MODELLING HUMAN ESCAPE FROM X-CHROMOSOME INACTIVATION IN MOUSE

by

Samantha Peeters

B.Sc., The University of Guelph, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Modelling human escape from X-chromosome inactivation in mouse

submitted by Samantha Peeters in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Medical Genetics

Examining Committee:

Dr. Carolyn Brown
Supervisor

Dr. Elizabeth Simpson
Supervisory Committee Member

Supervisory Committee Member

Dr. Dixie Mager
University Examiner

Dr. LeAnn Howe
University Examiner

Additional Supervisory Committee Members:

Dr. Matthew Lorincz
Supervisory Committee Member

Dr. Elizabeth Rideout
Supervisory Committee Member

Abstract

A long-standing question concerning X-chromosome inactivation has been how some genes avoid the otherwise stable chromosome-wide heterochromatinization of the inactive X. As 20% or more of human X-linked genes escape from inactivation, such genes are important contributors to sex differences in gene expression, and identifying the mechanism by which these exceptions occur will inform our understanding of X-inactivation and broader questions of epigenetic regulation. While bioinformatic studies have generated a list of candidate features, the nature of the elements or definitive evidence that any one particular element is necessary or sufficient for a gene to escape, is still elusive and requires experimental validation. Mouse models offer a well-characterized and readily manipulated system in which to study X-inactivation and escape, but have far fewer genes and gene clusters that escape than humans. Given these differences, it was unclear whether the mechanism of escape gene regulation is conserved between species, and thus, this thesis addresses conservation of the escape process and the potential to model human escape gene regulation using mouse systems.

Bacterial artificial chromosomes carrying genes known to escape from X-inactivation in humans were targeted to the *Hprt* locus and studied on the inactive X in mice. They were examined for escape by expression and inactivation-associated DNA methylation of promoter CpG islands. Expression from the inactive X and corresponding low

promoter DNA methylation of human gene *RPS4X* demonstrated that the mouse system is capable of recognizing human elements. Furthermore, the escape status of the transgene remained stable between developmental time points, tissues, and individual females. A second human escape gene, *KDM5C*, was targeted to the *Hprt* locus and was surprisingly subject to inactivation, suggesting that its mechanism of escape was not conserved or that the critical elements for escape were not contained in the transgene. To further interrogate the escape elements involved in both human genes analyzed, as well as additional constructs of interest, a docking site at *Hprt* was generated in a female mouse embryonic stem cell line. Overall, this thesis contributed to the development of approaches to examine human escape from inactivation, and characterization of two human escape regions.

Lay Summary

In general, human females have two X chromosomes, while males have only one X chromosome in addition to a comparatively gene-poor Y chromosome. To balance gene expression between the sexes, one X in the female is randomly inactivated early in development and remains inactive in all descendant cells. Although the majority of genes on the X chromosome are silenced, a significant portion is able to escape inactivation and continue to be expressed, contributing to differences between the sexes including susceptibility to or protection from disease. X-chromosome inactivation and escape is mainly studied using mouse models as they offer a readily manipulated system and access to developmental time points. However, as it is likely that there are differences in the regulatory processes between species, human escape genes have been modelled in mouse to demonstrate conservation between species, and a cell line model has been generated for further studies.

Preface

Approval for the generation and breeding of all mice discussed in this thesis was obtained from the University of British Columbia Committee on Animal Care (A10-0267, A10-0268, A14-0294, and A17-0276).

Chapter 2

A version of this chapter was published as:

Peeters SB, Korecki AJ, Simpson EM, Brown CJ. (2018). Human cis-acting elements regulating escape from X-chromosome inactivation function in mouse. Human Molecular Genetics, Volume 27, Issue 7: 1252–1262.

and reproduced by permission of Oxford University Press.

I performed all experiments and analysis in this chapter except for generation of the mouse strains (BAC electroporation and ESC microinjection), breeding, and tissue collection which were performed by collaborators in the Simpson Lab (AJ Korecki, and TC Lengyell of the Mouse Animal Production Services team). I generated all figures and wrote the manuscript with feedback from co-authors.

Chapter 3

A version of Chapter 3 was published as:

Peeters SB, Korecki AJ, Baldry SEL, Yang C, Tosefsky K, Balaton BP, Simpson EM, Brown CJ. (2019). How do genes that escape from X-chromosome inactivation contribute to Turner Syndrome? American Journal of Medical Genetics Part C: Seminars in Medical Genetics, Volume 181(1):28-35.

and reproduced by permission of Wiley Online Library.

RPS4X transgenic mice were previously generated in (Peeters et al, 2018) with tissue collection by AJ Korecki of the Simpson Lab. I performed all experiments and analysis in this chapter except for Western blotting which was performed by SEL Baldry of the Brown Lab. The manuscript was written with CJ Brown and I generated all figures reproduced in this thesis.

