are not linked to smoking, and one group have calculated that 35% of driver gene mutations in lung cancer are mutations due to replication errors not hereditary or environmental factors (Tomasetti et al., 2017). Additionally with declining rates of smoking, the relative proportion of lung cancers in never-smokers is increasing, and if considered as a separate entity lung cancer in never-smokers is the eighth most common cause of cancer related death in the UK (Bhopal et al., 2019) and approximately 25% of lung cancer cases worldwide are not attributed to tobacco use (Sun et al., 2007). Analysis in a large cohort of over a million female never-smokers in the UK only revealed three risk factors associated with increased lung cancer development, this included non-white ethnicity, a taller stature and asthma that requires treatment (Pirie et al., 2016), and there was little association with sociodemographic factors and dietary intake. Thus, from this study no lifestyle changes can be recommended to patients to aid in the prevention of development of lung cancer.

Considering this, there is an increasing demand for development of biomarkers that permit early-stage detection, and to date there is no uniform blood biomarker in clinical practise that is capable of detecting it at an early stage in asymptomatic individuals (Zamay et al., 2017). The use of low-dose computed tomographic (CT) screening could be utilised in a national screening programme, as it has been shown to reduce mortality. However of the 24.2% of positive results monitored 96.4% were false positives, thus making it an inefficient test for diagnosis (National Lung Screening Trial Research et al., 2011), and emphasising the need for biomarkers that can rule our patients with false positive (benign) CT results.

5.1.4. CIZ1B as a lung cancer biomarker

The CIZ1B protein is overexpressed in lung cancer patients, and by testing for its presence in circulating plasma it is possible to discriminate between patients with stage I disease compared to non-cancerous lung abnormalities such as benign lung nodules, asthma and chronic obstructive pulmonary disease (COPD) (Higgins et al., 2012). Considering this, an assay has been developed to allow quantification of CIZ1B fragments in patient plasma in a clinical setting (Coverley et al., 2017).

Additionally, we know that CIZ1B is a driver of cancer development, rather than a downstream passenger consequence, since upon CIZ1B loss cancer cell growth was restrained (Higgins et al., 2012). So, whilst CIZ1B could be an indispensable biomarker for diagnosis in the future, its potential for use as a therapeutic target remains underexplored. Here, I explore the properties of CIZ1B itself, and its potential to influence downstream pathways that contribute to the formation of cancer.

5.2. Aims

- Determine if CIZ1 and CIZ1B exist in the same native state *in vitro*
- Identify if there are any binding partner differences between CIZ1 and CIZ1B
- Monitor CIZ1 and CIZ1B behaviour via transient transfection in murine cells
- Measure *CIZ1B* transcript levels in lung and breast cancer compared to their respective normal tissue

5.3. Experimental Design

Since very little was known about CIZ1B function prior to the start of my PhD, several different experimental techniques have been utilised here to facilitate generation of a large breadth of different data. This chapter includes data from both human and murine CIZ1; human because of the obvious clinical relevance, and mouse because results can be aligned with this lab's body of functional data. They have been presented together here to facilitate analysis of the AcD and the B variant region of CIZ1. The Coverley lab considers a murine model to be similar enough to human model that they can be compared directly, since domain organisation is fully conserved (Figure 5.1), with up to 96% amino-acid sequence conservation in conserved domains, and 65% overall. All the techniques listed below are described in depth in the Materials and Methods chapter.

Domain	PLD1	PLD2	ZF1	ZF2	ZF3	AcD	MH3	**Overall**
Sequence identity (%)	78	71	92	86	96	67	79	**65**
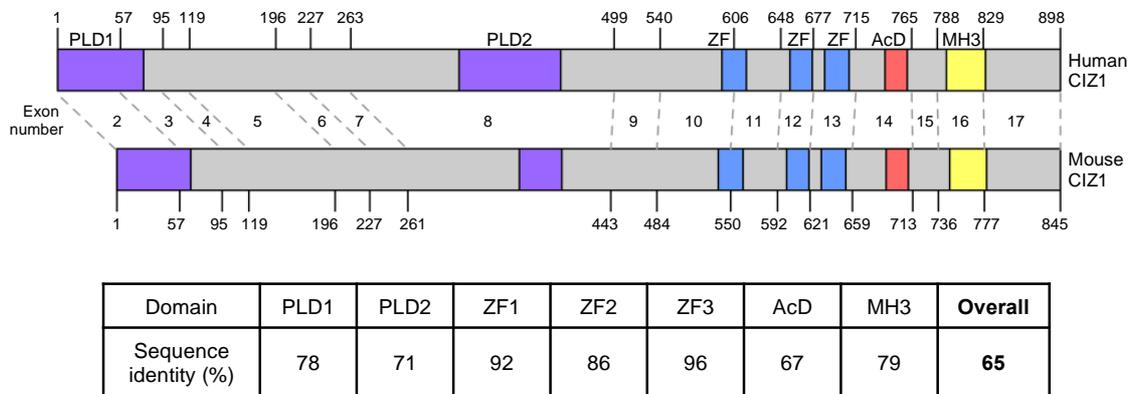
Figure 5.1. Alignment of murine and human CIZ1
Top. Alignment of the murine (UniProt ID Q8VEH2) and human (UniProt ID Q9ULV3) CIZ1 protein. Numbers corresponded to the approximate amino acid numbers at exon boundaries (rounded up or down accordingly when the boundary falls within a triplicate nucleotide set corresponding to an amino acid residue). The domains highlighted are: prion like domain 1 and 2 (PLD1 and PLD2) as documented in (Sofi et al., 2022); three zinc fingers (ZF1, ZF2 and ZF3) (ZnF_C2H2 SM00355, ZF_C2H2 sd00020 and ZF_C2H2 sd00020); an acidic domain (AcD) containing a highly concentrated area of aspartates and glutamates; and a matrin-3 like domain (MH3) (ZnF_U1 smart00451).
Bottom. Comparison of amino acid sequence identity between human and murine CIZ1, documenting that higher conservation is observed in domains compared to the overall conservation.

5.3.1. Biochemical analysis

CIZ1 C-terminal constructs were expressed in bacteria and purified for biochemical analysis. The human and murine forms of CIZ1 used for biochemical analysis in this chapter are the same molecular weight and start from the same point in CIZ1. They are referred to as H720 and C181 respectively, which refers to the starting amino acid in human (amino acids 720-898 and thus the terminal 179 amino acids), or the total number of amino acids in the mouse construct (amino acids 665-845 and thus the terminal 181 amino acids). This different nomenclature was conceived prior to my PhD to allow for easy discrimination of murine and human CIZ1 forms. Purified forms of recombinant CIZ1 underwent size exclusion chromatography (SEC), size exclusion chromatography multiple laser light scattering (SEC-MALLS) and mass spectrometry (MS) analysis.

5.3.2. Transient transfection in murine cells

Existing full-length forms of human CIZ1 and CIZ1B with an N-terminal GFP tag were transiently transfected for 24 hours into female murine fibroblast cell line (D3T3 cells) to monitor differences in behaviour at the Xi. Cells underwent immunofluorescence processing for co-staining with repressive marks found enriched at the Xi (H2AK119Ub1 and H3K27me3) to confirm Xi status and location.

5.3.3. Bioinformatics

With the help of bioinformatician Dr Andrew Mason, RNA-seq data from 433 lung cancer patients available from TCGA was accessed and analysed to measure expression levels of *CIZ1* exons, and normalised relative to each individual canonical CIZ1 exon 7 as described in Chapter 3. Quantification of *CIZ1B* transcript was also carried out using the same 433 lung cancer patients, as well as 1078 breast cancer patients. *CIZ1* exon expression in breast patients presented in Chapter 3 was edited to align with the breast cancer patient set analysed above. Comparison of *CIZ1B* transcript levels in normal breast and lung tissue was also conducted using RNA-seq data from the Genotype-Tissue Expression (GTEx) project portal (GTEx, 2013). In all cases the quantity of *CIZ1B* variant transcript was calculated as a percentage of the total reads in the 14/15 junction to calculate variant allele frequency (VAF).

5.4. Results

5.4.1. Murine and human C-terminal CIZ1 both exist in a multimeric state

The domains contained in human H720 and murine C181 C-terminal fragments of CIZ1 have been extensively described in the Introduction, but in summary, include the negatively charged acidic domain (AcD) and the matrin 3-homology domain (MH3) which contains a matrin like zinc finger (Figure 5.2A). They have a theoretical molecular weight of 20kDa, but when analysed in SDS-PAGE migrate aberrantly and are observed at approximately 25-30kDa (Figure 5.2B). This is hypothesised to be due to the high negative charge affecting SDS coating of the polypeptide, and this phenomenon is routinely observed by all members of the Coverley lab. The existence of two forms in both cases is hypothesised to be caused by the cleavage of the terminal 37 amino acids occurring during bacterial expression, as confirmed by MS analysis indicating the existence of these fragments and loss of epitope of extreme C-terminus CIZ1 antibodies (data not shown).

When analysed under native conditions, neither construct behaves as a single species, and instead migrates as multiple isoforms (Figure 5.2C). Electrophoresing these fragments under native conditions allows for qualitative analysis of CIZ1 forms, but as can be seen by the differences in two different protein molecular weight ladders, it does not facilitate accurate molecular weight estimates. Thus, alternative assays for definitive molecular weight estimates were exploited.

In Chapter 3 I showed that the MH3 domain is responsible for dimerisation of C-terminal fragments. Here I aimed to identify if the AcD also contributes to this, or whether it might influence conformation. The H720B construct was already in existence prior to my PhD so this was compared to H720. To complement this, I created a form of CIZ1 that lacked all of the AcD in the murine context (C181 ΔAcD) and compared this to C181. Together these two comparisons allow observation of potential effects of the AcD, verses a cancer-associated human splice variant that affects the AcD region. A schematic representation of these proteins can be seen in Figure 5.2D.

Size exclusion chromatography (SEC) profiles shown in Figure 5.2D are cropped to focus on the CIZ1 peak of interest, and full uncropped traces are presented in Appendix A. The SEC analysis of the human forms of CIZ1 was conducted in collaboration with Lewis Byrom, a fellow PhD student in the Coverley lab. The SEC analysis of CIZ1 alongside their respective mutants identified that the elution time of human and murine CIZ1 was different, with murine CIZ1 eluting earlier than human CIZ1. More significantly it also identified that loss of the AcD caused a shift in elution time, with H720B showing the more conservative change consistent with deletion of less of the AcD. (Figure 5.2D).

Figure 5.2. Analysis of CIZ1 C-terminal fragments

A. Diagram of full-length CIZ1 and small C-terminal fragments used in biochemical analysis in this chapter (C181/H720). C181 and H720 incorporates the AcD domain and the MH3 domain only, and are equivalent fragments of CIZ1 from murine and human respectively. C181 incorporates the terminal 181 amino acids, (including amino acids 665-845), and H720 incorporates the terminal 179 amino acids (amino acids 720-898). The amino acid sequence of C181 and H720 is shown, and the AcD and MH3 domain is highlighted in colours corresponding to the diagram above. The B variant region is underlined in green.

B. Denatured SDS-PAGE analysis of C181 and H720. Coomassie staining allows visualisation of all protein forms expressed in the bacterial system after purification steps. The same denatured molecular weight ladder is used in both gels and in part C.

C. Native PAGE analysis of C181 and H720. The ladder on the left is a native ladder, the ladder on the right is a denatured ladder.

D. Size exclusion chromatography (SEC) analysis of human and murine CIZ1. All samples had the sample volume loaded, the column type and flow rate matched to allow direct comparison. Protein presence is detected by absorption at wavelength 280nm

121

5.4.2. CIZ1 exists as a dimer regardless of acidic domain presence, however confirmational changes are observed

Size exclusion chromatography multi angle laser light scattering (SEC-MALLs) was conducted on the same forms of CIZ1 (Figure 5.3A). This identified that all forms are estimated to be 37-47kDa in size (Figure 5.3B). When compared to the monomeric sizes of 17-20kDa this indicates that all forms of CIZ1, regardless of AcD presence, exist as a dimer (Figure 5.3B). Thus, we can also conclude from this that the loss of the AcD leads to conformational changes in CIZ1, since the shift in retention time was due to changes in shape rather than changes in multimeric weight. The full uncropped SEC-MALLs traces are provided in Appendix B.

Modelling these forms of CIZ1 using AlphaFold2 software (Mirdita et al., 2022) showed that there are predicted secondary structure changes upon loss of the AcD domain (Figure 5.3C). They have been modelled as a monomer here to aid visualisation of conformational changes rather than multimeric changes. The more conservative H720B overlayed with H720 showed more minor changes, with the N terminus of the proteins overlaying regardless of the eight amino acid loss, however the C-terminus appears to have an altered structure, with potential changes in internal salt bridge formation, which could leave H720B with greater flexibility. The complete removal of the AcD in murine CIZ1 shows a more dramatic change. The N and C-terminus of C181 are predicted to sit close together in three-dimensional space, with the potential for two salt bridges to form similarly to the human forms. In contrast, the N and C-terminus of C181 ΔAcD are positioned much further apart. This more extreme change in secondary and tertiary structure could explain the shift in elution time of C181 ΔAcD compared to C181. It should be noted that all the salt bridges were predicted under loose conditions and range from 10-12Å in length, thus whilst they could exist, they can be treated as circumstantial evidence and would need to be tested directly. Nonetheless, regardless of the existence of these salt bridges, bioinformatic predictions in Figure 5.3C agree with the experimentation conducted in Figure 5.3A
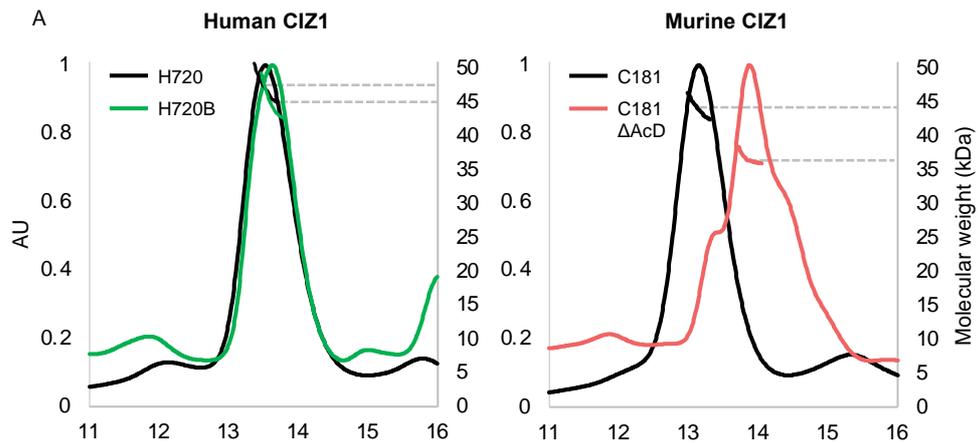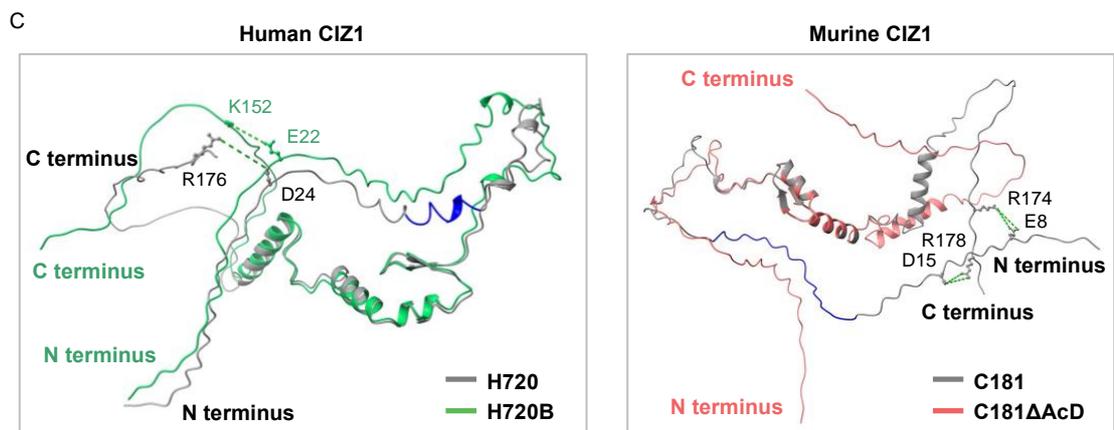
Figure 5.3. Further analysis of CIZ1 C-terminal fragments

A. Size exclusion chromatography (SEC-MALLS) analysis of human and murine CIZ1. The peaks correspond to 280 nm measurements that are read off the left hand side y-axis, the lines correspond to light scattering measurements that are read off the right hand side y-axis.

B. Molecular weight estimates produced from internal light scattering measurements taken in A. Multimeric state can be inferred from accurate molecular weight estimates.

C. AlphaFold2 (Mirdita et al., 2022) modelling of murine and human C-terminal fragments. The polypeptide colours correspond to colours used in part A to identify WT and mutant forms. Polypeptides were overlayed in ChimeraX (Pettersen et al., 2021) for visualisation of structure changes. Blue regions on the WT forms of fragments show the positions of the region removed in the respective mutants (VEEELCKQ and AcD in human and mouse respectively). Single letter amino acid code highlights residues involved in intra salt bridge connections.

5.4.3. CIZ1 and CIZ1B interaction partner study

To test whether the suspected confirmational changes occurring in CIZ1 upon AcD domain deletion influences the repertoire of protein-protein interactions supported by the C-terminus of CIZ1, I participated in a collaborative experiment whose main aim was to generate the first list of C-terminal interaction partners. Previous work from the Coverley lab has used GST-labelled N-terminal fragments of CIZ1 in pull-down assays using lysates derived from HeLa cells (Thacker et al., 2020), but no information on C-terminal interaction partners was available. We (myself, Lewis Byrom, Dawn Coverley, and two undergraduate project students) compared H720, H720B and GST alone using the workflow shown in Figure 5.4, and described in detail in the Materials and Methods chapter.
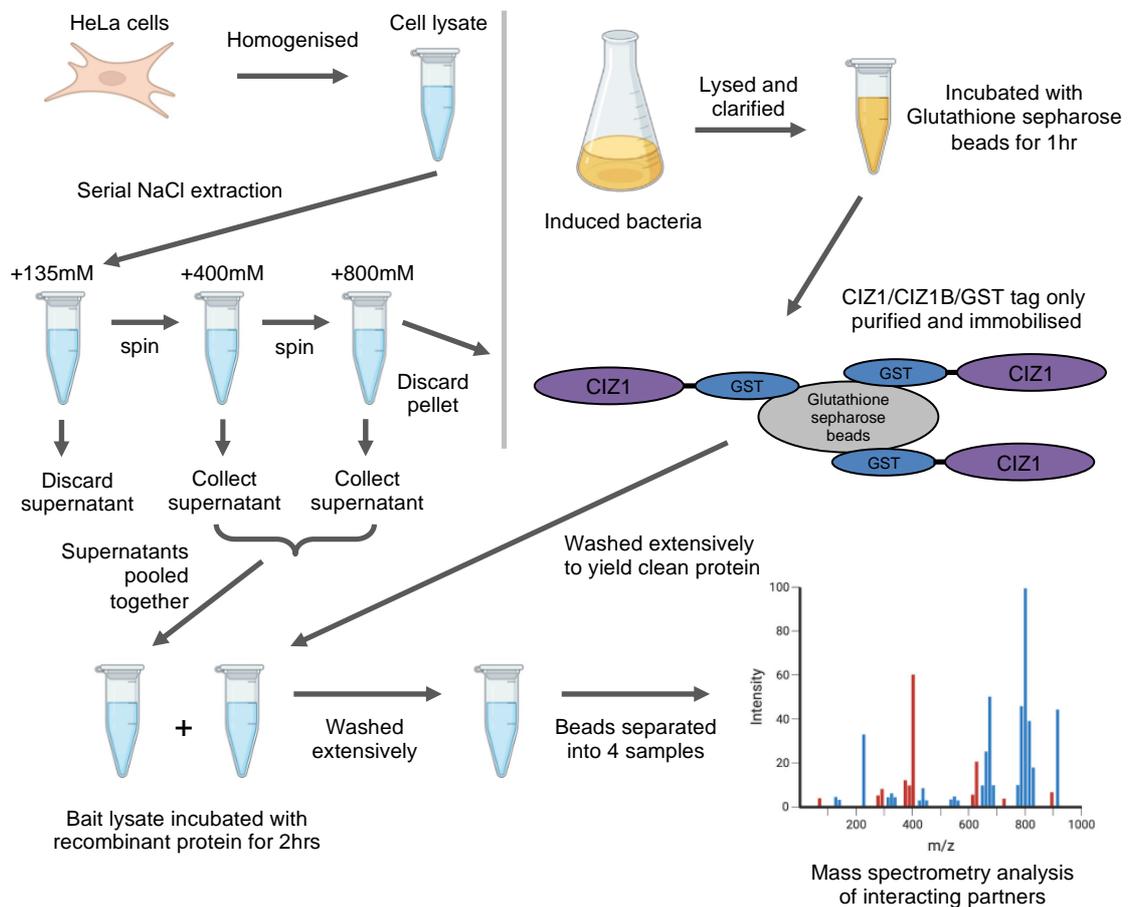
Figure 5.4. Workflow of sample preparation for mass spectrometry analysis
The left hand panel depicts the serial salt extraction steps to yield a pool of bait proteins that were loosely bound within the cell or tightly bound such as histones to chromatin. The right hand panel depicts expression and purification of recombinant protein used in the interaction study. The bottom shows the incubation of the sample and the final steps of the experimentation.

5.4.4. Most binding partners identified interact with a higher affinity to H720 compared to H720B

139 binding partners were identified in this study, after appropriate thresholding and significance analysis to remove spurious results, the full list is provided in Appendix C. The precursor ion area allowed us to quantify how much of each binding partner was present in each technical replicate, and is represented in a heat map (Figure 5.5). This shows that most of the binding partners show low or no binding to the GST only control (pGEX), and that the CIZ1 and CIZ1B groups cluster together more closely than to the control, thus giving us confidence that any interactions we are observing are not caused by non-specific interaction with the GST tag. Further analysis of the 121 binding partners that bound to CIZ1 forms more strongly than the negative control (Appendix C) using PANTHER pathway (Mi and Thomas, 2009) only identified DNA replication with a false discovery rate<0.05. This is unsurprising based on the published information detailing CIZ1 involvement in G1/S transition to promote DNA replication (Coverley et al., 2005), but does not yield any useful information.

The heat map also identified that there was likely some sample loss in the H720B 3 sample compared to the other H720B samples. To overcome this, all CIZ1 samples were normalised to their respective total ion content to control for any potential loss or gain of binding partner quantity due to loss or gain of overall protein recovered, and then re-compared to each-other to identify differences. To facilitate creation of a binding partner list that was simple to interpret, stringent thresholding was applied and only proteins with an adjusted t-test value of q<0.005 were considered. In addition to this, proteins previously identified as binding pGEX with equal or higher affinity were also removed. This reduced the initial set to a total of 44 binding partners, with 37 interacting with a higher affinity to H720 compared to H720B and 7 interacting with a higher affinity to H720B compared H720. The binding partners are plotted graphically for visualisation, where proteins to the left of the y-axis correspond to H720 binding proteins, and proteins to the right of the y-axis H720B binding proteins, with the Log2 fold change values depicting the average quantification differences between the two groups (Figure 5.6A). Proteins of interest that are discussed further have been highlighted. Additionally, all the interaction partners are listed based on their subcellular location in Figure 5.6B. These are shown as their respective gene names rather than protein accession numbers to facilitate interpretation. The full list of all binding partners measured after normalisation is provided in Appendix D.
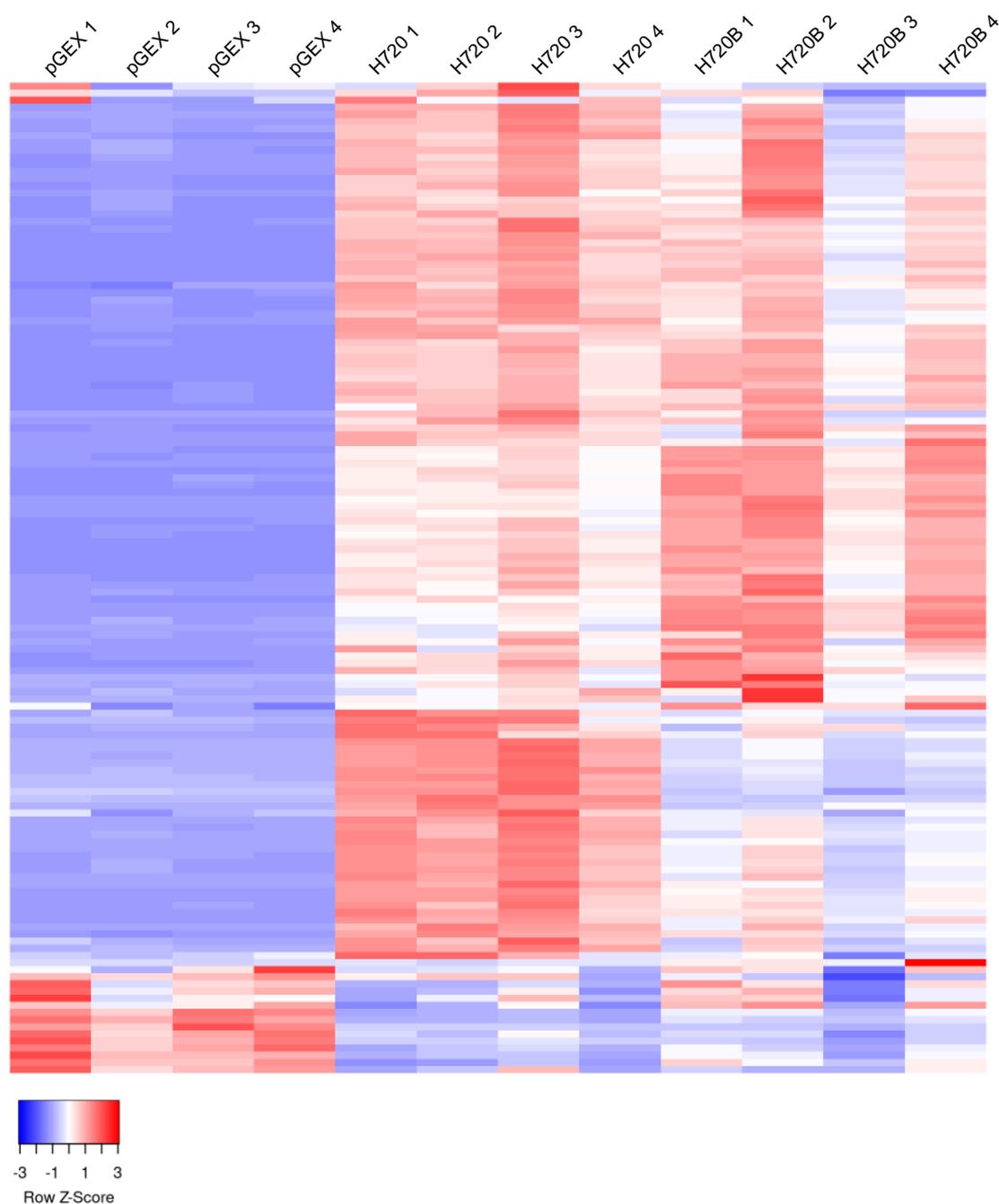
Figure 5.5. Heat map of interaction partner recovery

139 CIZ1 binding partners identified in the mass spectrometry analysis. High precursor ion areas are depicted in red, and lower precursor ion areas are depicted in blue. The clustering method used was average linkage and distance measurement method used was Pearson correlation coefficient. The Heat map was generated using Heatmapper (Babicki et al., 2016).

A. Visualisation of 44 H720 and H720B binding partners at the q<0.005 threshold.
Log2FC values represent affinity changes between the two groups, where negative
values represent greater affinity to H720 and positive values greater affinity to H720B.
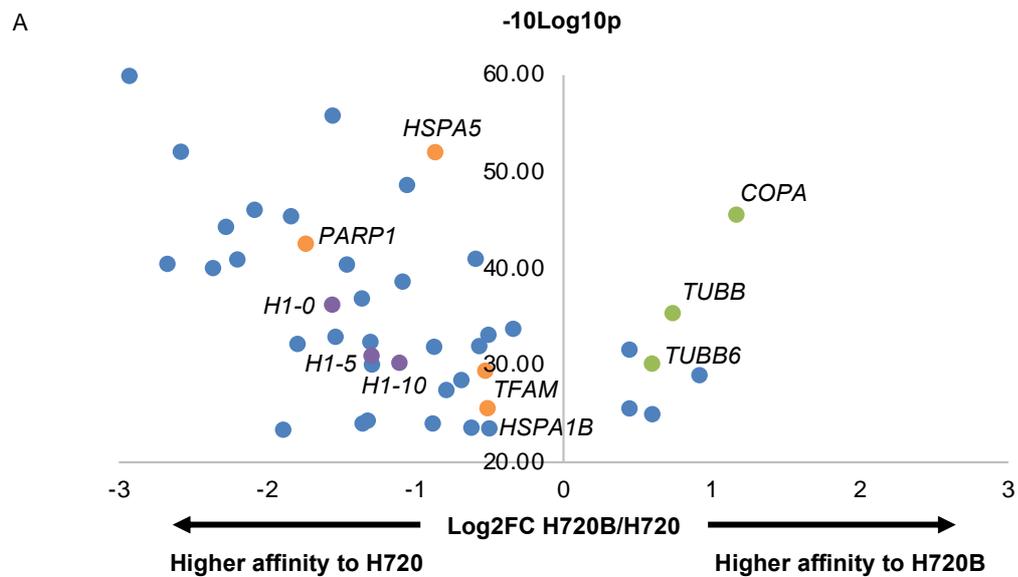
B.

Nucleolar	Nucleus
NOP2	SRSF3
CCDC86	HSPA5
UBTF	HP1BP3
RSL1D1	PARP1
DDX21	GAPDH (nuclear envelope)
RPL31	TOP1 (and nucleolus)
SRP14	H1-0
DHX9	CIZ1
RPL35	LUC7L2
DDX27	RBMX
BRIX1	H1-5
RPL15	H1-10 (and nucleolus)
RPS26	TFAM (and mitochondria)
RBM28	HSPA1B
RPS8	
RPL11	
PDCD11	
RPL3	
RPL18	
NOP58	
RPS7	
SRP72	
RPL19	

Nucleolar	Nucleus
RRP1B	COPA
RPS11	TUBB
DDX24	TUBB6 (and cytosol)
	DDX1 (and cytosol)

Figure 5.6. Final list of H720 and H720B interaction partners
A. Visualisation of 44 H720 and H720B binding partners at the q<0.005 threshold.
Log2FC values represent affinity changes between the two groups, where negative
values represent greater affinity to H720 and positive values greater affinity to H720B.
B. Protein shown above listed with their respective subcellular localisation. This was
identified based on data from the human protein atlas (Thul et al., 2017), UniProt
(UniProt, 2021) and COMPARTMENTS (Binder et al., 2014).

5.4.5. GFP-CIZ1B accumulates at the Xi at a higher frequency than GFP-CIZ1

To explore possible behavioural differences between CIZ1 and CIZ1B, GFP tagged CIZ1 plasmids were introduced into cycling D3T3 cells by transient transfection. Cells that have been transfected with CIZ1 fall into two distinct phenotypes, that either do or do not contain a GFP-CIZ1 marked Xi (Figure 5.7A). In previous experimentation conducted in the Coverley lab up to 20% of cells in a rapidly cycling population do not contain a detectable endogenous CIZ1-Xi assembly, typically validated by detection of H2AK119Ub1, H3K27me3 or Xist (Ridings-Figueroa et al., 2017). The exact number depends on the cell population and its rate of cycling. As shown in the previous chapter, cells with unmarked Xi's are mostly in early G1 and yet to rebuild their CIZ1 assemblies after exiting mitosis. Thus, in a cycling population the majority, but not all, cells will have a CIZ1 marked Xi. When monitoring localisation of ectopic CIZ1 a similar result was seen, however when comparing full length CIZ1 or CIZ1B (Figure 5.7B) I observed some small but reproducible differences. Notably, an increase in number of cells with assemblies at Xi for GFP-CIZ1B compared to GFP-CIZ1. This is represented inversely as a difference in the number of cells which do not have ectopic CIZ1 at Xi (Figure 5.7C). Although not explored further, this result begins to suggest differences in the way that CIZ1B is assembled and disassembled as cell pass through the cell cycle.
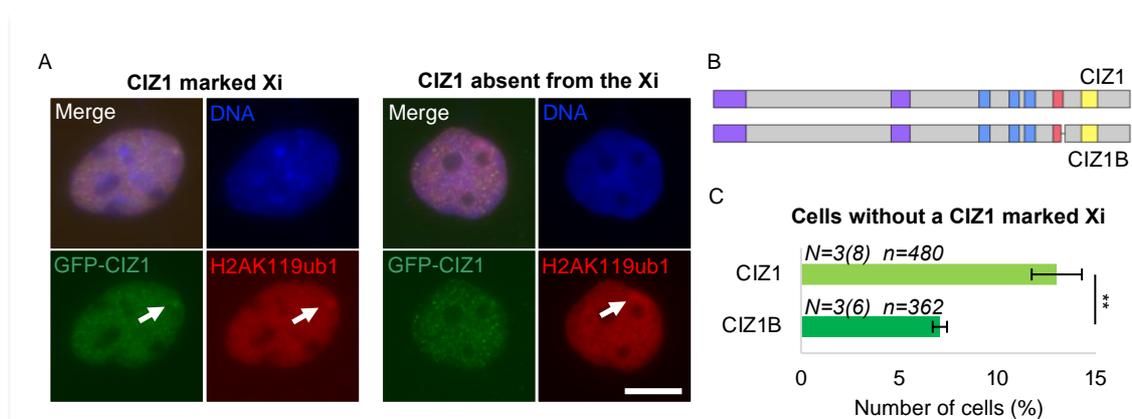
Figure 5.7. Behaviour of ectopic CIZ1 and CIZ1B in murine cells
A. Example immunofluorescence pictures of cells transfected with GFP-tagged CIZ1 or CIZ1B. Two different cell populations are observed, either with or without a CIZ1 at the inactive X chromosome (Xi). Cells are co-stained with a H2AK119ub1, a well characterised repressive mark that is observed at the Xi to corroborate CIZ1 Xi status.
B. Diagram of human CIZ1 forms used in experimentation. In contrast to biochemical experimentation full length forms of CIZ1 were utilised in this experimentation.
C. Visualisation of the percentage of cells that do not have a CIZ1 marked Xi in CIZ1 and CIZ1B transfected cells. Analysis was conducted in D3T3 cells. N=biological replicates and within this the total number of technical replicates, n=number of cells measured. Error bars show SEM. Statistical comparison is by t test, p=0.002.

5.4.6. *CIZ1B* and *CIZ1* transcript profile in breast and lung cancer

To date CIZ1B has only been documented in lung cancer (Higgins et al., 2012), but could also occur at transcript and protein level in other tissue types. In Chapter 3 I presented 3' overexpression of *CIZ1* in TCGA breast samples and breast cancer cell lines. Here, we measured the *CIZ1B* variant allele frequency (VAF), calculated as a percentage of the total reads in the 14/15 junction (i.e. the normal 14/15 junction (referred to as *CIZ1A*) and the *CIZ1B* reads added together). This analysis was conducted in breast normal and cancer samples and in lung normal and cancer samples, to allow internal comparisons.

Analysis of lung samples revealed a very minor increase in *CIZ1B* VAF in cancer tissue compared to normal (8.6% increased to 8.8%) (Figure 5.8A). This was surprising based on previous data that implicated approximately 40 fold change elevation of *CIZ1B* transcript in lung tumours and cancer cell lines (Higgins et al., 2012). Further analysis conducted to measure *CIZ1* exon expression, revealed there was also 3' overexpression of *CIZ1* transcripts in lung cancer patients staring from exon 10 (Figure 5.8A), as seen in breast cancer patients in Chapter 3 (presented again in Figure 5.8B). It should be noted that since we do not have data on *CIZ1* exon expression levels in normal adjacent lung samples we can't be sure that 3' elevation is a cancer specific feature in lung tissue, however based on the reduced 3' overexpression in normal adjacent breast tissue we are assuming this to be true. Since in Higgins et al. *CIZ1B* was quantified relative to actin, and here it was quantified relative to total *CIZ1A* and *CIZ1B*, this explains this discrepancy. Since elevation of the normal 14/15 junction too could mask *CIZ1B* elevation. Measurement of *CIZ1B* VAF in breast samples did not match the result in lung samples, and contrastingly there was a drop in *CIZ1B* VAF in the cancer samples relative to the normal samples, from ~8% to ~4% (Figure 5.8B).

When comparing total CIZ1 expression (transcript per million values) in normal adjacent samples and cancer samples from TCGA using GEPIA (Tang et al., 2017), I found that in breast cancer there is an overall decrease in CIZ1 expression, however in lung cancer there is an overall increase in CIZ1 expression (Figure 5.8C). This raises the possibility that there are different aberrant CIZ1 expression events occurring in the cancer types. For example, whilst there is strong 3' overexpression in breast cancers, the downregulation of the 5' of *CIZ1* (and possibly downregulation of full-length forms) may supersede this. If indeed *CIZ1B* is only present within full length *CIZ1*, this could explain why there is a drop in *CIZ1B* VAF in breast cancer, but in lung cancer there is a small increase. Since in breast cancer there could be 3' overexpression of *CIZ1* transcripts not containing the *CIZ1B* junction, and also a fall in full length transcript containing *CIZ1B*, thus leading to an overall *CIZ1B* VAF reduction. Whilst measuring *CIZ1B* VAF based on the total reads at the 14/15 junction seemed appropriate at the time of analysis, in the future quantification of *CIZ1B* VAF should be based on an alternative control. This could be a housekeeping gene like actin or an upstream exon of CIZ1 such as exon 6/7 that is not subjected to 3' elevation, and this would facilitate quantification of *CIZ1B* independent of *CIZ1* 3' expression levels. However, this still provides interesting preliminary data that *CIZ1* expression profiles in cancer may not be consistent across tissue types.
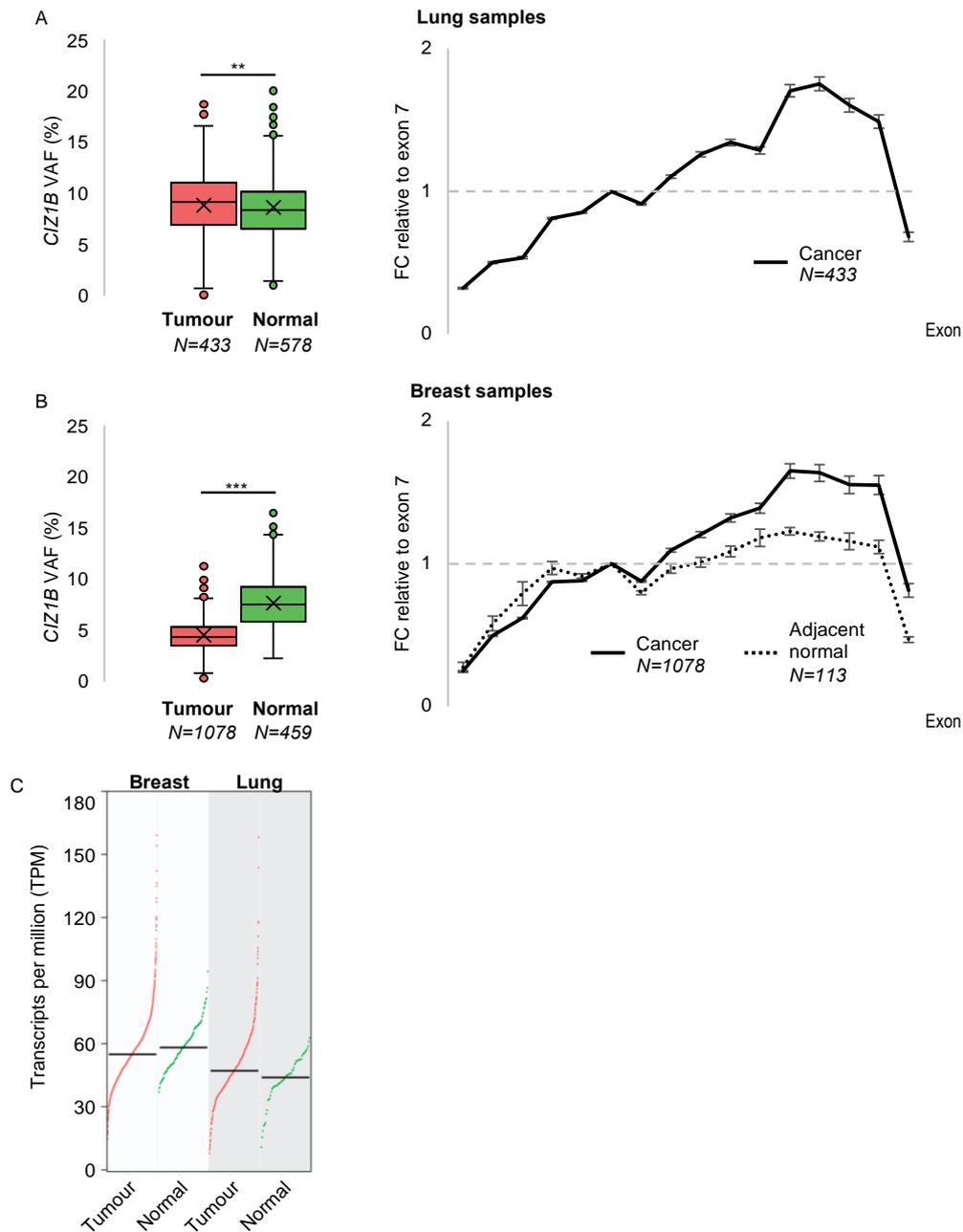
Figure 5.8. *CIZ1B* and *CIZ1* transcript profile in breast and lung cancer

A. Lung samples *CIZ1B* VAF relative to normal junction 14/15 expression (left) and CIZ1 exon expression relative to individual exon 7 expression (right). *CIZ1B VAF* graph cropped to 25% led to visual omission of two samples from the normal set. The total number of outliers for the tumour and normal set is 6 and 15 respectively. In both A and B all cancer and normal adjacent data is extracted from the TCGA, whereas normal data is extracted from GTEx. Error bars show SEM. N=number of patients. Statistical analysis of *CIZ1B VAF* was conducted using a Mann Whitney U test, p=0.003

B. Breast samples *CIZ1B* VAF relative to normal junction 14/15 expression (left) and CIZ1 exon expression relative to individual exon 7 expression (right). *CIZ1B VAF* graph cropped to 25% led to visual omission of one sample from the tumour set. The total number of outliers for the tumour and normal set is 30 and 3 respectively. Statistical analysis of *CIZ1B VAF* was conducted using a Mann Whitney U test, p=8.71x10^{-114}

C. CIZ1 TPM values in normal adjacent and cancer samples from breast and lung tissue. Graph generated using GEPIA utilising TCGA data. In order, N=1085,112, 483 and 59.

5.5. Discussion

5.5.1 CIZ1 AD nucleic acid interactions

A functional impact of CIZ1B could be caused by loss or gain of function affecting protein-nucleic acid interaction or protein-protein interaction, or caused by structural changes.