Chapter 4

This chapter is unpublished work. The CanEuCre transgenic mice were previously generated by the Simpson Lab and collaborators (Korecki et al, 2019; EM Simpson unpublished material) with tissue collection for this data analysis by AJ Korecki. The 129/Cast F1 2-1 ES cell line was gifted from Dr. Joost Gribnau. I performed all experiments and analyses described in this chapter, and generated all the figures.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

129: *Mus musculus* 129/Sv(ImJ)

3'UTR: three prime untranslated region

6-TG: 6-thioguanine

B6: *Mus musculus* C57BL/6J

BAC: bacterial artificial chromosome

bp: base pair

Cast: *Mus castaneus* Cast/EiJ

cDNA: complementary deoxyribonucleic acid

CTCF: CCCTC binding factor protein

CanEuCre: Canadian partnership with the European Commission for the development of new cre-driver resources

CRISPR: clustered regularly interspaced short palindromic repeats

DNA: deoxyribonucleic acid

DNAm: DNA methylation

dNTP: deoxyribonucleotide

E: embryonic day

ESC: embryonic stem cell

gDNA: genomic deoxyribonucleic acid

gRNA: guide ribonucleic acid

HAT: hypoxanthine aminopterin thymidine

kb: kilo base pairs

lnc: long non-coding

LINE: long interspersed nuclear element

Mb: mega base pairs

MEF: mouse embryonic fibroblast

NPC: neural progenitor cell

PAR: pseudoautosomal region

(q)PCR: (quantitative) polymerase chain reaction

RNA: ribonucleic acid

RT: reverse transcriptase

SNP: single nucleotide polymorphism

TAD: topologically associated domain

TSS: transcription start site

UCSC: University of California, Santa Cruz

Xa: active X chromosome

Xce: *X-controlling element*

XCI: X-chromosome inactivation

Xi: inactive X chromosome

Xic/*XIC*: X-inactivation center

Y: Y chromosome

List of Gene Names

Note that only genes which are discussed in the body of the thesis are included in this list. Gene symbols in all capital letters refer to human genes, mouse genes have a capital first letter followed by lowercase letters, and bacterial genes are written in all lowercase letters.

AGTR1: Angiotensin II Receptor Type 1

AMOTL1: Angiomotin like 1

camR: Chloramphenicol resistance gene

CARTPT: CART Prepropeptide

CLDN5: Claudin 5

CITED1/*Cited1*: Cbp/P300 Interacting Transactivator With Glu/Asp Rich Carboxy-Terminal Domain 1

CLVS2: Clavesin 2

CRH: Corticotropin Releasing Hormone

ERCC6L/*Ercc6l*: ERCC Excision Repair 6Like, Spindle Assembly Checkpoint Helicase

Flna: Filamin, alpha

GABRA6: Gamma-Aminobutyric Acid Type A Receptor Alpha6 Subunit

Hbb-bs: Hemoglobin, beta adult s chain

HDAC8: Histone Deacetylase 8

HPRT/*Hprt*: Hypoxanthine Guanine Phosphoribosyl Transferase

HTR1B: 5-Hydroxytryptamine Receptor 1B

IQSEC2/*Iqsec2*: IQ motif and Sec7domain 2

KANTR: KDM5C Adjacent Transcript

KCNA4: Potassium Voltage-Gated Channel Subfamily A Member 4

KDM5C/*Kdm5c*: Lysine-Specific Demethylase 5C

KDM6A/*Kdm6a*: Lysine Demethylase 6A

MAOA: Monoamine Oxidase A

MKI67: Marker of Proliferation Ki-67

NEUROD6: Neuronal Differentiation 6

NGFR: Nerve Growth Factor Receptor

NKX6-1: NK6 Homeobox 1

NOV: Nephroblastoma Overexpressed

NPY2R: Neuropeptide Y Receptor Y2

NR2E1: Nuclear Receptor Subfamily 2, Group E, Member 1

Pgk1: Phosphoglycerate kinase 1

PHB: Prohibitin

Phf6: PHD Finger Protein 6

PIN4: Peptidylprolyl Cis/Trans Isomerase, NIMA-Interacting 4

POU4F2: POU Class 4 Homeobox 2

puroR: puromycin resistance gene

RPS4X/*Rps4x*: Ribosomal Protein S4 X-linked

SLITRK6: SLIT And NTRK Like Family Member 6

SMC1A/*Smc1a*: Structural Maintenance of Chromosomes 1A

SOX3: SRY-Box 3

SPRY1: Sprouty RTK Signaling Antagonist 1

SRY: Sex Determining Region Y

Taf1: TATA-box binding protein associated factor 1

TSPYL2/*Tspyl2*: TSPY-like 2

XIST/*Xist*: Xi-Specific Transcripts

Zfx: Zinc finger protein X-linked

Acknowledgements

First and foremost I'd like to express my gratitude to my supervisor Dr. Carolyn Brown, I feel incredibly fortunate to have had the opportunity to learn from her over the years. I'm grateful for her patience, guidance, and for consistently inspiring me with her curiosity and passion for science. Thank you for investing in me.

I would like to extend my appreciation to my committee members Dr. Matthew Lorincz, Dr. Elizabeth Rideout, and particularly Dr. Elizabeth Simpson, for their time and feedback helping to shape this thesis and my progress as a researcher. This work would not have been possible without our collaborators in the Simpson lab, especially the contributions of Andrea Korecki whose mentorship, teamwork, and technical assistance was indispensable. Additional thanks to past and present members of the Brown lab and Molecular Epigenetics Group for all of their insights and advice, and for providing a supportive work environment.

My experience as a graduate student would not have been the same without the group of people I was lucky enough to be surrounded by, and I am forever grateful for the endless support, life talks, science talks, laughter, and for giving me a place to belong. Jenny Huang, Grace Tharmarajah, Amanda Ha, Christine Yang, Tom Dixon-McDougall,

Bradley Balaton, Julien Richard Albert, Aaron Bogutz, Adam Chruscicki, and Nicolas Coutin, thank you for your friendship.

Lastly, I would like to acknowledge my parents Joanne Wasiluk and Finley Peeters, and my grandmother Barbara Wasiluk, for their unwavering love and encouragement. Thank you for your confidence in me, and for teaching me that nothing is out of reach.

Dedication

To my childhood self who wanted to write a book one day, I don't think this is quite what she had in mind.

Chapter 1: Introduction

1.1 Thesis overview

X-chromosome inactivation (XCI) epigenetically silences one X chromosome (Xi) in every cell in female mammals, but approximately 12-27% of human and 3-7% of mouse X-linked genes escape from XCI and continue to be expressed albeit at levels lower than their active X (Xa) copy. Determining which genes escape from inactivation reveals an important source of sexually dimorphic gene expression, and identifying the mechanism by which these exceptions occur will inform our understanding of XCI and broader questions of selective epigenetic repression of genes. Several mechanisms have been suggested to contribute to escape, with some of the strongest evidence being for an intrinsic “escape element” in the DNA sequence in or near the gene. Additional components that likely facilitate escape from XCI include a lack of waystations to boost the silencing signal along the X, and boundary elements that function to inhibit spread of both euchromatin and heterochromatin.

Given the complexities of human research and the need to recapitulate early time points in development, mouse has been the leading model for study of XCI and escape. Known differences between the species suggest that mechanisms of XCI and escape may not be completely conserved, and so work in this thesis is focused on investigating conservation of mechanisms between the species in order to use knowledge gained in mouse to inform our understanding of escapee biology in humans. Specifically, the XCI

statuses of human transgenes docked at an X-linked locus in mice are explored using expression and DNA methylation (DNAm) assays to directly and indirectly assess their potential for escape. Escape of human genes in mouse is characterized at an organismal level and then regions of interest are pursued in an *in vitro* model for further study.

1.2 XCI and dosage compensation

Primary sex-determination in most mammals, including humans, makes use of the inheritance patterns of two distinct chromosomes, the X and the Y. In general, mammalian females have two X chromosomes, while males have only one X in addition to the male-determining Y chromosome. Once an ordinary pair of autosomes, the X and Y began to diverge when the Y acquired the *SRY* gene (sex-determining region Y) which encodes a transcription factor initiating testis development (Gubbay et al, 1990; Sinclair et al, 1990). The Y began to accumulate more genes involved in spermatogenesis, and was progressively degraded after a series of inversions and deletions suppressed its ability to recombine with the X during meiosis (save for two segments at either end of the X and Y chromosomes termed pseudoautosomal regions, or PAR). Consequently, the present X chromosome carries about ten times the number of protein-coding genes as the Y (reviewed in Bachtrog, 2013) (Figure 1.1).

With females having two copies of the more gene-rich X chromosome, it is hypothesized that XCI evolved to balance X-linked gene expression between the sexes. XCI is an epigenetic process initiated early in development by the long non-coding (lnc)RNA XIST (X-inactive specific transcript) (Borsani et al,1991; Brockdorf et al, 1991; Brown et al, 1991). *XIST* is expressed from the X-inactivation center (*XIC*, the region of the X that is required in *cis* for XCI to occur) exclusively on the Xi where it coats the chromosome in *cis* and works in conjunction with a number of silencing pathways to establish and lock-in XCI (reviewed in Balaton et al, 2018). Accumulation of repressive chromatin marks such as histone modifications and DNA hypermethylation at CpG island-containing promoters result in gene silencing, chromatin condensation, and late-replication of the chromosome (reviewed in Peeters et al, 2014; Dixon-McDougall and Brown, 2015).

1.2.1 XCI in human and mouse: differences in choice and timing

The initial choice of which X to inactivate (or which X to remain active) in humans is generally random and clonally maintained through cell division, resulting in females being mosaics for allelic expression of X-linked genes. Deviations from a random 50/50 split of cells, where the maternal or the paternal X is inactivated more often in a tissue or an entire organism, is called skewing. Skewing can result from chance, choice (for example, through manipulation of critical genes involved in XCI) or selection over the lifetime of a female (for example, of cells that have a gene expressed on the Xa that confers a growth advantage) (reviewed in Peeters et al, 2016).

In contrast, there are two well-defined deviations from initial random inactivation of either X in mouse. The first is the completely non-random (imprinted) inactivation of the paternal X in extraembryonic lineages (Tagaki and Sasaki, 1975), a process also at work in several other species (Table 1.1). Mary Lyon was one of the first to suggest that the timing of inactivation was an important factor in non-random choice, observing that species (or tissues) that are the earliest to show an Xi during development are the ones that have a silent paternal X (Lyon, 1993). Several lines of evidence suggest that the presence of an imprint laid on the maternal X during oogenesis makes it extremely resistant to chromosome silencing, although the possible presence of an imprint laid on the paternal X predisposing it to silencing has not been completely excluded (reviewed in Sado, 2017).

The imprinted XCI is then erased in the inner cell mass of the mouse embryo giving rise to the epiblast, where a second wave of XCI occurs. Here the deviation from random XCI is much more subtle, and involves competing alleles of the X-controlling element (*Xce*), whereby strains with a stronger allele in hybrid crosses will be more likely to remain the Xa in a larger population of cells (reviewed in Galupa and Heard, 2015; Peeters et al, 2016). Although the nature of *Xce* locus and its mechanisms of action are still under investigation, six functional alleles have been described in mouse, with additional evidence that the parent of origin of the allele plays a role in the degree of skewing (Calaway et al, 2013). However, there are still skewing differences in mice that cannot be accounted for by *Xce* or parental inheritance alone, suggesting there are

likely additional mechanisms at play (Crowley et al, 2015; Wu et al, 2014). To date, evidence for an *XCE* in humans is lacking, suggesting that the possibility of directed inactivation is unlikely.