It would not be anticipated that there would be any differences in nucleic acid binding as a direct consequence of the eight amino acid loss, since despite this loss there is still a net negative charge at the AcD that would repel the negatively charged phosphate backbone of RNA and/or DNA. It is also known that the C-terminus as an entity does not bind RNA as effectively as the N-terminus (Sofi et al., 2022), and as described in Chapter 3 the small amount of RNA binding that occurs is at the extreme C-terminus of CIZ1. Thus, it is unlikely that there are any differences in nucleic acid binding between CIZ1 and CIZ1B, but this would have to be tested.

5.5.2. Changes in CIZ1B structure?

The analysis in this chapter identified no differences in quaternary structure between H720 and H720B, as both exist in the same dimeric state. However, complete deletion of the AcD from murine CIZ1 did appear to cause a confirmational change as seen by a shift in SEC elution time and structural differences predicted when using the AlphaFold2 programme (Mirdita et al., 2022). It should be stated that whilst SEC is an assay that should be insensitive to charge changes, the dramatic pI change that occurs upon deletion of the AcD (4.97 to 9.17) could be contributing to the shift in elution time. The manufacturer suggests that higher salt concentrations can be used to suppress non-specific ionic interactions between positively charged proteins and the stationary phase, so this could be tested to ensure that this is not a contributing factor here. To further analyse potential changes in secondary structure upon AcD loss other techniques could be explored such as circular dichroism, which measures the absorption of polarised light to identify what secondary structure is present in a protein of interest.

However, since it is unlikely that CIZ1B displays any differences in protein-nucleic acid interactions and the changes that were identified in protein structure are minor, this suggests that protein-protein interactions (loss or gain) are the most likely candidate to facilitate CIZ1B corruption leading to tumourigenesis.

5.5.3. CIZ1 epigenetic maintenance and nucleolar proteins

The MS analysis in this chapter identified many binding partners, and differences were seen between H720 and H720B.

We have chosen not to focus on the large number of nucleolar proteins that were identified, but they may be relevant to the mechanism CIZ1 plays in epigenetic maintenance. We have previously shown that CIZ1 null cells show a genome wide reduction in the repressive marks H3K27me3 and H2AK119ub1 and a subsequent de-regulation of polycomb regulated genes (Stewart et al., 2019). We suggest that this is caused by the loss of movement of the Xi (and indeed other foci), from the periphery of the nucleus to sites adjacent to nucleoli during S-phase. This is supported by the emerging evidence that the nucleolus plays additional roles in maintenance of genomic stability, including the DNA damage response (DDR) and epigenetic regulation (Lindstrom et al., 2018). Perhaps Xi movement is facilitated by CIZ1's affinity with nucleolar proteins, and reduction in these interactions for H720B could compromise this mechanism. Since we did not include RNase in our sample preparation steps, we cannot fully rule out that some of these interactions could be indirect interactions bridged by RNA in our binding reactions (derived from bacterial or HeLa cells). However, whilst murine C-terminal fragments have been shown to be capable of binding RNA promiscuously, they did not interact with 18S rRNA (Sofi et al., 2022), which would argue against nucleolar RNA as a bridge between CIZ1 and other RNA binding proteins. An ongoing project is currently underway to investigate human CIZ1 C-terminal protein:protein interactions using an alternative H720 mutant which lacks the extreme C-terminus that we now know to be implicated in CIZ1 C-terminal RNA interactions.

5.5.4. CIZ1 interaction with linker histones

Histones were extensively identified as CIZ1 interaction partners in this study. In eukaryotes DNA is packaged into nucleosomes that consist of DNA wrapped around eight core histones (four dimers of H2A, H2B, H3 and H4), resembling "beads on a string" (Olins and Olins, 1974). These nucleosomes are further compacted by linker histones that bind at DNA entry and exit sites. There are 11 protein isoforms of H1, including the 7 somatic histones H1.0-H1.5, H1.10, and 4 germ cell specific histones H1oo, H1t, H1T2 and HILS1 (Hergeth and Schneider, 2015). The somatic H1 variants are further categorised based on their expression profile, as either replication dependent (H1.1-H1.5) or replication independent (H1.0 and H1.10) (Mayor et al., 2015). Here we identified that H720 and H720B had the same affinity to the core histones (Appendix D), but contrastingly H720 interacted more strongly with linker histones H1.0, H1.5 and H1.10, compared to H720B. It should also be noted that H1.2 and H1.3 were also identified in this screen, but these did not meet the criteria of at least two unique peptide sequences so were subsequently removed. This is likely due to the high sequence identity of the replication dependent H1 subtypes (Hergeth and Schneider, 2015) making it less likely that the peptides will be unique to a particular subtype and instead that amino acid sequence is conserved across H1 types. This together suggests that CIZ1 has no preferential binding to replication dependent or replication independent linker histones, but that H720B shows decreased affinity to linker histones generally.

The precise role of linker histones compared to core histones remains a subject of intensive study. Upon H1 loss there is an increased propensity to DNA breaks and DNA:RNA hybrids in heterochromatin (Bayona-Feliu et al., 2017), suggesting that it plays a protective role. Additionally linker histones are implicated in protein:protein interactions, with suggestions that they could act as recruitment hubs for proteins involved in regulation of chromatin status (McBryant et al., 2010, Kalashnikova et al., 2016), such as interaction with DNA methyltransferases DNMT1 and DNMT3B in embryonic stem cells (Yang et al., 2013) and polycomb repressive complex 2 (PRC2) for H3K27me3 modifications in somatic cells (Willcockson et al., 2021). Finally, linker histones have been shown to promote compaction by liquid-liquid phase separation (LLPS) (Gibson et al., 2019), which CIZ1 has also been implicated in (Sofi et al., 2022). Attempting to draw a conclusion of the significance of H1.0, H1.5 and H1.10 being identified in the screen is challenging since the exact individual roles of H1 subtypes remains quite unknown. H1.0 has been shown to interact with proteins located in the nucleolus (Kalashnikova et al., 2013) and H1.10 demonstrates G1 nucleolar accumulation (Stoldt et al., 2007), thus CIZ1 interaction with these subtypes could be relevant given the high propensity of nucleolar proteins identified in this interaction study.

Interestingly H1.0 has also been implicated in cell differentiation (Zlatanova and Doenecke, 1994), and in cancer H1.0 levels correlate with tumour differentiation status, where silencing facilitates long term proliferative potential (Torres et al., 2016). Additionally, melanoma cells have been shown to excrete H1.0 to escape differentiation (Schiera et al., 2016). However, the biological significance of the reduced affinity associated with alternative spicing of the AcD in CIZ1B remains to be explored further.

5.5.5. CIZ1 involvement in the DNA damage response

Other interactions identified in this study that are of note include two heat shock proteins (Hsp) and PARP-1. PARP-1 inhibitors have been utilised in the clinic for patients with BRCA mutations to preferentially target cancer cells based on their compromised DNA repair pathways (Chen, 2011). Additionally, it has been identified that Hsp70 translocates to the nucleus and binds to PARP-1 to protect HeLa cells from single stranded breaks (Kotoglou et al., 2009). Thus, it of interest here that a cancer associated form of CIZ1 shows reduced interaction with these proteins. It is already known that CIZ1 is involved in the DDR for protection from double stranded breaks (DSBs), as seen by an increase in DSBs in CIZ1 null mice compared to WT mouse embryonic fibroblasts (Khan et al., 2018). The interactions identified here provide preliminary evidence to suggest that CIZ1 could also be involved in protection against single stranded breaks by interacting with multiple players in the DDR.

In addition to this, mitochondrial transcription factor A (TFAM) was identified as a H720 interactor. TFAM is also found in the nucleus anchored to chromatin (Lee et al., 2014). In the mitochondria TFAM protects mitochondrial DNA (mtDNA) from damage by reducing production of reactive oxygen species (ROS) (Xu et al., 2009, Kunkel et al., 2018), however too much TFAM can lead to excessive repression of mtDNA and deficient oxidative phosphorylation (Bonekamp et al., 2021), so there is a fine balance to be maintained. In the nucleus TFAM does not appear to bind robustly to any sites (Wang et al., 2013), but is involved in negative self-regulation via suppression of NRF-1 leading to down regulation of TFAM expression (Lee et al., 2014). If indeed CIZ1 is interacting with TFAM in the nucleus *in vivo* it could be implicated in TFAM self-regulation, and thus upon loss or upregulation of cancer associated CIZ1 forms acting in a dominant negative mechanism the fine tuning of this mechanism could be affected. Additional evidence to support this is the involvement of p21 in TFAM expression, whereby loss of p21 leads to upregulation of TFAM expression (Kim et al., 2013). Since CIZ1 modulates p21 (Mitsui et al., 1999), in a setting where CIZ1 is corrupted this could lead to the opposite outcome and a downregulation of TFAM via upregulation of p21. This could leave mtDNA susceptible to damage via ROS, as is seen with nuclear DNA upon CIZ1 loss (Khan et al., 2018).

5.5.6. Mechanism of CIZ1 excretion into plasma?

The CIZ1B biomarker is detected in blood plasma from lung cancer patients (Higgins et al., 2012, Coverley et al., 2017), however the process by which it passes from the nucleus of cells into the blood stream has not been explored and remains an important open question. Coatomer subunit alpha (COPA) was returned as the most significant interaction partner with preference for H720B (q=1.9×10^{-4}).

COPA is a subunit of the coat protein complex 1 (COPI). The accepted common function of COPI is in Golgi to endoplasmic reticulum (ER) recycling (retrograde transport), whereas COPII has been implicated in ER to Golgi transport (anterograde transport) (Arakel and Schwappach, 2018). However, there are still many details of COPI function and structure that remain unknown, and indeed early literature claimed COPI potentially had a role in anterograde transport (Gaynor and Emr, 1997, Orci et al., 1997, Schekman and Mellman, 1997, Orci et al., 2000) and there was Golgi to ER protein recycling that was COPI independent (Storrie et al., 2000). A more recent publication has implicated COPI in bidirectional protein transport, under the regulation of GTPase CDC42 (Park et al., 2015). Regardless of the direction COPI movement, microtubules are required for anterograde (Fourriere et al., 2020) and retrograde (Pietrantoni et al., 2020) transport by providing a scaffold for kinesin or dynein attachment, thus it is unsurprising that two tubulin proteins were also identified in our screen. If COPI is involved in anterograde protein transport this could provide a mechanism for CIZ1 excretion into plasma. Thus heightened affinity for CIZ1B, could underpin CIZ1B abundance in plasma. Interestingly COPA has itself been implicated in cancer development via the mTOR signalling pathway in hepatocellular carcinoma (Song et al., 2021), and is also found to be upregulated in cervical cancer where it is being explored as a biomarker (Bao et al., 2022).

COPA is also a precursor protein from which the gastrointestinal hormone Xenin is believed to be produced (Chow and Quek, 1997). Xenin is secreted into plasma after eating (Feurle et al., 1992). The proteases implicated in the cleavage event that produces Xenin could be implicated in CIZ1 cleavage to create the CIZ1B fragments identified in patient plasma (Coverley et al., 2017).

5.5.7. Concluding remarks and future work

The MS interaction study did not identify any protein interactions that upon inspection of the relevant literature could provide an obvious rationale for the elevated CIZ1B Xi accumulation frequency compared to WT CIZ1. Perhaps in the context of full length CIZ1 there are additional changes in protein:protein interactions that have not been identified in the CIZ1 C-terminal fragments. In Chapter 3 I showed the CIZ1 C-terminal fragments exist as a dimer via interactions mediated by the MH3 domain, and here I showed that the partial or full loss of the AcD had no effect on multimeric state, thus the differences in interaction partners observed is not due to changes in CIZ1 multimeric state. An interesting avenue of future work would be to monitor difference in CIZ1 C-terminal interaction partners upon MH3 loss to measure if CIZ1 dimerisation is required for any of these partners.

Whilst there were no differences in CIZ1 quaternary structure upon AcD loss, there was preliminary evidence to suggest changes in CIZ1 secondary or tertiary structure, and perhaps these changes are responsible for the decreased binding affinity to the majority of biding partners. Whilst there were lots of binding partners of interest that were observed here, before a particular avenue is explored further, it would be pertinent to confirm a CIZ1 binding partner using a Co-IP assay to confirm that this interaction occurs *in vivo* with endogenous CIZ1, not just in an *in vitro* CIZ1 overexpression context.

Finally, I presented preliminary evidence to suggest that there are different events of aberrant CIZ1 expression occurring in different cancer tissue types. This has implications when attempting to deconvolute what mechanisms are being corrupted in early cancer development in different cancers, and suggests that we may not be able to extrapolate data from one cancer to another, since we do not know if there are any shared downstream consequences of *CIZ1B* expression compared to *CIZ1* 3' transcript elevation. Extended profiling of the events occurring and subsequent consequences in different tissue types is required before we can draw any conclusions, and this further highlights the complexity of altered CIZ1 expression in cancer.

6. Discussion

6.1. CIZ1 as a protector of epigenome and genome

Previous work has shown that deletion of CIZ1 caused loss of H2AK119ub1 and H3K27me3 at the inactive X chromosome (Xi) (Stewart et al., 2019), and the introduction of ectopic full length CIZ1 in CIZ1 null cells led to recovery of H2AK119ub1 and H3K27me3 (Sofi et al., 2022). Loss of these epigenetic marks in CIZ1 null cells correlated with deregulation of genome wide polycomb regulated genes (Stewart et al., 2019). Here I showed that the expression of dominant negative (DN) fragments also resulted in the loss of these epigenetic marks (Chapter 3 and 4), thus these DN fragments could mirror the environment of a CIZ1 depleted cell, and profiling the changes in gene expression in CIZ1 null cells and cells expressing CIZ1 DN fragments and comparing the results presents as an interesting follow up analysis. Since CIZ1 is a gene that is prone to alternative splicing events (Rahman et al., 2010), some of the documented cases of CIZ1 overexpression could actually be examples of alternative splicing leading to DN effects that could essentially mimic CIZ1 loss. In fact if only the 3' CIZ1 mRNA had been profiled in this body of work, it could have been wrongly concluded that it was another example of CIZ1 overexpression. This highlights the need for rigorous profiling of aberrant gene expression before conclusions can be drawn regarding how it is deviated from normal expression.

Since upon long term culturing of CIZ1 null cells there was partial recovery of H2AK119ub1 and H3K27me3 accumulation at the Xi (Stewart et al., 2019), this suggests the existence of a compensatory mechanism to restore proper gene regulation. This, in line with the observation that the Xi is highly stable and has only been re-activated upon cell reprogramming to an early developmental stage (Wutz, 2011), could provide a rationale for why there was only activation of a subset of X-linked genes in primary embryonic fibroblasts (PEFs) (3.6%) upon CIZ1 loss (Ridings-Figueroa et al., 2017). Contrastingly, in agreement with the suggestion that some cell types are more reliant on *Xist* for maintenance of the Xi (Loda et al., 2022), CIZ1 loss in female mice spleens led to dysregulation of 25% of the X-linked genes by greater than two fold, (alongside other genome wide changes), possibly implicating this CIZ1 associated X-linked de-regulation with development of the lymphoproliferative disorder observed. Whilst PEFs exposed to long term culture with chronic CIZ1 loss potentially developed a compensatory mechanism, this may not be the case with cells exposed to DN fragments, since as shown in Chapter 4 DN fragments behave in a dynamic manner and are only able to exert their function in early G1. Thus, perhaps they could be more deleterious than a CIZ1 null environment if the cell is not able to overcome this corruption.

It has been documented that PRC1/2 targets are often deregulated in cancer (reviewed in (Parreno et al., 2022)) and that there is aberrant gene activity from the Xi in primary breast tumours (Chaligné et al., 2015). Since the DN fragments led to loss of epigenetic modifications from the Xi, it could be postulated that gene deregulation might also be observed at the Xi in this instance. To measure this RNA extraction samples collected from cells expressing DN fragments are currently in the process of being analysed.

In CIZ1 null cells there is evidence of DNA damage and aberrant cell cycle progression (Khan et al., 2018) and leukaemias (Nishibe et al., 2013), so clearly CIZ1 plays an important role in correct cell cycle progression and preventing DNA damage. Additionally, it has been shown that mice are tolerant to *Xist* deletion, until exposed to carcinogens and inflammatory agents, suggesting *Xist* also plays a protective role (Yang et al., 2020b), possibly mediated by CIZ1 and other proteins. Here I observed that the C-terminus of CIZ1 directly interacts with some of the components of the DNA damage response (DDR) (Chapter 5). Thus, the overexpression of C-terminal DN fragments in cancers (Chapter 3) could be perturbing this important fidelity mechanism, by sequestering interaction partners. Additionally the C-terminus of CIZ1 is responsible for interactions with p21 (Mitsui et al., 1999), which restrains cell cycle progression and is a known tumour suppressor (Shamloo and Usluer, 2019). Thus, overexpression of C-terminal fragments could also be leading to increased inhibition of p21 activation, which could also be leading to the accumulation of damaged DNA. Indeed it has been shown that a reduction in p21 has been associated with an upregulation of stemness and EMT in cells (Rohnalter et al., 2015). Therefore, overall C-terminal DN fragments could be leading to genetic and epigenetic changes, but this requires follow up analysis.

6.2. C-terminal fragments implicated in nuclear matrix breakdown in cancer?

The corruption of the nuclear matrix (NM) in tumours is well documented (reviewed in (Lever and Sheer, 2010)), and since NM proteins are implicated in core processes such as DNA replication, RNA synthesis and processing, and nuclear transport, aberrations could have reaching effects. Additionally it has been documented that there are changes in protein immobilisation to the NM in differentiated and undifferentiated cells, for example cyclin E has been shown to be immobilised on the NM in a variety of differentiated vertebrate cells, but not in undifferentiated and cancer cells (Munkley et al., 2011). Since cancer can be thought of as a reversion to an undifferentiated state (Carvalho, 2020), this suggests that the solubility (and diffusibility) of cell cycle factors such as cyclin E could be implicated in the plasticity of cancer cells. This is interesting when considering the role of CIZ1 in cyclin delivery during the G1/S transition (Copeland et al., 2010), and suggests that aberrant CIZ1 expression could be associated with this loss of NM binding.

Polycomb group proteins (PcGs) are key regulators of developmental genes in human (Lee et al., 2006) and mice (Boyer et al., 2006) embryonic stem cells, and as mentioned above PRC1/2 targets are often deregulated in tumours (Parreno et al., 2022). It has been shown that PcGs exert their repressive effects by epigenetic modifications of histones, but also by causing architectural changes to chromatin, as observed by PRC2 mediated long-range intra- and inter-chromosomal interactions (Tiwari et al., 2008). Tying together the observations that CIZ1 loss led to corruption of PcG mediated regulation (Stewart et al., 2019), and the observation that CIZ1 overexpression led to a reduced accumulation of *Xist* at the Xi (Sunwoo et al., 2017) where it was hypothesised to be saturating the NM binding sites for *Xist* accumulation, this could suggest that under normal circumstances and at normal levels, CIZ1 could act as a bridging protein for PcG and *Xist* anchoring to the NM. This suggestion is supported by the observation that upon CIZ1 loss, EZH2 (a subunit of PRC2) had increased extractability in high salt, RNAse and DNase, which in primary WT cells it was resistant to (Stewart et al., 2019).

Considering this, expression of CIZ1 DN fragments could corrupt NM binding by saturating NM binding sites as described by Sunwoo et al. Since it is the C-terminal AD of CIZ1 that is implicated in NM attachment (Ainscough et al., 2007) this seems plausible. If this is indeed occurring it could exacerbate NM corruption in tumours, or indeed could even be one of the driving events of NM breakdown. Therefore an interesting future line of experimentation would be to identify if there is an increase in protein extractability when DN fragments are present, indicating reduced immobilisation on the NM.

6.3. Effect of CIZ1B splicing on CIZ1 function

One of my goals was to explore, and derive insight into, the effect that partial exclusion of exon 14 might have on CIZ1 function, and to understand whether it might play a role in promoting lung cancer. The interaction study in Chapter 5 comparing the binding affinity of a WT and CIZ1B human C-terminal fragments identified perturbed binding affinity of CIZ1B to several proteins. As discussed, this included components of the DDR (HSPA1B, HSPA5 and PARP1), and H1.0, H1.5 and H1.10 variants of the H1 linker histone.

Interestingly CIZ1 and CIZ1B interacted with other histone subtypes with equal affinity, including MACROH2A1, that is well known for being enriched at the Xi (Costanzi and Pehrson, 1998), therefore suggesting CIZ1B corruption implicates a mechanism that is specific to the H1 subtype. As described in Chapter 5, linker histones have been suggested to act as recruitment hubs for proteins involved in regulation of chromatin status, including PRC2 (Willcockson et al., 2021), and have also been shown to promote compaction by liquid-liquid phase separation (LLPS) (Gibson et al., 2019). Since CIZ1 has phase separation qualities (Sofi et al., 2022), and is crucial for PRC2 function at the Xi (Stewart et al., 2019), perhaps it is implicated in both of these roles of linker histones. Loss of H1 has also been associated with DNase hypersensitivity of pluripotency factors such as *Oct4* and *Sox2*, suggesting that it is normally enriched at these sites (Geeven et al., 2015), and in addition it has been shown to be enriched at the Xi (Chadwick and Willard, 2003). If CIZ1 is somehow involved in this enrichment, and this is subsequently perturbed by expression of CIZ1B, this could lead to reactivation of pluripotency factors and Xi repression erosion, as described above when considering the possible implications of the CIZ1 C-terminal fragments. Finally, loss of H1 has been associated with changes in chromosome compaction and an increase in nuclear volume (Hashimoto et al., 2010). This mirrors the phenotype observed when CIZ1 is absent, and cells undergo subsequent rounds of entering and exiting the cell cycle (Dobbs et al, unpublished). This could suggest that similarly to CIZ1 C-terminal fragments, CIZ1B also shares features of CIZ1 null cells.

Thus overall, upon CIZ1B expression with a perturbed ability to interact with linker histones or DDR components, this could be highly determinantal for epigenetic maintenance, and could also lead to accumulation of somatic mutations. So, as suggested above with CIZ1 C-terminal fragment overexpression, CIZ1B expression could also be implicated in both genetic instability and changes to epigenetic regulation, but via different mechanisms.

6.4. Relationship between CIZ1 expression and histone content

Histone chaperones are upregulated in a variety of cancers, and are categorised as a tumour promoting factors (Ray-Gallet and Almouzni, 2022), and histone density has been observed to be aberrated in cancer cell lines (Bruhn et al., 2022). An early observation in this body of work and others in the lab, is that histone quantity changes in response to changes in CIZ1 expression. It was observed that upon lentiviral transduction of CIZ1 C-terminal fragments, this led to overexpression of histone content relative to actin (Chapter 4). Additionally, when monitoring histone content in a normal breast epithelium cell line and in breast cancer cell lines, compared to primary breast epithelial cells, it was noted that histone levels were consistently higher in the cell lines relative to actin levels (Chapter 3). This was initially attributed to the altered karyotype of cancer cells, which can be approaching tetraploid (Nicholson and Cimini, 2013), and the acknowledgment that normal cell lines upon long term culturing can have phenotypes that are intermediate between normal and cancer tissue (Khanuja et al., 1993), and also have altered karyotypes (Bloomfield and Duesberg, 2015). However, coupling these observations with the observation that CIZ1 AD interacts directly with histones (Chapter 5), and that upon CIZ1 loss there appears to perturbations in histone levels relative to CIZ1 WT cells (Coverley lab ongoing work), suggests an intimate relationship between CIZ1 and histones.

Since CIZ1 has been implicated in the movement of the Xi towards the nucleolus in S phase (Stewart et al., 2019), which is required to maintain the epigenetic state of the Xi (Zhang et al., 2007), perhaps CIZ1 somehow mediates interactions between the replicating Xi and these histone chaperone complexes. Since it has been shown that there is an inverse correlation between histone content and RNA pol II occupancy (Das and Tyler, 2012), perhaps upon CIZ1 loss or dispersal from the Xi, the cell senses deregulation of polycomb regulated genes, and upregulates histone production as a compensatory mechanism. It should be stated that this is speculation in an attempt to rationalise how altered CIZ1 could affect cellular histone content, and this requires rigorous formal testing.

A further possible functional relationship relates to histone production. CIZ1 is implicated in cyclin E/A delivery to activate CDK2 (Copeland et al., 2010) and cyclin E/CDK2 has been shown to phosphorylate S phase-specific substrates such as NPAT, which is involved in the activation of histone gene transcription (Zhao et al., 1998). This could provide a link between CIZ1 overexpression and increased histone content.

6.5. Concluding remarks

Here I attempted to deconvolute the conflicting evidence in the literature of CIZ1 over-expression and under-expression both being associated with tumourigenesis, and how mutated forms of CIZ1 also sit in these pathways. I showed that whilst over-expression of CIZ1 C-terminal fragments and expression of the CIZ1B cancer associated splice variants presented with different phenotypes, they may both be involved in corruption of epigenetic maintenance and an impaired DDR. This does not seem unreasonable given the existing literature of epigenetic corruption (Stewart et al., 2019) and accumulation of DNA damage (Khan et al., 2018) upon CIZ1 loss. However, it should be stated that this is a preliminary hypothesis and is subject to rigorous follow up testing to confirm.

Work presented here and elsewhere indicated that cells are sensitive to subtle *in vivo* changes in CIZ1, and that there is a fine balance that has to be achieved for homeostasis. This suggests that it would likely be too difficult to design pharmaceuticals that target CIZ1 misexpression, without being counterproductive and creating the inverse effect. However, the use of CIZ1B as a biomarker is under development (Coverley et al., 2017), and if follow up work to characterise CIZ1B levels in other cancer tissue types is conducted and confirmed, this could convert the use of the CIZ1B biomarker assay for lung cancer diagnosis into a far reaching diagnostic tool.

In addition to this if the uncoupled expression of CIZ1 RD and AD documented in breast cancers is characterised in other tissue types, this could present as an additional biomarker opportunity, but of course would be harder to design and implement. Regardless, a better understanding of the role of CIZ1 in tumourigenesis, especially a more in-depth profiling of both aberrant expression and downstream consequences, could eventually improve patient outcomes, so is a path worth pursuing.

Appendices

Appendix A. Full SEC traces presented in Chapter 3 and Chapter 5

Murine SEC traces
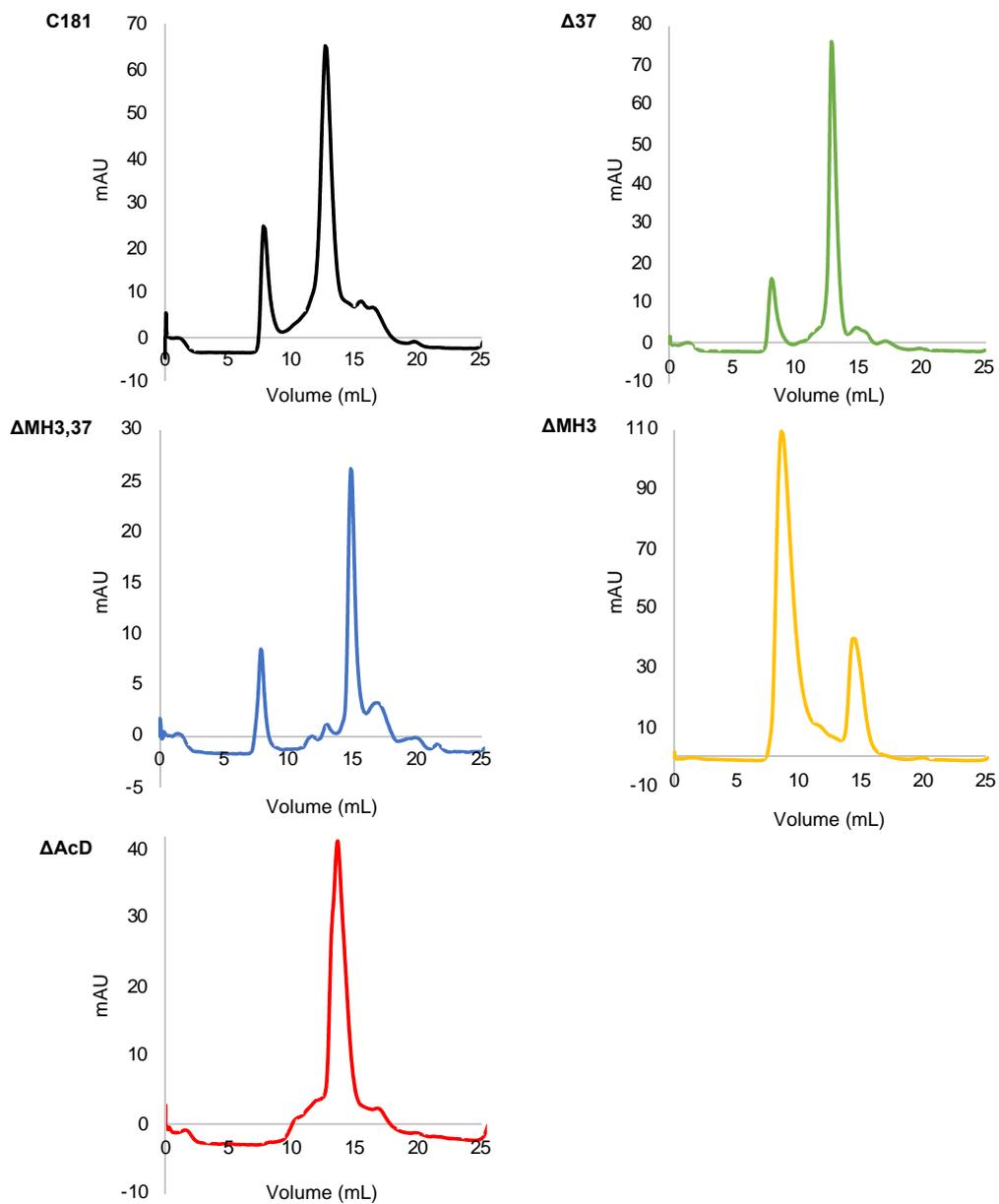
Human SEC traces
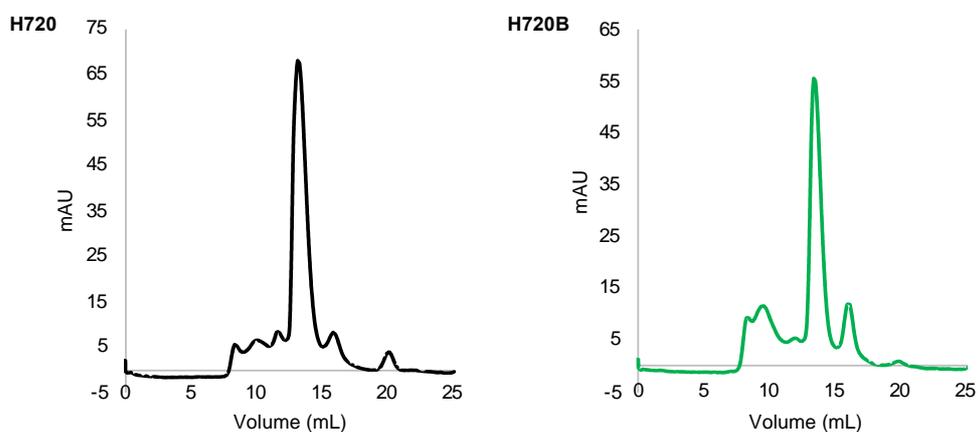
Appendix B. Full SEC-MALLs traces presented in Chapter 3 and Chapter 5

Murine SEC-MALLs traces
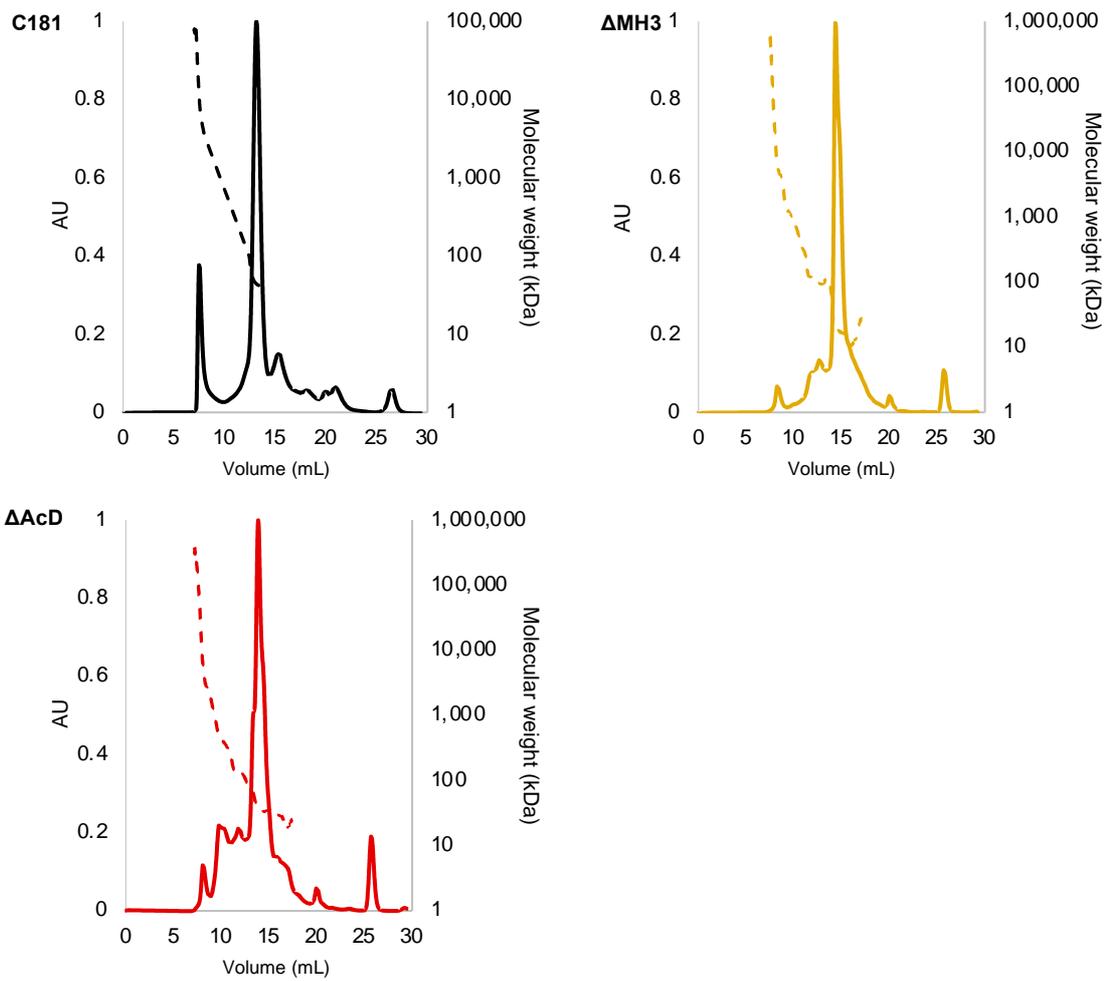
Human SEC-MALLs traces
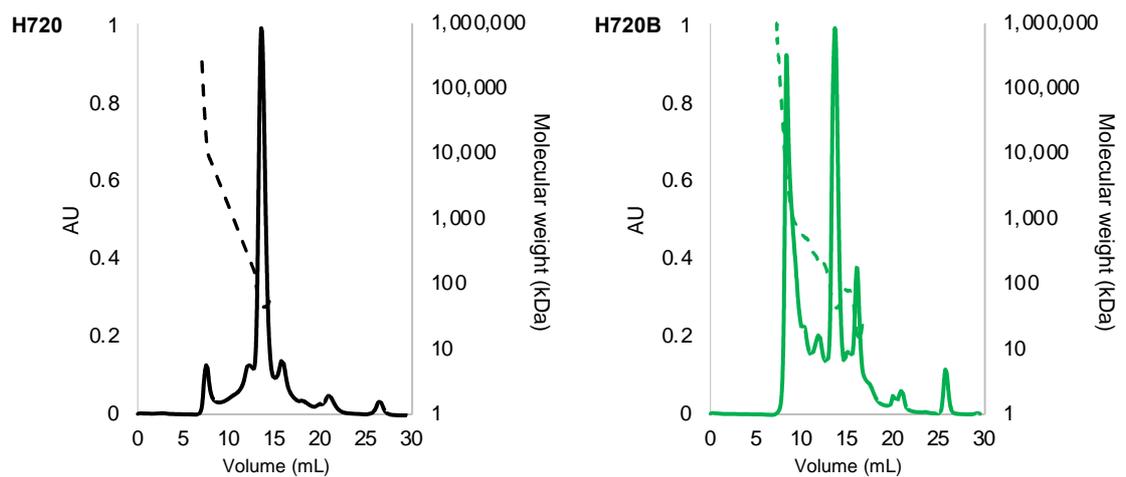
Appendix C. Full list of interaction partners identified by mass spectrometry in Chapter 5

Interaction partners are categorised as: binding CIZ1 more strongly than the negative control (CIZ1>NC), binding the negative control more strongly than CIZ1 (NC>CIZ1) or no differences between the three sets, (NC=CIZ1) respectively.

Number	Name	Highest mean condition	3-Way ANOVA q value	Group
1	DDX56	WT/B	1.4×10^{-11}	CIZ1>NC
2	UBTF	WT/B	1.4×10^{-11}	CIZ1>NC
3	CIZ1	WT/B	3.2×10^{-10}	CIZ1>NC
4	SRP14	WT/B	9.4×10^{-9}	CIZ1>NC
5	DDX31	WT/B	1.9×10^{-8}	CIZ1>NC
6	RPS26	WT/B	3.1×10^{-8}	CIZ1>NC
7	BSG	WT/B	5.8×10^{-8}	CIZ1>NC
8	TOP1	WT/B	9.6×10^{-8}	CIZ1>NC
9	NOP2	WT/B	9.4×10^{-8}	CIZ1>NC
10	DDX18	WT/B	9.6×10^{-8}	CIZ1>NC
11	RPL23A	WT/B	1.1×10^{-7}	CIZ1>NC
12	SRSF1	WT/B	1.0×10^{-7}	CIZ1>NC
13	TFAM	WT/B	1.7×10^{-7}	CIZ1>NC
14	SRSF2	WT/B	1.7×10^{-7}	CIZ1>NC
15	MYBBP1A	WT/B	1.9×10^{-7}	CIZ1>NC
16	GAPDH	WT/B	1.9×10^{-7}	CIZ1>NC
17	SPATS2L	WT/B	2.6×10^{-7}	CIZ1>NC
18	FBL	WT/B	2.5×10^{-7}	CIZ1>NC
19	DDX27	WT/B	2.6×10^{-7}	CIZ1>NC
20	RPS7	WT/B	2.6×10^{-7}	CIZ1>NC
21	HP1BP3	WT/B	4.1×10^{-7}	CIZ1>NC
22	RPL11	WT/B	4.7×10^{-7}	CIZ1>NC
23	DDX21	WT/B	5.5×10^{-7}	CIZ1>NC
24	CSNK2A3	WT/B	6.0×10^{-7}	CIZ1>NC
25	RBM28	WT/B	6.3×10^{-7}	CIZ1>NC
26	RPS6	WT/B	8.3×10^{-7}	CIZ1>NC
27	RPL23	WT/B	1.1×10^{-6}	CIZ1>NC
28	H1-10	WT/B	1.3×10^{-6}	CIZ1>NC
29	RPS14	WT/B	1.3×10^{-6}	CIZ1>NC
30	H2BC18	WT/B	1.3×10^{-6}	CIZ1>NC
31	RPS8	WT/B	1.3×10^{-6}	CIZ1>NC
32	SLTM	WT/B	1.3×10^{-6}	CIZ1>NC
33	DNAJC9	WT/B	1.4×10^{-6}	CIZ1>NC
34	H1-5	WT/B	1.4×10^{-6}	CIZ1>NC
35	SUPT16H	WT/B	1.5×10^{-6}	CIZ1>NC
36	RPL3	WT/B	1.6×10^{-6}	CIZ1>NC
37	RPS2	WT/B	2.2×10^{-6}	CIZ1>NC
38	H1-0	WT/B	2.8×10^{-6}	CIZ1>NC
39	RPS3A	WT/B	2.8×10^{-6}	CIZ1>NC
40	DHX9	WT/B	3.6×10^{-6}	CIZ1>NC
41	RPS16	WT/B	3.8×10^{-6}	CIZ1>NC
42	SRSF3	WT/B	3.9×10^{-6}	CIZ1>NC
43	RPL31	WT/B	4.5×10^{-6}	CIZ1>NC
44	TOP2A	WT/B	5.8×10^{-6}	CIZ1>NC
45	RPS4X	WT/B	6.6×10^{-6}	CIZ1>NC
46	HNRNPC	WT/B	6.8×10^{-6}	CIZ1>NC
47	RPL7A	WT/B	1.1×10^{-5}	CIZ1>NC
48	H2AZ1	WT/B	1.2×10^{-5}	CIZ1>NC
49	RBMX	WT/B	1.5×10^{-5}	CIZ1>NC
50	RPS3	WT/B	1.6×10^{-5}	CIZ1>NC
51	RPL7	WT/B	1.7×10^{-5}	CIZ1>NC

Number	Name	Highest mean condition	3-Way ANOVA q value	Group
52	PARP1	WT/B	2.3×10^{-5}	CIZ1>NC
53	H1-1	WT/B	2.6×10^{-5}	CIZ1>NC
54	RPL27	WT/B	2.8×10^{-5}	CIZ1>NC
55	RACK1	WT/B	2.8×10^{-5}	CIZ1>NC
56	NOP56	WT/B	4.0×10^{-5}	CIZ1>NC
57	RPS27	WT/B	4.4×10^{-5}	CIZ1>NC
58	RSL1D1	WT/B	5.1×10^{-5}	CIZ1>NC
59	RPS5	WT/B	7.6×10^{-5}	CIZ1>NC
60	H2BC11	WT/B	8.1×10^{-5}	CIZ1>NC
61	RPSA	WT/B	8.4×10^{-5}	CIZ1>NC
62	PDCD11	WT/B	8.8×10^{-5}	CIZ1>NC
63	RPL19	WT/B	1.2×10^{-4}	CIZ1>NC
64	H4C8	WT/B	1.3×10^{-4}	CIZ1>NC
65	NIFK	WT/B	1.6×10^{-4}	CIZ1>NC
66	GNL3	WT/B	2.3×10^{-4}	CIZ1>NC
67	RPLP0	WT/B	2.3×10^{-4}	CIZ1>NC
68	SRP72	WT/B	2.7×10^{-4}	CIZ1>NC
69	SMARCA5	WT/B	3.3×10^{-4}	CIZ1>NC
70	H3C13	WT/B	3.7×10^{-4}	CIZ1>NC
71	RPS18	WT/B	4.9×10^{-4}	CIZ1>NC
72	NCL	WT/B	5.0×10^{-4}	CIZ1>NC
73	SSRP1	WT/B	5.1×10^{-4}	CIZ1>NC
74	RPL15	WT/B	7.3×10^{-4}	CIZ1>NC
75	CCDC86	WT/B	8.4×10^{-4}	CIZ1>NC
76	RPL30	WT/B	8.7×10^{-4}	CIZ1>NC
77	MACROH2A1	WT/B	1.3×10^{-3}	CIZ1>NC
78	RPL35	WT/B	1.4×10^{-3}	CIZ1>NC
79	RPS19	WT/B	1.6×10^{-3}	CIZ1>NC
80	RPS13	WT/B	2.2×10^{-3}	CIZ1>NC
81	BRIX1	WT/B	2.4×10^{-3}	CIZ1>NC
82	EEF1B2	WT/B	2.4×10^{-3}	CIZ1>NC
83	HSPA5	WT/B	2.7×10^{-3}	CIZ1>NC
84	NOP58	WT/B	3.0×10^{-3}	CIZ1>NC
85	SLC25A5	WT/B	7.2×10^{-3}	CIZ1>NC
86	RPL18	WT/B	8.5×10^{-3}	CIZ1>NC
87	BAZ1B	WT/B	2.0×10^{-2}	CIZ1>NC
88	RTRAF	WT/B	4.6×10^{-9}	CIZ1>NC
89	FAM98A	WT/B	4.0×10^{-9}	CIZ1>NC
90	TUBB	WT/B	1.0×10^{-8}	CIZ1>NC
91	DIMT1	WT/B	3.2×10^{-8}	CIZ1>NC
92	COPA	WT/B	7.3×10^{-8}	CIZ1>NC
93	NAT10	WT/B	9.1×10^{-8}	CIZ1>NC
94	REXO4	WT/B	8.7×10^{-8}	CIZ1>NC
95	RPS15A	WT/B	1.0×10^{-7}	CIZ1>NC
96	SAFB	WT/B	9.4×10^{-8}	CIZ1>NC
97	DDX24	WT/B	1.1×10^{-7}	CIZ1>NC
98	RRP1B	WT/B	1.1×10^{-7}	CIZ1>NC
99	NUMA1	WT/B	2.6×10^{-7}	CIZ1>NC
100	RPS24	WT/B	5.1×10^{-7}	CIZ1>NC
101	RTCB	WT/B	9.0×10^{-7}	CIZ1>NC
102	SAFB2	WT/B	1.4×10^{-6}	CIZ1>NC
103	RPS11	WT/B	1.4×10^{-6}	CIZ1>NC
104	RPL13A	WT/B	1.9×10^{-6}	CIZ1>NC
105	CSNK2A2	WT/B	2.0×10^{-6}	CIZ1>NC
106	KNOP1	WT/B	2.0×10^{-6}	CIZ1>NC
107	RPL17	WT/B	2.3×10^{-6}	CIZ1>NC
108	RPL13	WT/B	3.0×10^{-6}	CIZ1>NC
109	RPL10	WT/B	1.7×10^{-5}	CIZ1>NC
110	TUBB6	WT/B	1.7×10^{-5}	CIZ1>NC