1.3 Escape from XCI in humans

Despite the vast heterochromatic environment of the Xi, a significant number of genes are able to escape silencing (termed escapees) and continue to be expressed from both X chromosomes, albeit at lower levels on the Xi than the Xa (averaging 33%, Tukiainen et al, 2017). The lower expression is likely due to the dampening effect of the heterochromatic neighbourhood on the Xi, indeed even the PAR1 genes that escape from XCI have lower expression in females than in males, as the Y copy is more fully expressed than the region on the Xi (Tukiainen et al, 2017). The majority of escapees in humans were identified in a benchmark study in 2005 based on expression from a human Xi in rodent somatic cell hybrids, and verified by biallelic expression of single nucleotide polymorphisms (SNPs) in female cell lines with skewed XCI (Carrel and Willard, 2005). This study initially defined genes that escape from XCI as having Xi gene expression at or above 10% of Xa levels, but escape definitions have since been expanded using statistical methods to validate even lower levels of Xi expression (Calabrese et al, 2012; Berletch et al, 2015) or by comparing male–female expression differences (Tukiainen et al, 2017). In addition to expression, these escape genes are epigenetically different than their silenced neighbours, being depleted in *Xist* RNA and typically having promoters marked by active histone modifications and hypomethylation

of CpG islands (Figure 1.2A). When expression data is not available, analysis can be extended to these epigenomic features that differentiate active and inactive genes as a proxy measurement of XCI status, often through comparison of male to female levels. For example, the inverse correlation between gene activity, and DNAm of CpG islands overlapping X-linked gene promoters, has been established as a sound test for predicting the inactivation status of an X-linked gene (Cotton et al, 2011; Cotton et al, 2014).

Aggregation of data from studies including somatic cell hybrids and females with non-random XCI, biallelic expression in tissues and single cells, and assessment of differences in expression, chromatin marks or DNAm between males and females, assigned a consensus XCI status call to 639 genes (81% of X-linked protein-coding genes expressed in somatic cells), of which 80 genes consistently escape (termed constitutive) and 93 genes variably escape from XCI (termed facultative) (Balaton et al, 2015). Escapees often cluster in regions (particularly on the short (p) arm) containing blocks of up to 15 genes. These escape genes are enriched in regions with shorter divergence times from the Y, and many of them encode for regulatory proteins involved in transcription and translation and therefore may be more sensitive to dosage (Bellot et al, 2014). Indeed, the list of escapees includes all characterized PAR1 genes (~22), the two most centromeric PAR2 genes (the two telomeric genes are silenced on both the Xi and Y, Ciccodicola et al, 2000) and 12 (of 14 informative, 17 total) genes with functional ancestral X-Y homologues outside the PAR (Bellot et al, 2014), leaving many additional

genes that escape but lack expressed Y homologues. The variability in number of genes escaping from XCI is largely accounted for by genes that show individual, strain (in mouse) or tissue-specific expression. Additionally, escape status calls for some genes are discordant between studies as escape classification is in part sensitive to technical and statistical methods.

1.3.1 Escape from XCI in mouse

In contrast to humans, considerably fewer genes escape inactivation in mice with data suggesting only 17 constitutive escape genes, and approximately 20 variable escape genes (reviewed in Balaton and Brown, 2016). A slightly larger number of variable genes have been found to escape in *in vitro* cell populations as opposed to tissue samples, which could reflect the acquisition of epigenetic changes leading to reactivation of X-linked genes in cell culture (Berletch et al, 2015; Marks et al, 2015). As many of the crosses analysing escape from XCI in mouse used the same parental strains, escape has only been examined for only a limited number of haplotypes relative to the diversity of X chromosomes profiled in humans. Additionally, the number of tissues evaluated in mouse is not yet as broad as in humans (having been extensively examined through expression based studies in Tukiainen et al, 2017) so the number of tissue-specific escape genes may increase with further inclusion of strains and tissue types. The genes located in the human PAR1 region that escape from XCI are autosomal in the mouse, which has a smaller independently evolved PAR (Perry et al, 2001).

Approximately one half of the mouse escape genes also escape in humans suggesting some conservation of the elements and mechanisms involved between species. Studies of XCI status have not been as comprehensive in other species; however, a limited number of human escape genes have been investigated for conservation of Xi expression across a diverse group of mammals (Jegalian and Page, 1998; Okamoto et al, 2011; Nadaf et al, 2011; Yen et al, 2001). In contrast to the escape domains in humans, mouse escapees are predominantly singletons or associated with a noncoding RNA gene (Reinius et al, 2010), although several variable genes have been identified to contribute to small regions of escape in cell lines (Marks et al, 2015). The lack of regions of escape in mouse is more suggestive of local regulatory elements driving expression (Calabrese et al, 2012; Giorgetti et al, 2016). Either these elements differ in humans, or there is the possibility of a larger bystander effect in humans whereby a single gene may be under strict regulatory control, but a lack of stable adjacent boundaries allows spread of open chromatin enabling neighbouring genes to also escape.

1.4 DNA elements model of escape

Human genes retaining an active homologue on the Y chromosome are often found in the escapee category, such as those in the PAR1, supporting the hypothesis that genes escape to achieve dosage compensation. However, there are genes in the PAR2 that have homologues on the Y that achieve dosage compensation through inactivation on both the Xi and the Y (Ciccodicola et al, 2000) demonstrating that evading inactivation is

not a basic characteristic of PAR genes, or restricted to genes with a Y homologue as discussed above. While most of the escape genes in humans cluster towards the distal p arm, some of the consistent escapees and most of the variable human escapees are found scattered throughout the rest of the chromosome, implying that there may be a different mechanism controlling their escape. This mechanism may be similar to that found in the mouse where there is less clustering of escapees.