Number	Name	Highest mean condition	3-Way ANOVA q value	Group
111	TUBA4A	WT/B	1.9×10^{-5}	CIZ1>NC
112	SRSF10	WT/B	2.0×10^{-5}	CIZ1>NC
113	GPATCH4	WT/B	1.1×10^{-4}	CIZ1>NC
114	TUBA1C	WT/B	1.7×10^{-4}	CIZ1>NC
115	MKI67	WT/B	2.1×10^{-4}	CIZ1>NC
116	DDX1	WT/B	2.8×10^{-4}	CIZ1>NC
117	RPLP2	WT/B	3.4×10^{-4}	CIZ1>NC
118	RPL4	WT/B	5.4×10^{-4}	CIZ1>NC
119	ACTB	WT/B	1.6×10^{-3}	CIZ1>NC
120	RBM34	WT/B	2.0×10^{-3}	CIZ1>NC
121	RPS9	WT/B	2.0×10^{-2}	CIZ1>NC
122	GSTM3	pGEX	1.7×10^{-5}	NC>CIZ1
123	LUC7L2	pGEX	1.4×10^{-4}	NC>CIZ1
124	LANCL1	pGEX	1.6×10^{-4}	NC>CIZ1
125	GSTM1	pGEX	2.5×10^{-4}	NC>CIZ1
126	HSPA1A	pGEX	5.9×10^{-4}	NC>CIZ1
127	HSPA9	pGEX	7.5×10^{-4}	NC>CIZ1
128	HSPA8	pGEX	1.5×10^{-3}	NC>CIZ1
129	CBR1	pGEX	7.6×10^{-3}	NC>CIZ1
130	AHCTF1	pGEX	1.7×10^{-2}	NC>CIZ1
131	TSFM	pGEX	3.8×10^{-2}	NC>CIZ1
132	XRCC5	WT/B	1.0×10^{-1}	NC=CIZ1
133	PRKDC	pGEX	1.1×10^{-1}	NC=CIZ1
134	EEF1A1P5	WT/B	1.2×10^{-1}	NC=CIZ1
135	HNRNPU	WT/B	2.3×10^{-1}	NC=CIZ1
136	EEF1G	WT/B	3.0×10^{-1}	NC=CIZ1
137	GSTP1	pGEX	3.8×10^{-1}	NC=CIZ1
138	XRCC6	pGEX	4.6×10^{-1}	NC=CIZ1
139	RBM39	pGEX	7.2×10^{-1}	NC=CIZ1

Appendix D. Comparison of H270 an H720B interaction partners identified by mass spectrometry after internal normalisation

Interaction partners are categorised as: binding H720 more strongly than H720B, binding H720B more strongly than H720, binding the negative control more strongly than H720 and H720B as identified in Appendix C, binding H720 more strongly than H720B but q>0.005, and binding H720B more strongly than H720 but q>0.005.

Number	Name	Highest mean condition	Fold change	t-test q value
1	NOP2	WT	7.6	3.7×10^{-5}
2	CCDC86	WT	3.0	5.5×10^{-5}
3	SRSF3	WT	6.0	7.6×10^{-5}
4	HSPA5	WT	1.8	7.6×10^{-5}
5	UBTF	WT	2.1	1.2×10^{-4}
6	RSL1D1	WT	4.2	1.9×10^{-4}
7	DDX21	WT	3.6	1.9×10^{-4}
8	HP1BP3	WT	4.9	2.2×10^{-4}
9	PARP1	WT	3.3	2.9×10^{-4}
10	GAPDH	WT	4.6	3.2×10^{-4}
11	RPL31	WT	1.5	3.2×10^{-4}
12	SRP14	WT	2.8	3.3×10^{-4}
13	DHX9	WT	6.4	3.3×10^{-4}
14	TOP1	WT	5.2	3.3×10^{-4}
15	RPL35	WT	2.1	4.4×10^{-4}
16	DDX27	WT	2.6	6.1×10^{-4}
17	H1-0	WT	3.0	6.9×10^{-4}
18	CIZ1	WT	1.3	9.9×10^{-4}
19	LUC7L2	WT	1.4	1.1×10^{-3}
20	BRIX1	WT	2.9	1.1×10^{-3}
21	RBMX	WT	2.5	1.2×10^{-3}
22	RPL15	WT	3.5	1.2×10^{-3}
23	RPS26	WT	1.5	1.3×10^{-3}
24	RBM28	WT	1.8	1.3×10^{-3}
25	H1-5	WT	2.5	1.5×10^{-3}
26	H1-10	WT	2.2	1.6×10^{-3}
27	RPS8	WT	2.5	1.6×10^{-3}
28	TFAM	WT	1.4	1.9×10^{-3}
29	RPL11	WT	1.6	2.1×10^{-3}
30	PDCD11	WT	1.7	2.5×10^{-3}
31	HSPA1B	WT	1.4	3.6×10^{-3}
32	RPL3	WT	2.5	4.4×10^{-3}
33	NOP58	WT	1.8	4.4×10^{-3}
34	RPL18	WT	2.6	4.4×10^{-3}
35	RPS7	WT	1.5	4.8×10^{-3}
36	SRP72	WT	1.4	4.9×10^{-3}
37	RPL19	WT	3.7	4.9×10^{-3}
38	COPA	B	2.2	1.9×10^{-4}
39	TUBB	B	1.7	7.6×10^{-4}
40	RRP1B	B	1.4	1.3×10^{-3}
41	TUBB6	B	1.5	1.6×10^{-3}
42	DDX1	B	1.9	2.0×10^{-3}
43	RPS11	B	1.4	3.6×10^{-3}
44	DDX24	B	1.5	3.9×10^{-3}
45	GSTM3	B (pGEX, appendix C)	1.9	7.3×10^{-4}
46	LANCL1	B (pGEX, appendix C)	1.7	1.1×10^{-3}
47	GSTM1	B (pGEX, appendix C)	1.7	3.1×10^{-3}

Number	Name	Highest mean condition	Fold change	t-test q value
48	H2BC11	WT (q>0.005)	1.7	5.4×10^{-3}
49	BSG	WT (q>0.005)	3.9	5.5×10^{-3}
50	RPL30	WT (q>0.005)	2.8	5.8×10^{-3}
51	HNRNPC	WT (q>0.005)	2.4	6.0×10^{-3}
52	RPL7	WT (q>0.005)	2.2	6.0×10^{-3}
53	SLC25A5	WT (q>0.005)	2.0	6.1×10^{-3}
54	SMARCA5	WT (q>0.005)	1.8	6.1×10^{-3}
55	H2BC18	WT (q>0.005)	1.6	6.8×10^{-3}
56	RPS18	WT (q>0.005)	2.4	7.0×10^{-3}
57	RPS2	WT (q>0.005)	1.6	9.9×10^{-3}
58	RPL23A	WT (q>0.005)	1.6	1.0×10^{-2}
59	DNAJC9	WT (q>0.005)	1.4	1.2×10^{-2}
60	RPS3A	WT (q>0.005)	1.4	1.2×10^{-2}
61	XRCC5	WT (q>0.005)	1.5	1.3×10^{-2}
62	TOP2A	WT (q>0.005)	2.4	1.5×10^{-2}
63	H1-1	WT (q>0.005)	3.6	1.5×10^{-2}
64	RPS6	WT (q>0.005)	1.4	1.6×10^{-2}
65	DDX18	WT (q>0.005)	1.2	2.1×10^{-2}
66	RACK1	WT (q>0.005)	1.9	2.1×10^{-2}
67	FBL	WT (q>0.005)	1.1	2.2×10^{-2}
68	HSPA8	WT (q>0.005)	1.3	2.3×10^{-2}
69	EEF1B2	WT (q>0.005)	1.9	2.4×10^{-2}
70	RPS16	WT (q>0.005)	1.5	2.4×10^{-2}
71	RPSA	WT (q>0.005)	1.7	2.5×10^{-2}
72	HNRNPU	WT (q>0.005)	2.1	3.1×10^{-2}
73	NCL	WT (q>0.005)	2.4	3.4×10^{-2}
74	SLTM	WT (q>0.005)	1.2	3.4×10^{-2}
75	RPLP0	WT (q>0.005)	1.6	3.8×10^{-2}
76	RPS5	WT (q>0.005)	2.1	3.8×10^{-2}
77	BAZ1B	WT (q>0.005)	1.6	3.9×10^{-2}
78	RPS3	WT (q>0.005)	1.4	4.0×10^{-2}
79	NOP56	WT (q>0.005)	1.2	4.0×10^{-2}
80	SSRP1	WT (q>0.005)	1.5	4.1×10^{-2}
81	SUPT16H	WT (q>0.005)	1.4	4.2×10^{-2}
82	RPS13	WT (q>0.005)	1.3	4.4×10^{-2}
83	RPL7A	WT (q>0.005)	1.3	4.7×10^{-2}
84	SRSF2	WT (q>0.005)	1.5	5.4×10^{-2}
85	MACROH2A1	WT (q>0.005)	1.4	6.2×10^{-2}
86	RPS19	WT (q>0.005)	1.4	6.7×10^{-2}
87	RPS14	WT (q>0.005)	1.1	7.5×10^{-2}
88	SRSF1	WT (q>0.005)	1.2	9.2×10^{-2}
89	XRCC6	WT (q>0.005)	1.1	9.9×10^{-2}
90	RPS27	WT (q>0.005)	1.3	1.1×10^{-1}
91	GNL3	WT (q>0.005)	1.2	1.2×10^{-1}
92	RPS4X	WT (q>0.005)	1.1	1.3×10^{-1}
93	H2AZ1	WT (q>0.005)	1.2	1.4×10^{-1}
94	H4C11	WT (q>0.005)	1.2	1.4×10^{-1}
95	MYBBP1A	WT (q>0.005)	1.1	1.5×10^{-1}
96	CSNK2A3	WT (q>0.005)	1.1	1.6×10^{-1}
97	RBM39	WT (q>0.005)	1.1	1.7×10^{-1}
98	H3C13	WT (q>0.005)	1.1	1.9×10^{-1}
99	NIFK	WT (q>0.005)	1.1	2.0×10^{-1}
100	DDX56	WT (q>0.005)	1.1	2.8×10^{-1}
101	SPATS2L	WT (q>0.005)	1.0	3.5×10^{-1}
102	RPS15A	B (q>0.005)	1.4	5.5×10^{-3}
103	FAM98A	B (q>0.005)	1.3	6.8×10^{-3}
104	SAFB2	B (q>0.005)	1.3	6.9×10^{-3}
105	RTCB	B (q>0.005)	1.3	7.0×10^{-3}
106	REXO4	B (q>0.005)	1.2	1.4×10^{-2}

Number	Name	Highest mean condition	Fold change	t-test q value
107	TUBA1C	B (q>0.005)	1.2	1.9×10^{-2}
108	RTRAF	B (q>0.005)	1.2	3.5×10^{-2}
109	EEF1G	B (q>0.005)	1.2	3.6×10^{-2}
110	RPS9	B (q>0.005)	1.2	4.6×10^{-2}
111	GPATCH4	B (q>0.005)	1.2	5.3×10^{-2}
112	KNOP1	B (q>0.005)	1.1	5.5×10^{-2}
113	RPL17	B (q>0.005)	1.2	6.2×10^{-2}
114	RPS24	B (q>0.005)	1.2	9.0×10^{-2}
115	RPL13	B (q>0.005)	1.2	1.3×10^{-1}
116	RBM34	B (q>0.005)	1.3	1.3×10^{-1}
117	CBR1	B (q>0.005)	1.1	1.4×10^{-1}
118	NAT10	B (q>0.005)	1.1	1.6×10^{-1}
119	RPL10	B (q>0.005)	1.1	1.6×10^{-1}
120	SRSF10	B (q>0.005)	1.4	1.8×10^{-1}
121	GSTP1	B (q>0.005)	1.1	2.3×10^{-1}
122	MKI67	B (q>0.005)	1.1	2.3×10^{-1}
123	RPLP2	B (q>0.005)	1.0	2.7×10^{-1}
124	DIMT1	B (q>0.005)	1.0	2.9×10^{-1}
125	CSNK2A2	B (q>0.005)	1.1	2.9×10^{-1}
126	HSPA9	B (q>0.005)	1.0	2.9×10^{-1}
127	ACTB	B (q>0.005)	1.1	3.0×10^{-1}
128	RPL4	B (q>0.005)	1.4	3.0×10^{-1}
129	PRKDC	B (q>0.005)	1.0	3.0×10^{-1}
130	SAFB	B (q>0.005)	1.0	3.1×10^{-1}
131	TUBA4A	B (q>0.005)	1.1	3.1×10^{-1}
132	RPL13A	B (q>0.005)	1.0	3.3×10^{-1}
133	NUMA1	B (q>0.005)	1.0	3.4×10^{-1}

Abbreviations

AcD	Acidic domain
AD	Anchor domain
ANOVA	Analysis of variance
APC/C	Anaphase promoting complex/cyclosome
AURKB	Aurora kinase B
BRCA1	Breast cancer type 1 susceptibility protein
BSA	Bovine serum albumin
°C	Degrees centigrade
CDK	Cyclin dependent kinase
CIZ1	Cip1-interacing zinc finger protein 1
CSK	Cytoskeletal buffer
DAPI	4′,6-diamidino-2-phenylindole
DDR	DNA damage response
DIG	Digoxigenin
DLC1	Dynein light chain
DMEM	Dulbecco's Modified Eagle Medium
DN	Dominant negative
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
DSB	Double stranded break
DTT	Dithiothreitol
DUB	Deubiquitinase
eCIZ1	Embryonic CIZ1
EDTA	Ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
EMSA	Electrophoretic mobility shift assay
EMT	Epithelial mesenchymal transition
ESR1/ESRα	Oestrogen receptor alpha
FC	Fold change
G0	Quiescence
G1 phase	Growth 1 phase
G2 phase	Growth 2 phase
GFP	Green fluorescent protein
H1	Linker histone 1
H2A,2B,3,4	Core histone 2A, 2B, 3 and 4
H2AK119ub1	Ubiquitination of histone H2A at lysine 119
H3K27me3	Tri-methylation of histone H3 at lysine 27
H3S10	Histone 3 serine 10
H3S28	Histone 3 serine 28
HDAC	Histone deacetylase
HEKs	Human embryonic kidney cells
HMEC	Human mammary epithelial cells
hnRNP	Heterogeneous nuclear ribonucleoprotein
HP1	Heterochromatin protein 1

kDa	Kilodalton
LLPS	Liquid-liquid phase separation
lncRNA	Long non-coding RNAs
M phase	Mitosis
MARs	Matrix attachment regions
MH3	Matrin 3-homologous domain 3
MS	Mass spectrometry
NLS	Nuclear localisation sequence
p53	Tumour protein 53
PBS	Phosphate buffered saline
PEF	Primary embryonic fibroblast
PFA	Paraformaldehyde
PIPES	Piperazine-N,N'bis(2-ethanesulfonic acid)
PLD	Prion like domain
PMSF	Phenylmethanesulfonylfluoride
polyQ	Polyglutamine
PP1/2A	Protein phosphatase 1 and 2A
pRb	Retinoblastoma protein
PRC1/2	Polycomb repressive complex 1/2
PSG	Penicillin streptomycin glutamine
PTM	Post-translational modification
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RD	Replication domain
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
RNase	Ribonuclease
ROS	Reactive oxygen species
RRM	RNA recognition motif
rRNA	Ribosomal RNA
S phase	DNA synthesis phase
S/MARs	Collective term for SARs and MARs
SARs	Scaffold attachment regions
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALLs	Size exclusion chromatography multiple angle laser light scattering
SEM	Standard error of the mean
TAD	Transactivation domain
TCGA	The Cancer Genome Atlas
TP53	Tumour protein 53
TSS	Transcription start site
UT	Untransfected
WT	Wild type
Xa	Active X chromosome
Xi	Inactive X chromosome
XIC	X-inactivation centre
Xist	X-inactive specific transcript
ZF	Zinc finger

References

Afonso, O., Castellani, C. M., Cheeseman, L. P., Ferreira, J. G., Orr, B., Ferreira, L. T., Chambers, J. J., Morais-de-Sa, E., Maresca, T. J. & Maiato, H. 2019. Spatiotemporal control of mitotic exit during anaphase by an aurora B-Cdk1 crosstalk. *Elife,* 8.

Aihara, A., Tanaka, S., Yasen, M., Matsumura, S., Mitsunori, Y., Murakata, A., Noguchi, N., Kudo, A., Nakamura, N., Ito, K. & Arii, S. 2010. The selective Aurora B kinase inhibitor AZD1152 as a novel treatment for hepatocellular carcinoma. *J Hepatol,* 52**,** 63-71.

Ainscough, J. F., Rahman, F. A., Sercombe, H., Sedo, A., Gerlach, B. & Coverley, D. 2007. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. *J Cell Sci,* 120**,** 115-24.

Ajiro, K., Yoda, K., Utsumi, K. & Nishikawa, Y. 1996. Alteration of cell cycle-dependent histone phosphorylations by okadaic acid. Induction of mitosis-specific H3 phosphorylation and chromatin condensation in mammalian interphase cells. *J Biol Chem,* 271**,** 13197-201.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., Schermelleh, L., Nesterova, T. B., Koseki, H. & Brockdorff, N. 2017. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science,* 356**,** 1081-1084.

Almlof, T., Gustafsson, J. A. & Wright, A. P. 1997. Role of hydrophobic amino acid clusters in the transactivation activity of the human glucocorticoid receptor. *Mol Cell Biol,* 17**,** 934-45.

Aoto, T., Saitoh, N., Sakamoto, Y., Watanabe, S. & Nakao, M. 2008. Polycomb group protein-associated chromatin is reproduced in post-mitotic G1 phase and is required for S phase progression. *J Biol Chem,* 283**,** 18905-15.

Arakel, E. C. & Schwappach, B. 2018. Formation of COPI-coated vesicles at a glance. *J Cell Sci,* 131.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

Attardi, L. D. & Tjian, R. 1993. Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. *Genes Dev,* 7**,** 1341-53.

Azzariti, A., Bocci, G., Porcelli, L., Fioravanti, A., Sini, P., Simone, G. M., Quatrale, A. E., Chiarappa, P., Mangia, A., Sebastian, S., Del Bufalo, D., Del Tacca, M. & Paradiso, A. 2011. Aurora B kinase inhibitor AZD1152: determinants of action and ability to enhance chemotherapeutics effectiveness in pancreatic and colon cancer. *Br J Cancer,* 104**,** 769-80.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A. & Wishart, D. S. 2016. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res,* 44**,** W147-53.

Babkoff, A., Cohen-Kfir, E., Aharon, H. & Ravid, S. 2021. Aurora-B phosphorylates the myosin II heavy chain to promote cytokinesis. *J Biol Chem,* 297**,** 101024.

Bai, Z. T., Bai, B., Zhu, J., Di, C. X., Li, X. & Zhou, W. C. 2018. Epigenetic actions of environmental factors and promising drugs for cancer therapy. *Oncol Lett,* 15**,** 2049-2056.

Baldin, V., Lukas, J., Marcote, M. J., Pagano, M. & Draetta, G. 1993. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes Dev,* 7**,** 812-21.

Bao, H., Li, X., Cao, Z., Huang, Z., Chen, L., Wang, M., Hu, J., Li, W., Sun, H., Jiang, X., Mei, P., Li, H., Lu, L. & Zhan, M. 2022. Identification of COPA as a potential prognostic biomarker and pharmacological intervention target of cervical cancer by quantitative proteomics and experimental verification. *J Transl Med,* 20**,** 18.

Barnum, K. J. & O'Connell, M. J. 2014. Cell cycle regulation by checkpoints. *Methods Mol Biol,* 1170**,** 29-40.

Barr, M. L. & Bertram, E. G. 1949. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature,* 163**,** 676.

Barr, M. L. & Moore, K. L. 1957. Chromosomes, sex chromatin, and cancer. *Proc Can Cancer Conf,* 2**,** 3-16.

Baylin, S. B. & Jones, P. A. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer,* 11**,** 726-34.

Bayona-Feliu, A., Casas-Lamesa, A., Reina, O., Bernues, J. & Azorin, F. 2017. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat Commun,* 8**,** 283.

Belgrader, P., Dey, R. & Berezney, R. 1991. Molecular cloning of matrin 3. A 125-kilodalton protein of the nuclear matrix contains an extensive acidic domain. *Journal of Biological Chemistry,* 266**,** 9893-9899.

Benoît, M.-H., Hudson, T., Maire, G., Squire, J., Arcand, S., Provencher, D., Mes-Masson, A.-M. & Tonin, P. 2007. Global analysis of chromosome X gene expression in primary cultures of normal ovarian surface epithelial cells and epithelial ovarian cancer cell lines. *International Journal of Oncology*.

Berezney, R. 1980. Fractionation of the nuclear matrix. I. Partial separation into matrix protein fibrils and a residual ribonucleoprotein fraction. *J Cell Biol,* 85**,** 641-50.

Berezney, R. & Coffey, D. S. 1974. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun,* 60**,** 1410-7.

Berezney, R. & Coffey, D. S. 1976. The nuclear protein matrix: isolation, structure, and functions. *Adv Enzyme Regul,* 14**,** 63-100.

Berg, J. M. 1988. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci U S A,* 85**,** 99-102.

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. & Allis, C. D. 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol,* 26**,** 2560-9.

Bertran-Alamillo, J., Cattan, V., Schoumacher, M., Codony-Servat, J., Giménez-Capitán, A., Cantero, F., Burbridge, M., Rodríguez, S., Teixidó, C., Roman, R., Castellví, J., García-Román, S., Codony-Servat, C., Viteri, S., Cardona, A.-F., Karachaliou, N., Rosell, R. & Molina-Vila, M.-A. 2019. AURKB as a target in non-small cell lung cancer with acquired resistance to anti-EGFR therapy. *Nature Communications,* 10**,** 1812.

Bhopal, A., Peake, M. D., Gilligan, D. & Cosford, P. 2019. Lung cancer in never-smokers: a hidden disease. *J R Soc Med,* 112**,** 269-271.

Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R. & Jensen, L. J. 2014. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford),* 2014**,** bau012.

Bird, A. 2007. Perceptions of epigenetics. *Nature,* 447**,** 396-398.

Bloomfield, M. & Duesberg, P. 2015. Karyotype alteration generates the neoplastic phenotypes of SV40-infected human and rodent cells. *Mol Cytogenet,* 8**,** 79.

Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C. & Allis, C. D. 2002. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet,* 30**,** 73-6.

Bonekamp, N. A., Jiang, M., Motori, E., Garcia Villegas, R., Koolmeister, C., Atanassov, I., Mesaros, A., Park, C. B. & Larsson, N. G. 2021. High levels of TFAM repress mammalian mitochondrial DNA transcription in vivo. *Life Sci Alliance,* 4.

Borah, V., Shah, P. N., Ghosh, S. N., Sampat, M. B. & Jussawalla, D. J. 1980. Further studies on the prognostic importance of Barr body frequency in human breast cancer: with discussion on its probable mechanism. *J Surg Oncol,* 13**,** 1-7.

Boulikas, T. 1993. Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J Cell Biochem,* 52**,** 14-22.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. & Jaenisch, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature,* 441**,** 349-53.

Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol,* 34**,** 525-7.

Brayer, K. J. & Segal, D. J. 2008. Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem Biophys,* 50**,** 111-31.

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D. & Rastan, S. 1991. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature,* 351**,** 329-31.

Brower, V. 2011. Epigenetics: Unravelling the cancer code. *Nature,* 471**,** S12-3.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. & Willard, H. F. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature,* 349**,** 38-44.

Brown, C. J. & Willard, H. F. 1994. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature,* 368**,** 154-6.

Bruhn, C., Bastianello, G. & Foiani, M. 2022. Cancer cell histone density links global histone acetylation, mitochondrial proteome and histone acetylase inhibitor sensitivity. *Commun Biol,* 5**,** 882.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. & Zhang, Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science,* 298**,** 1039-43.

Cao, R. & Zhang, Y. 2004. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev,* 14**,** 155-64.

Carmena, M. & Earnshaw, W. C. 2003. The cellular geography of aurora kinases. *Nat Rev Mol Cell Biol,* 4**,** 842-54.

Carvalho, J. 2020. Cell Reversal From a Differentiated to a Stem-Like State at Cancer Initiation. *Front Oncol,* 10**,** 541.

Cassola, A., Noe, G. & Frasch, A. C. 2010. RNA recognition motifs involved in nuclear import of RNA-binding proteins. *RNA Biol,* 7**,** 339-44.

Cha, H. J., Uyan, O., Kai, Y., Liu, T., Zhu, Q., Tothova, Z., Botten, G. A., Xu, J., Yuan, G. C., Dekker, J. & Orkin, S. H. 2021. Inner nuclear protein Matrin-3 coordinates cell differentiation by stabilizing chromatin architecture. *Nat Commun,* 12**,** 6241.

Chadwick, B. P. & Willard, H. F. 2003. Barring gene expression after XIST: maintaining facultative heterochromatin on the inactive X. *Semin Cell Dev Biol,* 14**,** 359-67.

Chaligné, R., Popova, T., Mendoza-Parra, M. A., Saleem, M. A., Gentien, D., Ban, K., Piolot, T., Leroy, O., Mariani, O., Gronemeyer, H., Vincent-Salomon, A., Stern, M. H. & Heard, E. 2015. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Res,* 25**,** 488-503.

Chan, H. L., Beckedorff, F., Zhang, Y., Garcia-Huidobro, J., Jiang, H., Colaprico, A., Bilbao, D., Figueroa, M. E., LaCava, J., Shiekhattar, R. & Morey, L. 2018. Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nat Commun,* 9**,** 3377.

Chelsky, D., Ralph, R. & Jonak, G. 1989. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol Cell Biol,* 9**,** 2487-92.

Chen, A. 2011. PARP inhibitors: its role in treatment of cancer. *Chin J Cancer,* 30**,** 463-71.

Chen, C. K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P. & Guttman, M. 2016. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science,* 354**,** 468-472.

Chenevix-Trench, G., Spurdle, A. B., Gatei, M., Kelly, H., Marsh, A., Chen, X., Donn, K., Cummings, M., Nyholt, D., Jenkins, M. A., Scott, C., Pupo, G. M., Dork, T., Bendix, R., Kirk, J., Tucker, K., McCredie, M. R., Hopper, J. L., Sambrook, J., Mann, G. J. & Khanna, K. K. 2002. Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst,* 94**,** 205-15.

Cheng, F., Liu, C., Lin, C. C., Zhao, J., Jia, P., Li, W. H. & Zhao, Z. 2015. A Gene Gravity Model for the Evolution of Cancer Genomes: A Study of 3,000 Cancer Genomes across 9 Cancer Types. *PLoS Comput Biol,* 11**,** e1004497.

Chow, V. T. & Quek, H. H. 1997. Alpha coat protein COPA (HEP-COP): presence of an Alu repeat in cDNA and identity of the amino terminus to xenin. *Ann Hum Genet,* 61**,** 369-73.

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. & Chang, H. Y. 2015. Systematic discovery of Xist RNA binding proteins. *Cell,* 161**,** 404-16.

Ciejek, E. M., Tsai, M. J. & O'Malley, B. W. 1983. Actively transcribed genes are associated with the nuclear matrix. *Nature,* 306**,** 607-9.

Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol,* 132**,** 259-75.

Clute, P. & Pines, J. 1999. Temporal and spatial control of cyclin B1 destruction in metaphase. *Nat Cell Biol,* 1**,** 82-7.

Cockerill, P. N. & Garrard, W. T. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell,* 44**,** 273-82.

Copeland, N. A., Sercombe, H. E., Ainscough, J. F. & Coverley, D. 2010. Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. *J Cell Sci,* 123**,** 1108-15.

Copeland, N. A., Sercombe, H. E., Wilson, R. H. & Coverley, D. 2015. Cyclin-A-CDK2-mediated phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. *J Cell Sci,* 128**,** 1518-27.

Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. 1999. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science,* 286**,** 1162-6.

Costanzi, C. & Pehrson, J. R. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature,* 393**,** 599-601.

Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell,* 59**,** 827-836.

Coverley, D., Higgins, G., West, D., Jackson, O. T., Dowle, A., Haslam, A., Ainscough, E., Chalkley, R. & White, J. 2017. A quantitative immunoassay for lung cancer biomarker CIZ1b in patient plasma. *Clin Biochem,* 50**,** 336-343.

Coverley, D., Laman, H. & Laskey, R. A. 2002. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. *Nat Cell Biol,* 4**,** 523-8.

Coverley, D., Marr, J. & Ainscough, J. 2005. Ciz1 promotes mammalian DNA replication. *J Cell Sci,* 118**,** 101-12.

CRUK. 2021. *Lung cancer incidence statistics* [Online]. Cancer Research UK. Available: https://www.cancerresearchuk.org/health-professional/cancer-

statistics/statistics-by-cancer-type/lung-cancer/incidence#heading-Zero [Accessed 22nd December 2022].

CRUK. 2022. *Lung cancer mortality statistics* [Online]. Cancer Research UK. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/mortality#heading-Zero [Accessed 22nd December 2022].

Csankovszki, G., Nagy, A. & Jaenisch, R. 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol,* 153**,** 773-84.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. & Jaenisch, R. 1999. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet,* 22**,** 323-4.

da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A. F., Wutz, A., Margueron, R. & Heard, E. 2014. Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell,* 53**,** 301-16.

Dahmcke, C. M., Buchmann-Moller, S., Jensen, N. A. & Mitchelmore, C. 2008. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol Cell Neurosci,* 38**,** 589-94.

Das, C. & Tyler, J. K. 2012. Histone exchange and histone modifications during transcription and aging. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1819**,** 332-342.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H. & Brockdorff, N. 2004. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell,* 7**,** 663-76.

Dean, P. N. & Jett, J. H. 1974. Mathematical analysis of DNA distributions derived from flow microfluorometry. *J Cell Biol,* 60**,** 523-7.

DeGregori, J., Kowalik, T. & Nevins, J. R. 1995. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes. *Mol Cell Biol,* 15**,** 4215-24.

den Elzen, N. & Pines, J. 2001. Cyclin A is destroyed in prometaphase and can delay chromosome alignment and anaphase. *J Cell Biol,* 153**,** 121-36.

den Hollander, P. & Kumar, R. 2006. Dynein light chain 1 contributes to cell cycle progression by increasing cyclin-dependent kinase 2 activity in estrogen-stimulated cells. *Cancer Res,* 66**,** 5941-9.

den Hollander, P., Rayala, S. K., Coverley, D. & Kumar, R. 2006. Ciz1, a Novel DNA-binding coactivator of the estrogen receptor alpha, confers hypersensitivity to estrogen action. *Cancer Res,* 66**,** 11021-9.

Dixon-McDougall, T. & Brown, C. J. 2022. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics Chromatin,* 15**,** 6.

Dobrinic, P., Szczurek, A. T. & Klose, R. J. 2021. PRC1 drives Polycomb-mediated gene repression by controlling transcription initiation and burst frequency. *Nat Struct Mol Biol,* 28**,** 811-824.

Dobson, J. R., Hong, D., Barutcu, A. R., Wu, H., Imbalzano, A. N., Lian, J. B., Stein, J. L., van Wijnen, A. J., Nickerson, J. A. & Stein, G. S. 2017. Identifying Nuclear Matrix-Attached DNA Across the Genome. *J Cell Physiol,* 232**,** 1295-1305.

Dong, P., Tada, M., Hamada, J., Nakamura, A., Moriuchi, T. & Sakuragi, N. 2007. p53 dominant-negative mutant R273H promotes invasion and migration of human endometrial cancer HHUA cells. *Clin Exp Metastasis,* 24**,** 471-83.

Downs, B. & Wang, S. M. 2015. Epigenetic changes in BRCA1-mutated familial breast cancer. *Cancer Genet,* 208**,** 237-40.

Drysdale, C. M., Duenas, E., Jackson, B. M., Reusser, U., Braus, G. H. & Hinnebusch, A. G. 1995. The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol Cell Biol,* 15**,** 1220-33.

Emanuele, M. J., Lan, W., Jwa, M., Miller, S. A., Chan, C. S. & Stukenberg, P. T. 2008. Aurora B kinase and protein phosphatase 1 have opposing roles in modulating kinetochore assembly. *J Cell Biol,* 181**,** 241-54.

Engelke, R., Riede, J., Hegermann, J., Wuerch, A., Eimer, S., Dengjel, J. & Mittler, G. 2014. The quantitative nuclear matrix proteome as a biochemical snapshot of nuclear organization. *J Proteome Res,* 13**,** 3940-56.

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E. S., Plath, K. & Guttman, M. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science,* 341**,** 1237973.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. & Bickmore, W. A. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell,* 38**,** 452-64.

Evans, R. P., Naber, C., Steffler, T., Checkland, T., Maxwell, C. A., Keats, J. J., Belch, A. R., Pilarski, L. M., Lai, R. & Reiman, T. 2008. The selective Aurora B kinase inhibitor AZD1152 is a potential new treatment for multiple myeloma. *Br J Haematol,* 140**,** 295-302.

Fackelmayer, F. O., Dahm, K., Renz, A., Ramsperger, U. & Richter, A. 1994. Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur J Biochem,* 221**,** 749-57.

Fan, H., Lv, P., Huo, X., Wu, J., Wang, Q., Cheng, L., Liu, Y., Tang, Q. Q., Zhang, L., Zhang, F., Zheng, X., Wu, H. & Wen, B. 2018. The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes. *Genome Res,* 28**,** 192-202.

Feurle, G. E., Hamscher, G., Kusiek, R., Meyer, H. E. & Metzger, J. W. 1992. Identification of xenin, a xenopsin-related peptide, in the human gastric mucosa and its effect on exocrine pancreatic secretion. *Journal of Biological Chemistry,* 267**,** 22305-22309.

Flemming, W. 1879. Beitrage zur Kenntniss der Zelle und ihrer Lebenserscheinungen. *Archiv für Mikroskopische Anatomie,* 16**,** 302-436.

Fourriere, L., Jimenez, A. J., Perez, F. & Boncompain, G. 2020. The role of microtubules in secretory protein transport. *J Cell Sci,* 133.

Ganesan, S., Silver, D. P., Greenberg, R. A., Avni, D., Drapkin, R., Miron, A., Mok, S. C., Randrianarison, V., Brodie, S., Salstrom, J., Rasmussen, T. P., Klimke, A., Marrese, C., Marahrens, Y., Deng, C.-X., Feunteun, J. & Livingston, D. M. 2002. BRCA1 Supports XIST RNA Concentration on the Inactive X Chromosome. *Cell,* 111**,** 393-405.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. & Reinberg, D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell,* 45**,** 344-56.

Gaynor, E. C. & Emr, S. D. 1997. COPI-independent anterograde transport: cargo-selective ER to Golgi protein transport in yeast COPI mutants. *J Cell Biol,* 136**,** 789-802.

Geeven, G., Zhu, Y., Kim, B. J., Bartholdy, B. A., Yang, S. M., Macfarlan, T. S., Gifford, W. D., Pfaff, S. L., Verstegen, M. J., Pinto, H., Vermunt, M. W., Creyghton, M. P., Wijchers, P. J., Stamatoyannopoulos, J. A., Skoultchi, A. I. & de Laat, W. 2015. Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol,* 16**,** 289.

Geng, Y., Yu, Q., Sicinska, E., Das, M., Schneider, J. E., Bhattacharya, S., Rideout, W. M., Bronson, R. T., Gardner, H. & Sicinski, P. 2003. Cyclin E ablation in the mouse. *Cell,* 114**,** 431-43.

Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S. & Schaffner, W. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science,* 263**,** 808-11.

Ghanizadeh-Vesali, S., Zekri, A., Zaker, F., Zaghal, A., Yousefi, M., Alimoghaddam, K., Ghavamzadeh, A. & Ghaffari, S. H. 2016. Significance of AZD1152 as a potential treatment against Aurora B overexpression in acute promyelocytic leukemia. *Ann Hematol,* 95**,** 1031-42.

Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S. & Rosen, M. K. 2019. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell,* 179**,** 470-484 e21.

Goldenson, B. & Crispino, J. D. 2015. The aurora kinases in cell cycle and leukemia. *Oncogene,* 34**,** 537-45.

Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D. H., Niall, H. D. & Boyse, E. A. 1975. Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Natl Acad Sci U S A,* 72**,** 11-5.

Goto, H., Yasui, Y., Nigg, E. A. & Inagaki, M. 2002. Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes Cells,* 7**,** 11-7.

GTEx 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet,* 45**,** 580-5.

Guo, Y., Zhao, S. & Wang, G. G. 2021. Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 'Readout', and Phase Separation-Based Compaction. *Trends Genet,* 37**,** 547-565.

Hall, L. L., Byron, M., Pageau, G. & Lawrence, J. B. 2009. AURKB-mediated effects on chromatin regulate binding versus release of XIST RNA to the inactive chromosome. *J Cell Biol,* 186**,** 491-507.

Hall, T. M. 2005. Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol,* 15**,** 367-73.

Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell,* 75**,** 805-16.

Harper, J. W., Elledge, S. J., Keyomarsi, K., Dynlacht, B., Tsai, L. H., Zhang, P., Dobrowolski, S., Bai, C., Connell-Crowley, L., Swindell, E. & et al. 1995. Inhibition of cyclin-dependent kinases by p21. *Mol Biol Cell,* 6**,** 387-400.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K. & Nakagawa, S. 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell,* 19**,** 469-76.

Hashimoto, H., Takami, Y., Sonoda, E., Iwasaki, T., Iwano, H., Tachibana, M., Takeda, S., Nakayama, T., Kimura, H. & Shinkai, Y. 2010. Histone H1 null vertebrate cells exhibit altered nucleosome architecture. *Nucleic Acids Res,* 38**,** 3533-45.

Hauf, S., Cole, R. W., LaTerra, S., Zimmer, C., Schnapp, G., Walter, R., Heckel, A., van Meel, J., Rieder, C. L. & Peters, J. M. 2003. The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J Cell Biol,* 161**,** 281-94.

He, D., Zeng, C. & Brinkley, B. R. 1995. Nuclear matrix proteins as structural and functional components of the mitotic apparatus. *Int Rev Cytol,* 162B**,** 1-74.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D. & Spector, D. L. 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell,* 107**,** 727-38.

Helbig, R. & Fackelmayer, F. O. 2003. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma,* 112**,** 173-82.

Helfrich, B. A., Kim, J., Gao, D., Chan, D. C., Zhang, Z., Tan, A. C. & Bunn, P. A., Jr. 2016. Barasertib (AZD1152), a Small Molecule Aurora B Inhibitor, Inhibits the Growth of SCLC Cell Lines In Vitro and In Vivo. *Mol Cancer Ther,* 15**,** 2314-2322.

Hergeth, S. P. & Schneider, R. 2015. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep,* 16**,** 1439-53.

Herskowitz, I. 1987. Functional inactivation of genes by dominant negative mutations. *Nature,* 329**,** 219-22.

Higgins, G., Roper, K. M., Watson, I. J., Blackhall, F. H., Rom, W. N., Pass, H. I., Ainscough, J. F. & Coverley, D. 2012. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc Natl Acad Sci U S A,* 109**,** E3128-35.

Hinds, P. W., Mittnacht, S., Dulic, V., Arnold, A., Reed, S. I. & Weinberg, R. A. 1992. Regulation of retinoblastoma protein functions by ectopic expression of human cyclins. *Cell,* 70**,** 993-1006.

Hirota, T., Lipp, J. J., Toh, B. H. & Peters, J. M. 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature,* 438**,** 1176-80.

Hyun, K., Jeon, J., Park, K. & Kim, J. 2017. Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine,* 49**,** e324-e324.

Ishii, K., Hirano, Y., Araki, N., Oda, T., Kumeta, M., Takeyasu, K., Furukawa, K. & Horigome, T. 2008. Nuclear matrix contains novel WD-repeat and disordered-region-rich proteins. *FEBS Lett,* 582**,** 3515-9.

Iuchi, S. 2001. Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci,* 58**,** 625-35.

Jackson, S. P. & Bartek, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-1078.

Jager, N., Schlesner, M., Jones, D. T., Raffel, S., Mallm, J. P., Junge, K. M., Weichenhan, D., Bauer, T., Ishaque, N., Kool, M., Northcott, P. A., Korshunov, A., Drews, R. M., Koster, J., Versteeg, R., Richter, J., Hummel, M., Mack, S. C., Taylor, M. D., Witt, H., Swartman, B., Schulte-Bockholt, D., Sultan, M., Yaspo, M. L., Lehrach, H., Hutter, B., Brors, B., Wolf, S., Plass, C., Siebert, R., Trumpp, A., Rippe, K., Lehmann, I., Lichter, P., Pfister, S. M. & Eils, R. 2013. Hypermutation of the inactive X chromosome is a frequent event in cancer. *Cell,* 155**,** 567-81.

Jeffery, H. M. & Weinzierl, R. O. J. 2020. Multivalent and Bidirectional Binding of Transcriptional Transactivation Domains to the MED25 Coactivator. *Biomolecules,* 10.

Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J. & Pavletich, N. P. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature,* 376**,** 313-20.

Jeon, Y. & Lee, J. T. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell,* 146**,** 119-33.