Since neither genomic location nor dosage compensation can fully explain why some genes can escape the influence of *XIST* and repression by the heterochromatic environment, evidence is building to support the idea that it is both the genomic context of DNA sequences and epigenetic interactions at a particular location that determines whether or not a gene can escape from XCI. DNA elements contributing to whether a gene escapes from XCI have been categorized as waystations, escape elements and boundaries (Figure 1.2B). Waystations are thought to propagate the silencing signal along the X chromosome, boundary elements act as insulators between active and inactive domains, and sequences termed escape elements allow the gene to avoid complete silencing. If all three are found in the right configuration, the gene is able to be expressed in the midst of an otherwise silent chromosome (discussed in Yang et al, 2011; Balaton and Brown, 2016; Peeters et al, 2019). Hypotheses and evidence have been presented for and against each of these features as well as expanded their definitions over the years, and so identification and testing to elucidate their roles is central to refining our understanding of the XCI escape model.

1.4.1 Waystations

Waystations, or booster elements, were initially postulated to explain how the chromosome condensation signal spreads from the *XIC* along the X chromosome (Gartler and Riggs,1983). If there is a lack of these waystations in a particular area, then the model predicts that there is nothing to amplify the signal and genes are not efficiently inactivated, allowing genes to undergo transcription because they are far enough away from the silencing signal of *XIST* and other factors. The most convincing choice for a waystation was proposed to be long interspersed nuclear elements (LINEs, Lyon, 1998) as they were shown to be enriched on the X chromosome as a whole compared to the autosomes (Boyle et al, 1990). This 'repeat hypothesis' is supported by bioinformatic studies showing that LINEs are enriched around the transcriptional start sites (TSS) of genes subject to XCI and depleted in regions that harbour genes escaping from XCI (Bailey et al, 2000; Carrel et al, 2006; Wang et al, 2006).

Further evidence for the correlation between LINE density and the efficiency of the spread of silencing comes from X;autosome translocations (Cotton et al, 2013; Tannan et al, 2013) as well as mouse *Xist* transgenes (Loda et al, 2017) where the spread of silencing is reduced relative to the extent of silencing seen on the X. In addition to a positive correlation between silencing and presence of LINE elements, the strongest correlation observed with subject genes was with pre-existing features of heterochromatin, most significantly for EZH2, a component of PRC2 (Cotton et al, 2013). Despite the reduced spread of silencing seen on the autosomes described

above, studies of autosomal bacterial artificial chromosome (BAC) integrations into the 5' end of the X-linked *Hprt* locus in mouse found all but one transgene to be subject to XCI, suggesting waystations could boost silencing across distances of up to 200 kb of integrated DNA (Yang et al, 2012).

It is challenging to separate the enrichment in LINEs correlated with lack of escape from the unique evolutionary history of the X chromosome. The X diverged away from the Y chromosome through reduced recombination into segments termed evolutionary strata (Lahn and Page, 1999). It was observed that the oldest evolutionary strata had picked up more LINEs over time than the younger strata on the p arm, which coincidentally holds most of the escape domains (Brown and Greally, 2003). If LINEs actually play such a key role in spreading XCI as suggested, they may have had less time to accumulate and inactivate this newer X-chromosome material.

Functionally, DNA and RNA fluorescent *in situ* hybridization at progressive time points in differentiating mouse embryonic stem cells (ESCs) suggested that LINEs may facilitate XCI at different levels (Chow et al, 2010). Silent LINEs could help to facilitate nucleation of a heterochromatic compartment by XIST RNA, thereby playing a role in the spacial segregation of inactivated and escaping genes via repeat-induced associations, and transcriptionally active LINEs may participate in local propagation of XCI into regions that would otherwise be prone to escape (Chow et al, 2010). Further

confounding evidence for LINEs as waystations is that lack of active LINEs does not interfere with XCI in some rodent species (Cantrell et al, 2009), and there is no evidence to date that XIST RNA is targeted to these sequences or other repeats (Engreitz et al, 2013; Simon et al, 2013).

1.4.2 Boundary elements and structural variation on the Xi

CTCF (CCCTC-Binding Factor) binding has been hypothesized to be responsible for setting boundaries around escapees, with differential binding between species creating the larger domains of escape genes in humans than in mouse, which tend to be solo escapees (Filippova et al, 2007). It has been suggested that shifting the binding site of CTCF in different cell lines can adjust which genes in a region are escaping (Berletch et al, 2015) and may be part of the explanation for genes that variably escape. However, given the abundance of CTCF sites available on the X**,** and that CTCF sites alone could not insulate a transgenic reporter gene from inactivation (Ciavatta et al, 2006), it is unlikely that CTCF acts independently in the escape mechanism. An analysis in mouse showed specific enrichment of CTCF at the TSSs of X-linked but not autosomal genes that escape ectopic XCI, pointing to an X chromosome-specific enrichment of CTCF at escaping loci (Loda et al, 2017).

Allelic ultrastructural studies of the X chromosome have revealed that in both humans and mice the Xi forms a distinctive superdomain structure rather than the topologically

associated domains (TADs) that are observed across the Xa and the autosomes (Deng et al, 2015; Giorgetti et al 2016). Yet evidence also suggests that continued transcription of genes escaping from XCI, along with binding of factors such as CTCF, may enable the maintenance or re-creation of TAD structures around regions containing escape genes (Giorgetti et al, 2016). Indeed, the spread of inactivation in human translocations showed a tendency for genes within a TAD to behave similarly as either subject or escape (130 of 195 domains, Cotton et al, 2013). In human female fibroblasts, all of the associated genes in 15 of the 17 TADs either escape XCI or are silenced, and rare clusters of mouse escapees (three regions) in neural progenitor cells (NPCs) co-localize with TADs (Marks et al, 2015) suggesting genomic architecture may be playing a role in escape. The escaping lncRNA loci involved in the superloop formation, including the superdomain hinge region *Dxz4*/*DXZ4*, are enriched in Xi-specific CTCF sites. However, deletions and inversions of this hinge region after XCI induction in human (Darrow et al, 2016) and mouse cells (Bonora et al, 2018) had little impact on gene expression, attesting to the multiple layers of control that ensure silencing of the Xi. Deletion of the *Dxz4* hinge region in mouse ESCs prior to XCI similarly did not impact constitutive escape genes, but did show loss of escape of variable escape genes in a single clone (Giorgetti et al, 2016).

1.4.3 Escape Elements

While waystations and boundary elements affect the spread of silencing or escape regions on the chromosome, the third component to the escape model suggests an

element that is enriched around escape genes, termed an ‘escape element’, that is somehow involved in specifically facilitating the ability of genes to break from the silence surrounding them. Sequences that have been proposed here include Alu repetitive elements and ncRNAs, as they have been found in the vicinity of escape genes through bioinformatic analysis of genomic environments (McNeil et al, 1991; Wang et al, 2006; Reinius et al, 2010). Alu elements can alter the distribution of DNAm, and possibly transcription of genes throughout the genome (reviewed in Batzer et al, 2002), as well as play a role in species-specific, repeat-driven expansions of CTCF binding (Schmidt et al, 2012). There is evidence for promoter-proximal elements being involved in escape (Calabrese et al, 2012; Mugford et al, 2014; Giorgetti et al, 2016), but whether they are the same regulatory sequences used by the expressed alleles on the Xa still needs to be determined. Some escape genes have alternative promoters that can differ in status (Goto and Kimura, 2009). *In silico* analysis of the TSSs of escape genes in humans has found significant over-representations of YY1 transcription factor binding motif and ChIP-seq peaks, and similar to CTCF, YY1 occupancy is significantly biased toward the Xi at loci that are frequent contacts of Xi-specific superloops (Chen et al, 2016). Additionally, components of the cohesin complex (RAD21, SMC3) have been found to co-localize with CTCF at these sites in both human and mouse female cells (Yang et al, 2015), again highlighting that there is likely a structural component to escape.

The strongest evidence for both an intrinsic escape element and boundary sequence comes from several studies in mouse cells of *Kdm5c*, a gene that escapes from XCI in

both humans and mouse (Agulnik et al, 1994). Random integrations of two overlapping BACs containing *Kdm5c*, as well as flanking genes normally subject to XCI, demonstrated ongoing escape from inactivation of *Kdm5c* at four different locations on the mouse X chromosome, while the flanking genes maintained their expected inactive state (Li and Carrel, 2008). A follow-up to this study analysed partially deleted integrations of the transgene and showed that deletion of the 3' end of *Kdm5c* extended the escape domain by disrupting proper silencing of three endogenous genes adjacent to the *Kdm5c* transgene integration (Horvath et al, 2013). These studies narrowed down a minimal region in the *Kdm5c* BAC containing an element(s) necessary for escape from XCI, and hinted at a missing boundary sequence when the 3' end was deleted.

1.5 Escape genes: contributions to sexual dimorphism and differences in disease

Escape genes are important mechanistically for basic research into epigenetics and gene regulation, but also have profound impacts on health including contributions to differences between the sexes in susceptibility to or protection from disease, and even differences within females as these genes have been shown to vary between tissues and individuals. The contribution of the second sex chromosome is most evident in Turner syndrome where over 99% of 45,X conceptuses fail to survive, with those that are viable having been suggested to have an undetected cryptic mosaicism, and therefore true 45,X is not viable (Hook and Warburton, 2014). In mice there is not as dramatic of an impact for X monosomy, which is in agreement with them having fewer

genes that escape from inactivation and a much smaller PAR with different genes than human (Perry et al, 2001). Indeed, species with larger PARs have lower frequencies of X0 births (Raudsepp et al, 2011). Therefore, the prime candidates for the phenotype in humans have been genes shared by the X and Y sex chromosomes (17 genes and PARs), in particular the nine in which Y (and Xi) expression is not shared with mouse (Bellot et al, 2014). Additionally, as isochromosome Xq is observed in over 15% of Turner syndrome karyotypes (Cameron-Pimblett et al, 2017), escape genes on the Xq arm are likely poor candidates for involvement in Turner syndrome, although incomplete dosage compensation due to over-expression rather than under-expression hasn't been ruled out.

As the products of X–Y conserved gene pairs are enriched in regulatory functions, their continued presence on the Y chromosome, and escape from inactivation on the X chromosome, argues for a strict requirement for continued dosage equivalence, an observation that is reinforced by the number of X-linked intellectual disability syndromes caused by haploinsufficiency of many of these genes (reviewed in Carrel and Brown, 2017). However, emerging evidence suggests a degree of functional divergence in such X–Y gene pairs as the Y homologue cannot always compensate for the X, and males who have a mutant X allele experience a more severe phenotype than heterozygous females (reviewed in Snell and Turner, 2018). Insufficiency of the Y homologue has also been established in recent work investigating the sex bias of many cancers, where it was shown that mutations in several genes that escape from XCI, called escape from X-

inactivation tumor suppressors, were more commonly found in males than females arguing that females are protected by expression from their second allele (Dunford et al, 2016). Therefore, the differential expression of escapees in females can offer a protective effect against *de novo* and inherited X-linked mutations, but has also been proposed to contribute to the over-representation of females for some complex traits. For example, autoimmune disorders are biased toward individuals having two X chromosomes, where overexpression of X-linked genes, specifically those that are variable between women, could impact disease predisposition (reviewed in Carrel and Brown, 2017).