Jones, D. T. & Cozzetto, D. 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics,* 31**,** 857-63.

Jones, P. A. & Baylin, S. B. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.

Kalashnikova, A. A., Rogge, R. A. & Hansen, J. C. 2016. Linker histone H1 and protein-protein interactions. *Biochim Biophys Acta,* 1859**,** 455-61.

Kalashnikova, A. A., Winkler, D. D., McBryant, S. J., Henderson, R. K., Herman, J. A., DeLuca, J. G., Luger, K., Prenni, J. E. & Hansen, J. C. 2013. Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res,* 41**,** 4026-35.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J. & Ding, L. 2013. Mutational landscape and significance across 12 major cancer types. *Nature,* 502**,** 333-339.

Kang, J., Lee, H. J., Jun, S. Y., Park, E. S. & Maeng, L. S. 2015a. Cancer-Testis Antigen Expression in Serous Endometrial Cancer with Loss of X Chromosome Inactivation. *PLoS One,* 10**,** e0137476.

Kang, J., Lee, H. J., Kim, J., Lee, J. J. & Maeng, L. S. 2015b. Dysregulation of X chromosome inactivation in high grade ovarian serous adenocarcinoma. *PLoS One,* 10**,** e0118927.

Kato, J., Matsushime, H., Hiebert, S. W., Ewen, M. E. & Sherr, C. J. 1993. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev,* 7**,** 331-42.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

Kettenbach, A. N., Schweppe, D. K., Faherty, B. K., Pechenick, D., Pletnev, A. A. & Gerber, S. A. 2011. Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. *Sci Signal,* 4**,** rs5.

Khan, M. M., Xiao, J., Patel, D. & LeDoux, M. S. 2018. DNA damage and neurodegenerative phenotypes in aged Ciz1 null mice. *Neurobiol Aging,* 62**,** 180-190.

Khanuja, P. S., Lehr, J. E., Soule, H. D., Gehani, S. K., Noto, A. C., Choudhury, S., Chen, R. & Pienta, K. J. 1993. Nuclear matrix proteins in normal and breast cancer cells. *Cancer Res,* 53**,** 3394-8.

Kim, A. J., Jee, H. J., Song, N., Kim, M., Jeong, S. Y. & Yun, J. 2013. p21(WAF(1)/C(1)P(1)) deficiency induces mitochondrial dysfunction in HCT116 colon cancer cells. *Biochem Biophys Res Commun,* 430**,** 653-8.

Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol,* 37**,** 907-915.

Kim, S., Leong, A., Kim, M. & Yang, H. W. 2022. CDK4/6 initiates Rb inactivation and CDK2 activity coordinates cell-cycle commitment and G1/S transition. *Sci Rep,* 12**,** 16810.

Kimura, M., Uchida, C., Takano, Y., Kitagawa, M. & Okano, Y. 2004. Cell cycle-dependent regulation of the human aurora B promoter. *Biochem Biophys Res Commun,* 316**,** 930-6.

Kitagawa, M., Higashi, H., Jung, H. K., Suzuki-Takahashi, I., Ikeda, M., Tamai, K., Kato, J., Segawa, K., Yoshida, E., Nishimura, S. & Taya, Y. 1996. The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2. *The EMBO Journal,* 15**,** 7060-7069.

Klug, A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys,* 43**,** 1-21.

Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J. W., Elledge, S., Nishimoto, T., Morgan, D. O., Franza, B. R. & Roberts, J. M. 1992. Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle. *Science,* 257**,** 1689-94.

Kolpa, H. J., Creamer, K. M., Hall, L. L. & Lawrence, J. B. 2022. SAF-A mutants disrupt chromatin structure through dominant negative effects on RNAs associated with chromatin. *Mamm Genome,* 33**,** 366-381.

Kolpa, H. J., Fackelmayer, F. O. & Lawrence, J. B. 2016. SAF-A Requirement in Anchoring XIST RNA to Chromatin Varies in Transformed and Primary Cells. *Dev Cell,* 39**,** 9-10.

Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A,* 106**,** 10171-6.

Kotoglou, P., Kalaitzakis, A., Vezyraki, P., Tzavaras, T., Michalis, L. K., Dantzer, F., Jung, J. U. & Angelidis, C. 2009. Hsp70 translocates to the nuclei and nucleoli, binds to XRCC1 and PARP-1, and protects HeLa cells from single-strand DNA breaks. *Cell Stress Chaperones,* 14**,** 391-406.

Kozar, K., Ciemerych, M. A., Rebel, V. I., Shigematsu, H., Zagozdzon, A., Sicinska, E., Geng, Y., Yu, Q., Bhattacharya, S., Bronson, R. T., Akashi, K. & Sicinski, P. 2004. Mouse development and cell proliferation in the absence of D-cyclins. *Cell,* 118**,** 477-91.

Kucherlapati, M. 2018. Examining transcriptional changes to DNA replication and repair factors over uveal melanoma subtypes. *BMC Cancer,* 18**,** 818.

Kunkel, G. H., Chaturvedi, P., Thelian, N., Nair, R. & Tyagi, S. C. 2018. Mechanisms of TFAM-mediated cardiomyocyte protection. *Can J Physiol Pharmacol,* 96**,** 173-181.

Kuriyama, H., Fukushima, S., Kimura, T., Okada, E., Ishibashi, T., Mizuhashi, S., Kanemaru, H., Kajihara, I., Makino, K., Miyashita, A., Aoi, J., Okada, S., Ihn, H. & Kita, K. 2020. Matrin-3 plays an important role in cell cycle and apoptosis for survival in malignant melanoma. *J Dermatol Sci,* 100**,** 110-119.

Laity, J. H., Lee, B. M. & Wright, P. E. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol,* 11**,** 39-46.

Lane, D. P. 1992. Cancer. p53, guardian of the genome. *Nature,* 358**,** 15-6.

Lee, E. J., Kang, Y. C., Park, W. H., Jeong, J. H. & Pak, Y. K. 2014. Negative transcriptional regulation of mitochondrial transcription factor A (TFAM) by nuclear TFAM. *Biochem Biophys Res Commun,* 450**,** 166-71.

Lee, J. T. & Jaenisch, R. 1997. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature,* 386**,** 275-9.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R. & Young, R. A. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell,* 125**,** 301-13.

Lei, L., Wu, J., Gu, D., Liu, H. & Wang, S. 2016. CIZ1 interacts with YAP and activates its transcriptional activity in hepatocellular carcinoma cells. *Tumour Biol,* 37**,** 11073-9.

Lever, E. & Sheer, D. 2010. The role of nuclear organization in cancer. *J Pathol,* 220**,** 114-25.

Lewis, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature,* 276**,** 565-570.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

Lindstrom, M. S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J. & Volarevic, S. 2018. Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. *Oncogene,* 37**,** 2351-2366.

Linnemann, A. K., Platts, A. E. & Krawetz, S. A. 2009. Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum Mol Genet,* 18**,** 645-54.

Liu, T., Ren, X., Li, L., Yin, L., Liang, K., Yu, H., Ren, H., Zhou, W., Jing, H. & Kong, C. 2015a. Ciz1 promotes tumorigenicity of prostate carcinoma cells. *Front Biosci (Landmark Ed),* 20**,** 705-15.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y. & Ren, J. 2015b. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics,* 31**,** 3359-61.

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., Mungall, C. J., Arner, E., Baillie, J. K., Bertin, N., Bono, H., de Hoon, M., Diehl, A. D., Dimont, E., Freeman, T. C., Fujieda, K., Hide, W., Kaliyaperumal, R., Katayama, T., Lassmann, T., Meehan, T. F., Nishikata, K., Ono, H., Rehli, M., Sandelin, A., Schultes, E. A., t Hoen, P. A., Tatum, Z., Thompson, M., Toyoda, T., Wright, D. W., Daub, C. O., Itoh, M., Carninci, P., Hayashizaki, Y., Forrest, A. R., Kawaji, H. & consortium, F. 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol,* 16**,** 22.

Loda, A., Collombet, S. & Heard, E. 2022. Gene regulation in time and space during X-chromosome inactivation. *Nat Rev Mol Cell Biol,* 23**,** 231-249.

Lose, F., Duffy, D. L., Kay, G. F., Kedda, M. A., Spurdle, A. B., Kathleen Cuningham Foundation Consortium for Research into Familial Breast, C. & Australian Ovarian Cancer Study Management, G. 2008. Skewed X chromosome

inactivation and breast and ovarian cancer status: evidence for X-linked modifiers of BRCA1. *J Natl Cancer Inst,* 100**,** 1519-29.

Lukasik, A., Uniewicz, K. A., Kulis, M. & Kozlowski, P. 2008. Ciz1, a p21 cip1/Waf1-interacting zinc finger protein and DNA replication factor, is a novel molecular partner for human enhancer of rudimentary homolog. *FEBS J,* 275**,** 332-40.

Lyon, M. F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature,* 190**,** 372-3.

Malik, A. M. & Barmada, S. J. 2021. Matrin 3 in neuromuscular disease: physiology and pathophysiology. *JCI Insight,* 6.

Malumbres, M. 2014. Cyclin-dependent kinases. *Genome Biol,* 15**,** 122.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev,* 11**,** 156-66.

Marescal, O. & Cheeseman, I. M. 2020. Cellular Mechanisms and Regulation of Quiescence. *Dev Cell,* 55**,** 259-271.

Markaki, Y., Gan Chong, J., Wang, Y., Jacobson, E. C., Luong, C., Tan, S. Y. X., Jachowicz, J. W., Strehle, M., Maestrini, D., Banerjee, A. K., Mistry, B. A., Dror, I., Dossin, F., Schoneberg, J., Heard, E., Guttman, M., Chou, T. & Plath, K. 2021. Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell,* 184**,** 6174-6192 e32.

Mattern, K. A., Humbel, B. M., Muijsers, A. O., Jong, L. d. & Driel, R. v. 1996. hnRNP proteins and B23 are the major proteins of the internal nuclear matrix of HeLa S3 cells. *Journal of Cellular Biochemistry,* 62**,** 275-289.

Mayor, R., Izquierdo-Bouldstridge, A., Millan-Arino, L., Bustillos, A., Sampaio, C., Luque, N. & Jordan, A. 2015. Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. *J Biol Chem,* 290**,** 7474-91.

McBryant, S. J., Lu, X. & Hansen, J. C. 2010. Multifunctionality of the linker histones: an emerging role for protein-protein interactions. *Cell Res,* 20**,** 519-28.

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K. & Guttman, M. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature,* 521**,** 232-6.

Melen, K., Kinnunen, L. & Julkunen, I. 2001. Arginine/lysine-rich structural element is involved in interferon-induced nuclear import of STATs. *J Biol Chem,* 276**,** 16447-55.

Mendiratta, G., Ke, E., Aziz, M., Liarakos, D., Tong, M. & Stites, E. C. 2021. Cancer gene mutation frequencies for the U.S. population. *Nat Commun,* 12**,** 5961.

Mermoud, J. E., Popova, B., Peters, A. H., Jenuwein, T. & Brockdorff, N. 2002. Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. *Curr Biol,* 12**,** 247-51.

Mi, H. & Thomas, P. 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol,* 563**,** 123-40.

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W. & Lee, J. T. 2015. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science,* 349.

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods,* 19**,** 679-682.

Mirkovitch, J., Mirault, M. E. & Laemmli, U. K. 1984. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell,* 39**,** 223-32.

Mitsui, K., Matsumoto, A., Ohtsuka, S., Ohtsubo, M. & Yoshimura, A. 1999. Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. *Biochem Biophys Res Commun,* 264**,** 457-64.

Moen, E. L., Litwin, E., Arnovitz, S., Zhang, X., Zhang, W., Dolan, M. E. & Godley, L. A. 2015. Characterization of CpG sites that escape methylation on the inactive human X-chromosome. *Epigenetics,* 10**,** 810-818.

Moindrot, B. & Brockdorff, N. 2016. RNA binding proteins implicated in Xist-mediated chromosome silencing. *Semin Cell Dev Biol,* 56**,** 58-70.

Monk, M. & Harper, M. I. 1979. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature,* 281**,** 311-3.

Mori, N., Ishikawa, C., Senba, M., Kimura, M. & Okano, Y. 2011. Effects of AZD1152, a selective Aurora B kinase inhibitor, on Burkitt's and Hodgkin's lymphomas. *Biochem Pharmacol,* 81**,** 1106-15.

Munkley, J., Copeland, N. A., Moignard, V., Knight, J. R., Greaves, E., Ramsbottom, S. A., Pownall, M. E., Southgate, J., Ainscough, J. F. & Coverley, D. 2011. Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells. *Nucleic Acids Res,* 39**,** 2671-7.

Murata-Hori, M., Tatsuka, M. & Wang, Y. L. 2002. Probing the dynamics and functions of aurora B kinase in living cells during mitosis and cytokinesis. *Mol Biol Cell,* 13**,** 1099-108.

Murata-Hori, M. & Wang, Y. L. 2002. The kinase activity of aurora B is required for kinetochore-microtubule interactions during mitosis. *Curr Biol,* 12**,** 894-9.

Murrell, A., Rakyan, V. K. & Beck, S. 2005. From genome to epigenome. *Hum Mol Genet,* 14 Spec No 1**,** R3-R10.

Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D. & Nagai, K. 2004. The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol,* 341**,** 185-98.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M. & Ito, T. 2008. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev,* 22**,** 37-49.

Nakai, K. & Horton, P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences,* 24**,** 34-35.

Nakayasu, H. & Berezney, R. 1991. Nuclear matrins: identification of the major nuclear matrix proteins. *Proc Natl Acad Sci U S A,* 88**,** 10312-6.

Nanney, D. L. 1958. Epigenetic Control Systems. *Proc Natl Acad Sci U S A,* 44**,** 712-7.

National Lung Screening Trial Research, T., Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M. & Sicks, J. D. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med,* 365**,** 395-409.

Nguyen Ba, A. N., Pogoutse, A., Provart, N. & Moses, A. M. 2009. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics,* 10**,** 202.

NHS. 2021. *Cancer Registration Statistics, England 2019* [Online]. NHS Digital. Available: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2019 [Accessed 22nd December 2022].

NHS. 2022. *Cancer Survival in England, cancers diagnosed 2015 to 2019, followed up to 2020* [Online]. NHS Digital. Available: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-survival-in-england/cancers-diagnosed-2015-to-2019-followed-up-to-2020#resources [Accessed 22nd December 2022].

Nicholson, J. M. & Cimini, D. 2013. Cancer karyotypes: survival of the fittest. *Front Oncol,* 3**,** 148.

Nigg, E. A. 2001. Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol,* 2**,** 21-32.

Nishibe, R., Watanabe, W., Ueda, T., Yamasaki, N., Koller, R., Wolff, L., Honda, Z., Ohtsubo, M. & Honda, H. 2013. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. *FEBS Lett,* 587**,** 1529-35.

Oakes, V., Wang, W., Harrington, B., Lee, W. J., Beamish, H., Chia, K. M., Pinder, A., Goto, H., Inagaki, M., Pavey, S. & Gabrielli, B. 2014. Cyclin A/Cdk2 regulates Cdh1 and claspin during late S/G2 phase of the cell cycle. *Cell Cycle,* 13**,** 3302-11.

Obaya, A. J. & Sedivy, J. M. 2002. Regulation of cyclin-Cdk activity in mammalian cells. *Cell Mol Life Sci,* 59**,** 126-42.

Ohno, S., Kaplan, W. D. & Kinosita, R. 1959. Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. *Experimental Cell Research,* 18**,** 415-418.

Olins, A. L. & Olins, D. E. 1974. Spheroid chromatin units (v bodies). *Science,* 183**,** 330-2.

Orci, L., Ravazzola, M., Volchuk, A., Engel, T., Gmachl, M., Amherdt, M., Perrelet, A., Sollner, T. H. & Rothman, J. E. 2000. Anterograde flow of cargo across the golgi stack potentially mediated via bidirectional "percolating" COPI vesicles. *Proc Natl Acad Sci U S A,* 97**,** 10400-5.

Orci, L., Stamnes, M., Ravazzola, M., Amherdt, M., Perrelet, A., Söllner, T. H. & Rothman, J. E. 1997. Bidirectional Transport by Distinct Populations of COPI-Coated Vesicles. *Cell,* 90**,** 335-349.

Pagano, M., Pepperkok, R., Verde, F., Ansorge, W. & Draetta, G. 1992. Cyclin A is required at two points in the human cell cycle. *EMBO J,* 11**,** 961-71.

Pageau, G. J., Hall, L. L. & Lawrence, J. B. 2007. BRCA1 does not paint the inactive X to localize XIST RNA but may contribute to broad changes in cancer that impact XIST and Xi heterochromatin. *J Cell Biochem,* 100**,** 835-50.

Painter, R. C., Osmond, C., Gluckman, P., Hanson, M., Phillips, D. I. & Roseboom, T. J. 2008. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *BJOG,* 115**,** 1243-9.

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W. R., Damianov, A., Chronis, C., Papp, B., Chen, C. K., McKee, R., Wang, X. J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D. L. & Plath, K. 2020. A protein assembly mediates Xist localization and gene silencing. *Nature,* 587**,** 145-151.

Papa, A., Wan, L., Bonora, M., Salmena, L., Song, M. S., Hobbs, R. M., Lunardi, A., Webster, K., Ng, C., Newton, R. H., Knoblauch, N., Guarnerio, J., Ito, K., Turka, L. A., Beck, A. H., Pinton, P., Bronson, R. T., Wei, W. & Pandolfi, P. P. 2014. Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell,* 157**,** 595-610.

Papini, D., Levasseur, M. D. & Higgins, J. M. G. 2021. The Aurora B gradient sustains kinetochore stability in anaphase. *Cell Rep,* 37**,** 109818.

Park, S. Y., Yang, J. S., Schmider, A. B., Soberman, R. J. & Hsu, V. W. 2015. Coordinated regulation of bidirectional COPI transport at the Golgi by CDC42. *Nature,* 521**,** 529-32.

Parreno, V., Martinez, A. M. & Cavalli, G. 2022. Mechanisms of Polycomb group protein function in cancer. *Cell Res,* 32**,** 231-253.

Pauzaite, T., Thacker, U., Tollitt, J. & Copeland, N. A. 2016. Emerging Roles for Ciz1 in Cell Cycle Regulation and as a Driver of Tumorigenesis. *Biomolecules,* 7.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H. & Bateman, A. 2022. InterPro in 2022. *Nucleic Acids Res*.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. 1996. Requirement for Xist in X chromosome inactivation. *Nature,* 379**,** 131-7.

Peters, A. H., Mermoud, J. E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N. & Jenuwein, T. 2002. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet,* 30**,** 77-80.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci,* 30**,** 70-82.

Pietrantoni, G., Ibarra-Karmy, R. & Arriagada, G. 2020. Microtubule Retrograde Motors and Their Role in Retroviral Transport. *Viruses,* 12.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B. A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T. B. & Brockdorff, N. 2017. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell,* 68**,** 955-969 e10.

Pirie, K., Peto, R., Green, J., Reeves, G. K., Beral, V. & Million Women Study, C. 2016. Lung cancer in never smokers in the UK Million Women Study. *Int J Cancer,* 139**,** 347-54.

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B. & Zhang, Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science,* 300**,** 131-5.

Qian, J., Beullens, M., Lesage, B. & Bollen, M. 2013. Aurora B defines its own chromosomal targeting by opposing the recruitment of the phosphatase scaffold Repo-Man. *Curr Biol,* 23**,** 1136-43.

Qian, J., Lesage, B., Beullens, M., Van Eynde, A. & Bollen, M. 2011. PP1/Repo-man dephosphorylates mitotic histone H3 at T3 and regulates chromosomal aurora B targeting. *Curr Biol,* 21**,** 766-73.

Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forne, I., Pichler, G., Horl, D., Fellinger, K., Spada, F., Bonapace, I. M., Imhof, A., Harz, H. & Leonhardt, H. 2015. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res,* 25**,** 911-29.

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

Rahman, F., Ainscough, J. F., Copeland, N. & Coverley, D. 2007. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. *Hum Mutat,* 28**,** 993-1004.

Rahman, F. A., Aziz, N. & Coverley, D. 2010. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. *BMC Cancer,* 10**,** 482.

Raposo, A. C., Casanova, M., Gendrel, A. V. & da Rocha, S. T. 2021. The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. *Biochem Soc Trans,* 49**,** 2549-2560.

Ray-Gallet, D. & Almouzni, G. 2022. H3-H4 histone chaperones and cancer. *Curr Opin Genet Dev,* 73**,** 101900.

Reithmeier, A., Panizza, E., Krumpel, M., Orre, L. M., Branca, R. M. M., Lehtio, J., Ek-Rylander, B. & Andersson, G. 2017. Tartrate-resistant acid phosphatase (TRAP/ACP5) promotes metastasis-related properties via TGFbeta2/TbetaR and CD44 in MDA-MB-231 breast cancer cells. *BMC Cancer,* 17**,** 650.

Richardson, A. L., Wang, Z. C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J. D., Livingston, D. M. & Ganesan, S. 2006. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell,* 9**,** 121-32.

Ridings-Figueroa, R., Stewart, E. R., Nesterova, T. B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S. A.,

Albadrani, G., Mansfield, W., Roulson, J. A., Brockdorff, N., Ainscough, J. F. X. & Coverley, D. 2017. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev,* 31**,** 876-888.

Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell,* 64**,** 615-23.

Rohnalter, V., Roth, K., Finkernagel, F., Adhikary, T., Obert, J., Dorzweiler, K., Bensberg, M., Muller-Brusselbach, S. & Muller, R. 2015. A multi-stage process including transient polyploidization and EMT precedes the emergence of chemoresistent ovarian carcinoma cells with a dedifferentiated and pro-inflammatory secretory phenotype. *Oncotarget,* 6**,** 40005-25.