1.6 Modelling human escape from XCI

Human XCI and escape has been predominately studied in human somatic cells and tissues, or hybrid mouse cells with a human Xi. However, measuring escape in human somatic cells and tissue samples requires skewed XCI or male data to make inferences about Xi expression, and is only useful for gathering data after inactivation has occurred. In order to test potential escape elements and mechanisms of escape, developmental models are needed that can be manipulated prior to XCI, and analysed afterward for effects on escape or silencing. Since the use of early human embryos is an ethically sensitive issue human pluripotent stem cell models have been developed, but they are extremely vulnerable to derivation and maintenance culture conditions leading them to have errant genetic and epigenetic regulation, including instability of XCI and high variability in patterns of XCI. This epigenetic instability has led to difficulty

in reproducing results and drawing definitive conclusions about XCI mechanisms and the inactivation status of genes in these lines (reviewed in Geens et al, 2017). Strides have been made in recent years optimizing culture conditions to more closely mimic early events in humans (Sahakyan et al, 2016; Guo et al, 2017) but such conditions have yet to be widely adopted by the field.

In contrast to human pluripotent stem cells, mouse ESCs have been well characterized and have more clearly defined culture conditions. However, establishing a stable XaXa ESC line can still be problematic as there is frequent loss of one of the two Xa chromosomes, although some lines have been validated that do tolerate having two Xas (Zvetkova et al, 2005). Despite the ability of some lines to retain both Xas during ESC culture, differentiation of these lines provides another opportunity to lose an X rather than inactivate. As it has been proposed that gene dosage from two Xas interferes with differentiation (Schulz et al, 2014), if XCI is not initiated in a robust manner the cell will tolerate, or even favour, losing an X in order to proceed.

Although some challenges remain in culturing mouse ESCs, full genome sequencing of mouse and information on SNP differences between strains has substantially increased the efficiency in perturbing one allele and measuring outcome. Sequence information has also proven crucial in setting up breeding schemes to have highly informative offspring with parent of origin information on active and silent alleles. Additionally,

mouse models offer experimental advantages over human including access to developmental time points and tissues at an organismal level. For these reasons, mouse systems continue to be a leading model for studying XCI in mouse, and to some extent humans (in the case of somatic cell hybrids or human transgenes in mouse).

1.7 *Hprt* docking site

While random insertion of transgenes has been useful for initial studies of escape from XCI (Li and Carrel, 2008), these cell lines can be challenging to characterize in terms of copy number and location in the genome, both factors that could influence expression of the transgene and therefore make comparisons between integrations problematic. To control for these influences, studies have made use of transgenes integrated at safe harbours such as the 5' end of the X-linked *Hprt* gene, which has been long described as a docking site for predictable homologous recombination of both plasmids and larger BACs, and has been shown to provide a relatively neutral environment in which the expression pattern of the knocked-in promoter is primarily maintained (Portales-Casamar et al, 2010; Schmouth et al 2013; de Leeuw et al, 2014).

Work from our lab described the assessment of more than 1.5 Mb of primarily autosomal human DNA integrated at the 5' end of the *Hprt* locus on the mouse X chromosome for evidence of escape from XCI (Yang et al, 2012). It was anticipated that escape might be frequent given that waystations have been suggested to be depleted

on autosomes, however only one truncated gene, *PHB*, was identified as being expressed from the Xi. The rarity of escape observed for BACs integrated 5' of *Hprt* may reflect an inability to capture the correct combination of escape elements and boundary factors in the constructs, the capability of XCI to spread across constructs up to 200 kb even if they lack additional waystations, or a resilience of the docking site to Xi expression. Since the majority of those transgenes were autosomal, a focus on integration of X-linked genes that typically escape from XCI is needed for proper analysis of intrinsic escape elements.

1.8 Thesis objectives

The goals of this thesis were to (1) investigate the conservation of escape elements between mouse and human by testing the ability of human genes to escape from XCI in mouse and (2) generate a female mouse cell line model to integrate candidate regions at a controlled locus, and functionally test regions for DNA elements regulating escape from XCI. Chapter 2 employs transgenic mouse models to demonstrate that a mouse system is capable of recognizing *cis*-acting human elements regulating escape from XCI of a primate-specific escape gene, and validates an X-linked locus as a permissible docking site for further study of escape transgenes. In Chapter 3, the transgenic model is further characterized throughout the lifespan of the mouse where the XCI statuses of the transgenes remain stable between time points, tissues, and individual females. In Chapter 4 the XCI status of a second human gene hypothesized to escape from XCI in mouse is explored with unexpected results. Generation of an *in vitro* mouse ESC model

is described for testing of additional constructs without the need for mouse generation and breeding time. Overall, this thesis advances our understanding of epigenetic silencing by studying exceptions to the rule: genes that escape from XCI.