Roseboom, T., de Rooij, S. & Painter, R. 2006. The Dutch famine and its long-term consequences for adult health. *Early Hum Dev,* 82**,** 485-91.

Ross, E. D., Minton, A. & Wickner, R. B. 2005. Prion domains: sequences, structures and interactions. *Nat Cell Biol,* 7**,** 1039-44.

Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C. D., Reinberg, D., Avner, P. & Heard, E. 2004. Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol,* 24**,** 5475-84.

Rubin, S. M., Sage, J. & Skotheim, J. M. 2020. Integrating Old and New Paradigms of G1/S Control. *Mol Cell,* 80**,** 183-192.

Rubinsztein, D., Wyttenbach, A. & Rankin, J. 1999. Intracellular inclusions, pathological markers in diseases caused by expanded polyglutamine tracts? *Journal of medical genetics,* 36**,** 265-270.

Russo, A. A., Jeffrey, P. D. & Pavletich, N. P. 1996. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol,* 3**,** 696-700.

Satyanarayana, A., Hilton, M. B. & Kaldis, P. 2008. p21 Inhibits Cdk1 in the absence of Cdk2 to maintain the G1/S phase DNA damage checkpoint. *Mol Biol Cell,* 19**,** 65-77.

Savinov, A. & Roth, F. P. 2021. Seeds of their own destruction: Dominant-negative peptide screening yields functional insight and therapeutic leads. *Cell Syst,* 12**,** 691-693.

Schekman, R. & Mellman, I. 1997. Does COPI Go Both Ways? *Cell,* 90**,** 197-200.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H. & Wanker, E. E. 1997. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell,* 90**,** 549-558.

Schiera, G., Di Liegro, C. M., Puleo, V., Colletta, O., Fricano, A., Cancemi, P., Di Cara, G. & Di Liegro, I. 2016. Extracellular vesicles shed by melanoma cells contain a modified form of H1.0 linker histone and H1.0 mRNA-binding proteins. *Int J Oncol,* 49**,** 1807-1814.

Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. & Wutz, A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J,* 25**,** 3110-22.

Shamloo, B. & Usluer, S. 2019. p21 in Cancer Research. *Cancers (Basel),* 11.

Sharma, S., Kelly, T. K. & Jones, P. A. 2010. Epigenetics in cancer. *Carcinogenesis,* 31**,** 27-36.

Sharp, J. A., Perea-Resa, C., Wang, W. & Blower, M. D. 2020. Cell division requires RNA eviction from condensing chromosomes. *J Cell Biol,* 219.

Sigler, P. B. 1988. Transcriptional activation. Acid blobs and negative noodles. *Nature,* 333**,** 210-2.

Simon, J. A. & Kingston, R. E. 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nature Reviews Molecular Cell Biology,* 10**,** 697-708.

Sirchia, S. M., Ramoscelli, L., Grati, F. R., Barbera, F., Coradini, D., Rossella, F., Porta, G., Lesma, E., Ruggeri, A., Radice, P., Simoni, G. & Miozzo, M. 2005. Loss of the inactive X chromosome and replication of the active X in BRCA1-defective and wild-type breast cancer cells. *Cancer Res,* 65**,** 2139-46.

Smith, H. C., Ochs, R. L., Lin, D. & Chinault, A. C. 1987. Ultrastructural and biochemical comparisons of nuclear matrices prepared by high salt or LIS extraction. *Mol Cell Biochem,* 77**,** 49-61.

Sofi, S., Williamson, L., Turvey, G. L., Scoynes, C., Hirst, C., Godwin, J., Brockdorff, N., Ainscough, J. & Coverley, D. 2022. Prion-like domains drive CIZ1 assembly formation at the inactive X chromosome. *J Cell Biol,* 221.

Soneson, C., Love, M. I. & Robinson, M. D. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res,* 4**,** 1521.

Song, Y., An, O., Ren, X., Chan, T. H. M., Tay, D. J. T., Tang, S. J., Han, J., Hong, H., Ng, V. H. E., Ke, X., Shen, H., Pitcheshwar, P., Lin, J. S., Leong, K. W., Molias, F. B., Yang, H., Kappei, D. & Chen, L. 2021. RNA editing mediates the functional switch of COPA in a novel mechanism of hepatocarcinogenesis. *J Hepatol,* 74**,** 135-147.

Spatz, A., Borg, C. & Feunteun, J. 2004. X-chromosome genetics and human cancer. *Nat Rev Cancer,* 4**,** 617-29.

Srivastava, S., Wang, S., Tong, Y. A., Hao, Z. M. & Chang, E. H. 1993. Dominant negative effect of a germ-line mutant p53: a step fostering tumorigenesis. *Cancer Res,* 53**,** 4452-5.

Staller, M. V., Holehouse, A. S., Swain-Lenz, D., Das, R. K., Pappu, R. V. & Cohen, B. A. 2018. A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst,* 6**,** 444-455 e6.

Staller, M. V., Ramirez, E., Kotha, S. R., Holehouse, A. S., Pappu, R. V. & Cohen, B. A. 2022. Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst,* 13**,** 334-345 e5.

Stewart, E. R., Turner, R. M. L., Newling, K., Ridings-Figueroa, R., Scott, V., Ashton, P. D., Ainscough, J. F. X. & Coverley, D. 2019. Maintenance of epigenetic landscape requires CIZ1 and is corrupted in differentiated fibroblasts in long-term culture. *Nat Commun,* 10**,** 460.

Stewart, S. & Fang, G. 2005. Destruction box-dependent degradation of aurora B is mediated by the anaphase-promoting complex/cyclosome and Cdh1. *Cancer Res,* 65**,** 8730-5.

Stoldt, S., Wenzel, D., Schulze, E., Doenecke, D. & Happel, N. 2007. G1 phase-dependent nucleolar accumulation of human histone H1x. *Biol Cell,* 99**,** 541-52.

Storrie, B., Pepperkok, R. & Nilsson, T. 2000. Breaking the COPI monopoly on Golgi recycling. *Trends in Cell Biology,* 10**,** 385-390.

Strahl, B. D. & Allis, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-45.

Strasburger, E. 1883. Die Controversen der indirecten Kerntheilung. *Archiv für Mikroskopische Anatomie,* 23**,** 246-304.

Strom, A. R. & Brangwynne, C. P. 2019. The liquid nucleome - phase transitions in the nucleus at a glance. *J Cell Sci,* 132.

Subik, K., Lee, J.-F., Baxter, L., Strzepek, T., Costello, D., Crowley, P., Xing, L., Hung, M.-C., Bonfiglio, T., Hicks, D. G. & Tang, P. 2010. The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. *Breast Cancer: Basic and Clinical Research,* 4.

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F. C., Ruderman, J. V. & Hershko, A. 1995. The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell,* 6**,** 185-97.

Sugiyama, K., Sugiura, K., Hara, T., Sugimoto, K., Shima, H., Honda, K., Furukawa, K., Yamashita, S. & Urano, T. 2002. Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene,* 21**,** 3103-11.

Sun, S., Schiller, J. H. & Gazdar, A. F. 2007. Lung cancer in never smokers--a different disease. *Nat Rev Cancer,* 7**,** 778-90.

Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. 2017. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc Natl Acad Sci U S A,* 114**,** 10654-10659.

Swarts, D. R. A., Stewart, E. R., Higgins, G. S. & Coverley, D. 2018. CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours. *Cell Cycle,* 17**,** 2268-2283.

Takagi, N. & Sasaki, M. 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature,* 256**,** 640-2.

Takeshita, M., Koga, T., Takayama, K., Ijichi, K., Yano, T., Maehara, Y., Nakanishi, Y. & Sueishi, K. 2013. Aurora-B overexpression is correlated with aneuploidy and poor prognosis in non-small cell lung cancer. *Lung Cancer,* 80**,** 85-90.

Tamburri, S., Lavarone, E., Fernandez-Perez, D., Conway, E., Zanotti, M., Manganaro, D. & Pasini, D. 2020. Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell,* 77**,** 840-856 e5.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C. & Zhang, Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res,* 45**,** W98-W102.

Tao, Y., Leteur, C., Calderaro, J., Girdler, F., Zhang, P., Frascogna, V., Varna, M., Opolon, P., Castedo, M., Bourhis, J., Kroemer, G. & Deutsch, E. 2009. The aurora B kinase inhibitor AZD1152 sensitizes cancer cells to fractionated irradiation and induces mitotic catastrophe. *Cell Cycle,* 8**,** 3172-81.

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., Speedy, H. E., Stefancsik, R., Thompson, S. L., Wang, S., Ward, S., Campbell, P. J. & Forbes, S. A. 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res,* 47**,** D941-D947.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S. & Brockdorff, N. 2012. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell,* 148**,** 664-78.

Thacker, U., Pauzaite, T., Tollitt, J., Twardowska, M., Harrison, C., Dowle, A., Coverley, D. & Copeland, N. A. 2020. Identification of DHX9 as a cell cycle regulated nucleolar recruitment factor for CIZ1. *Scientific Reports,* 10.

Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L. M., Backstrom, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C. M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J. M., Sivertsson, A., Sjostedt, E., Skogs, M., Stadler, C., Sullivan, D. P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Ponten, F., von Feilitzen, K., Lilley, K. S., Uhlen, M. & Lundberg, E. 2017. A subcellular map of the human proteome. *Science,* 356.

Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E. & Baylin, S. B. 2008. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res,* 18**,** 1171-9.

Tomasetti, C., Li, L. & Vogelstein, B. 2017. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science,* 355**,** 1330-1334.

Torres, C. M., Biran, A., Burney, M. J., Patel, H., Henser-Brownhill, T., Cohen, A. S., Li, Y., Ben-Hamo, R., Nye, E., Spencer-Dene, B., Chakravarty, P., Efroni, S., Matthews, N., Misteli, T., Meshorer, E. & Scaffidi, P. 2016. The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science,* 353.

Umene, K., Banno, K., Kisu, I., Yanokura, M., Nogami, Y., Tsuji, K., Masuda, K., Ueki, A., Kobayashi, Y., Yamagami, W., Nomura, H., Tominaga, E., Susumu, N. & Aoki, D. 2013. Aurora kinase inhibitors: Potential molecular-targeted drugs for gynecologic malignant tumors. *Biomed Rep,* 1**,** 335-340.

UniProt, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res,* 49**,** D480-D489.

Veiga, D. F. T., Nesta, A., Zhao, Y., Deslattes Mays, A., Huynh, R., Rossi, R., Wu, T. C., Palucka, K., Anczukow, O., Beck, C. R. & Banchereau, J. 2022. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. *Sci Adv,* 8**,** eabg6711.

Veitia, R. A., Caburet, S. & Birchler, J. A. 2018. Mechanisms of Mendelian dominance. *Clin Genet,* 93**,** 419-428.

Verheijen, R., van Venrooij, W. & Ramaekers, F. 1988. The nuclear matrix: structure and composition. *J Cell Sci,* 90 (Pt 1)**,** 11-36.

Viatour, P. 2012. Bridges between Cell Cycle Regulation and Self-Renewal Maintenance. *Genes Cancer,* 3**,** 670-7.

Vigneron, S., Sundermann, L., Labbe, J. C., Pintard, L., Radulescu, O., Castro, A. & Lorca, T. 2018. Cyclin A-cdk1-Dependent Phosphorylation of Bora Is the Triggering Factor Promoting Mitotic Entry. *Dev Cell,* 45**,** 637-650 e7.

Vincent-Salomon, A., Ganem-Elbaz, C., Manie, E., Raynal, V., Sastre-Garau, X., Stoppa-Lyonnet, D., Stern, M. H. & Heard, E. 2007. X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res,* 67**,** 5134-40.

Waddington, C. H. 1957. *The Strategy of the Genes*, George Allen & Unwin.

Wang, D. Q., Wang, K., Yan, D. W., Liu, J., Wang, B., Li, M. X., Wang, X. W., Liu, J., Peng, Z. H., Li, G. X. & Yu, Z. H. 2014. Ciz1 is a novel predictor of survival in human colon cancer. *Exp Biol Med (Maywood),* 239**,** 862-870.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. & Zhang, Y. 2004. Role of histone H2A ubiquitination in Polycomb silencing. *Nature,* 431**,** 873-878.

Wang, W., Qin, J. J., Voruganti, S., Nag, S., Zhou, J. & Zhang, R. 2015. Polycomb Group (PcG) Proteins and Human Cancers: Multifaceted Functions and Therapeutic Implications. *Med Res Rev,* 35**,** 1220-67.

Wang, X., Xie, H., Zhu, Z., Zhang, J. & Xu, C. 2022. Molecular basis for the recognition of CIZ1 by ERH. *FEBS J.*

Wang, Y., Li, X., Zhang, J., Liu, Q., Gao, P., Li, D., Zhang, S. & Liu, J. 2019. CIZ1 Expression Is Upregulated in Hemangioma of the Tongue. *Pathol Oncol Res,* 25**,** 1653-1658.

Wang, Y. E., Marinov, G. K., Wold, B. J. & Chan, D. C. 2013. Genome-wide analysis reveals coating of the mitochondrial genome by TFAM. *PLoS One,* 8**,** e74513.

Wang, Z., Yu, Z., Wang, G. H., Zhou, Y. M., Deng, J. P., Feng, Y., Chen, J. Q. & Tian, L. 2020. AURKB Promotes the Metastasis of Gastric Cancer, Possibly by Inducing EMT. *Cancer Manag Res,* 12**,** 6947-6958.

Warder, D. E. & Keherly, M. J. 2003. Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0-2)YYAC. *J Biomed Sci,* 10**,** 406-17.

Warfield, L., Tuttle, L. M., Pacheco, D., Klevit, R. E. & Hahn, S. 2014. A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proc Natl Acad Sci U S A,* 111**,** E3506-13.

Warrick, J. M., Paulson, H. L., Gray-Board, G. L., Bui, Q. T., Fischbeck, K. H., Pittman, R. N. & Bonini, N. M. 1998. Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila. *Cell,* 93**,** 939-949.

Weng, M. T. & Luo, J. 2013. The enigmatic ERH protein: its role in cell cycle, RNA splicing and cancer. *Protein Cell,* 4**,** 807-12.

Willcockson, M. A., Healton, S. E., Weiss, C. N., Bartholdy, B. A., Botbol, Y., Mishra, L. N., Sidhwani, D. S., Wilson, T. J., Pinto, H. B., Maron, M. I., Skalina, K. A., Toro, L. N., Zhao, J., Lee, C. H., Hou, H., Yusufova, N., Meydan, C., Osunsade, A., David, Y., Cesarman, E., Melnick, A. M., Sidoli, S., Garcia, B. A., Edelmann, W., Macian, F. & Skoultchi, A. I. 2021. H1 histones control the epigenetic landscape by local chromatin compaction. *Nature,* 589**,** 293-298.

Willis, A., Jung, E. J., Wakefield, T. & Chen, X. 2004. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene,* 23**,** 2330-8.

Wilson, R. H., Hesketh, E. L. & Coverley, D. 2016. The Nuclear Matrix: Fractionation Techniques and Analysis. *Cold Spring Harb Protoc,* 2016**,** pdb top074518.

Winham, S. J., Larson, N. B., Armasu, S. M., Fogarty, Z. C., Larson, M. C., McCauley, B. M., Wang, C., Lawrenson, K., Gayther, S., Cunningham, J. M., Fridley, B. L. & Goode, E. L. 2019. Molecular signatures of X chromosome inactivation and associations with clinical outcomes in epithelial ovarian cancer. *Hum Mol Genet,* 28**,** 1331-1342.

Wu, J., Huen, M. S., Lu, L. Y., Ye, L., Dou, Y., Ljungman, M., Chen, J. & Yu, X. 2009. Histone ubiquitination associates with BRCA1-dependent DNA damage response. *Mol Cell Biol,* 29**,** 849-60.

Wu, J., Lei, L., Gu, D., Liu, H. & Wang, S. 2016. CIZ1 is upregulated in hepatocellular carcinoma and promotes the growth and migration of the cancer cells. *Tumour Biol,* 37**,** 4735-42.

Wutz, A. 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet,* 12**,** 542-53.

Wutz, A., Rasmussen, T. P. & Jaenisch, R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet,* 30**,** 167-74.

Xiao, C., Sharp, J. A., Kawahara, M., Davalos, A. R., Difilippantonio, M. J., Hu, Y., Li, W., Cao, L., Buetow, K., Ried, T., Chadwick, B. P., Deng, C. X. & Panning, B. 2007. The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell,* 128**,** 977-89.

Xiao, J., Uitti, R. J., Zhao, Y., Vemula, S. R., Perlmutter, J. S., Wszolek, Z. K., Maraganore, D. M., Auburger, G., Leube, B., Lehnhoff, K. & LeDoux, M. S. 2012. Mutations in CIZ1 cause adult onset primary cervical dystonia. *Ann Neurol,* 71**,** 458-69.

Xiao, J., Vemula, S. R., Xue, Y., Khan, M. M., Kuruvilla, K. P., Marquez-Lona, E. M., Cobb, M. R. & LeDoux, M. S. 2016. Motor phenotypes and molecular networks associated with germline deficiency of Ciz1. *Exp Neurol,* 283**,** 110-20.

Xiao, J. & Zhang, Y. 2021. AURKB as a Promising Prognostic Biomarker in Hepatocellular Carcinoma. *Evol Bioinform Online,* 17**,** 11769343211057589.

Xiao, X., Liu, C., Pei, Y., Wang, Y. Z., Kong, J., Lu, K., Ma, L., Dou, S. X., Wang, P. Y., Li, G., Chen, P. & Li, W. 2020. Histone H2A Ubiquitination Reinforces Mechanical Stability and Asymmetry at the Single-Nucleosome Level. *J Am Chem Soc,* 142**,** 3340-3345.

Xu, S., Zhong, M., Zhang, L., Wang, Y., Zhou, Z., Hao, Y., Zhang, W., Yang, X., Wei, A., Pei, L. & Yu, Z. 2009. Overexpression of Tfam protects mitochondria against beta-amyloid-induced oxidative damage in SH-SY5Y cells. *FEBS J,* 276**,** 3800-9.

Yang, J., Lee, S. J., Kwon, Y., Ma, L. & Kim, J. 2020a. Tumor suppressive function of Matrin 3 in the basal-like breast cancer. *Biol Res,* 53**,** 42.

Yang, L., Kirby, J. E., Sunwoo, H. & Lee, J. T. 2016. Female mice lacking Xist RNA show partial dosage compensation and survive to term. *Genes Dev,* 30**,** 1747-60.

Yang, L., Yildirim, E., Kirby, J. E., Press, W. & Lee, J. T. 2020b. Widespread organ tolerance to Xist loss and X reactivation except under chronic stress in the gut. *Proc Natl Acad Sci U S A,* 117**,** 4262-4272.

Yang, S. M., Kim, B. J., Norwood Toro, L. & Skoultchi, A. I. 2013. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci U S A,* 110**,** 1708-13.

Yin, J., Wang, C., Tang, X., Sun, H., Shao, Q., Yang, X. & Qu, X. 2013. CIZ1 regulates the proliferation, cycle distribution and colony formation of RKO human colorectal cancer cells. *Mol Med Rep,* 8**,** 1630-4.

Zamay, T. N., Zamay, G. S., Kolovskaya, O. S., Zukov, R. A., Petrova, M. M., Gargaun, A., Berezovski, M. V. & Kichkailo, A. S. 2017. Current and Prospective Protein Biomarkers of Lung Cancer. *Cancers (Basel),* 9.

Zarkowska, T. & Mittnacht, S. 1997. Differential phosphorylation of the retinoblastoma protein by G1/S cyclin-dependent kinases. *J Biol Chem,* 272**,** 12738-46.

Zeitz, M. J., Malyavantham, K. S., Seifert, B. & Berezney, R. 2009. Matrin 3: chromosomal distribution and protein interactions. *J Cell Biochem,* 108**,** 125-33.

Zeng, X., Chen, S. & Huang, H. 2011. Phosphorylation of EZH2 by CDK1 and CDK2: a possible regulatory mechanism of transmission of the H3K27me3 epigenetic mark through cell divisions. *Cell Cycle,* 10**,** 579-83.

Zhang, D., Wang, Y., Dai, Y., Wang, J., Suo, T., Pan, H., Liu, H., Shen, S. & Liu, H. 2015. CIZ1 promoted the growth and migration of gallbladder cancer cells. *Tumour Biol,* 36**,** 2583-91.

Zhang, H. S., Gavin, M., Dahiya, A., Postigo, A. A., Ma, D., Luo, R. X., Harbour, J. W. & Dean, D. C. 2000. Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF. *Cell,* 101**,** 79-89.

Zhang, J., Lin, X., Wu, L., Huang, J. J., Jiang, W. Q., Kipps, T. J. & Zhang, S. 2020. Aurora B induces epithelial-mesenchymal transition by stabilizing Snail1 to promote basal-like breast cancer metastasis. *Oncogene,* 39**,** 2550-2567.

Zhang, L. F., Huynh, K. D. & Lee, J. T. 2007. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell,* 129**,** 693-706.

Zhao, J., Dynlacht, B., Imai, T., Hori, T. & Harlow, E. 1998. Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. *Genes Dev,* 12**,** 456-61.

Zhao, Y. & Garcia, B. A. 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol,* 7**,** a025064.

Zhou, X., Liu, Q., Wada, Y., Liao, L. & Liu, J. 2018. CDKN1A-interacting zinc finger protein 1 is a novel biomarker for lung squamous cell carcinoma. *Oncol Lett,* 15**,** 183-188.

Zlatanova, J. & Doenecke, D. 1994. Histone H1 zero: a major player in cell differentiation? *FASEB J,* 8**,** 1260-8.

Zylicz, J. J., Bousard, A., Zumer, K., Dossin, F., Mohammad, E., da Rocha, S. T., Schwalb, B., Syx, L., Dingli, F., Loew, D., Cramer, P. & Heard, E. 2019. The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell,* 176**,** 182-197 e23.