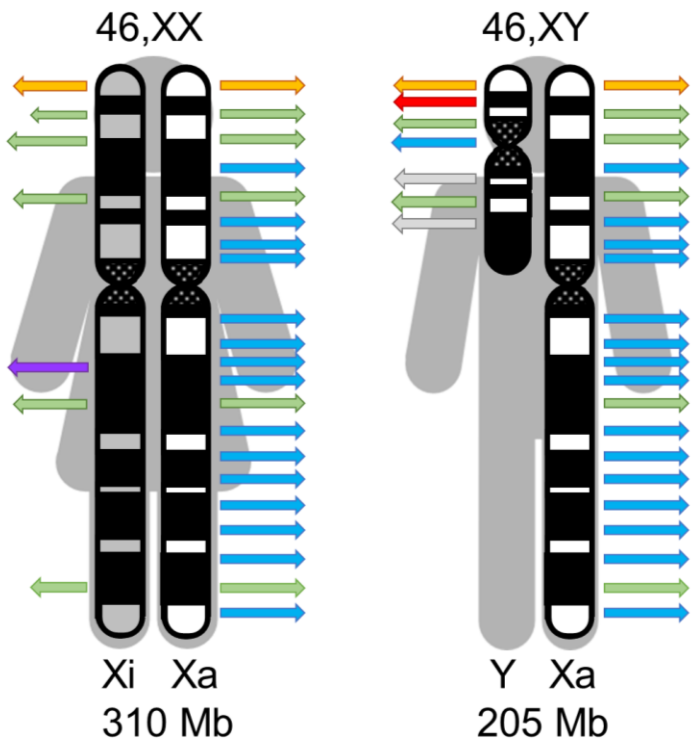


Figure 1.1 Gene expression from the sex chromosomes

Schematic of expression of genes from the X and Y chromosomes in 46,XX females (left), 46,XY males (right) to highlight differences between the sexes. Arrows reflect classes of genes expressed from each sex chromosome and shorter length corresponds to reduced gene expression from the Xi compared to Xa. Orange: PAR1; green: genes that escape from XCI; blue: genes that are subject to XCI; grey: Y-specific genes. The purple arrow indicates *XIST,* which is expressed only from the Xi, while the red indicates *SRY* which is the male sex-determining gene. In humans, the euchromatic part of the Y chromosome is about 23 Mb in size and contains 78 protein-coding genes, compared to 150 Mb of euchromatin and about 800 protein-coding genes on the X (reviewed in Bachtrog, 2013).

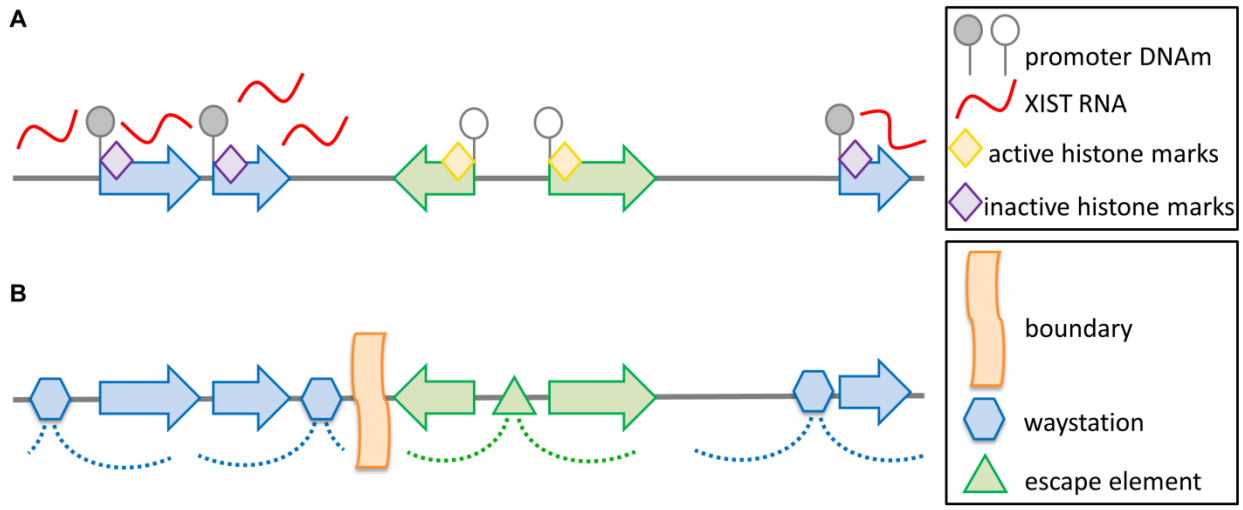


Figure 1.2 Features contributing to escape from XCI

A) Escapees (green) differ from silenced genes (blue) with respect to active and inactive epigenetic features such as active (yellow diamond) and inactive (purple diamond) histone marks, the presence of XIST RNA (red wavy line) and promoter DNAm (white lollipop = unmethylated, grey lollipop = methylated). **B)** DNA sequences such as waystations (blue hexagon), escape elements (green triangle) and boundary elements (orange rectangle) have been hypothesized to account for genes that are subject to and escape from XCI. Adapted from (Yang et al, 2001; Peeters et al, 2014).

Table 1.1 Timing of XCI in mammals

Columns adapted from Peeters et al, 2016.

Species	Non-random XCI (paternal imprint)	XCI in extra-embryonic tissue (days)[b]	XCI in embryonic tissue (days)[b]
Human	No[a] (de Mello et al, 2010; Okamoto et al, 2011)	10-12	12-20
Horse	No[a] (Wang et al, 2012)	7.5-10.5	11.5-12.5
Porcine	No (Ramos-Ibeas et al, 2019)		10-11
Rabbit	No (Okamoto et al, 2011)	5	5
Bovine	Yes; extra-embryonic (Xue et al, 2012) and morula[a] (Ferreira et al, 2010) No; extra-embryonic[a] (Xue et al, 2012) and blastocysts[a] (Ferreira et al, 2010; Min et al 2017)	7-9	14-15
Marsupial	Yes; embryonic and extra-embryonic (Cooper et al, 1971; Sharman, 1971)	unilaminar blastocyst	
Mouse	Yes; extra-embryonic (Tagaki et al, 1975)	3.5	5.5-6.5
Rat	Yes; extra-embryonic (Wake et al, 1976)	6	7

[a] assisted reproduction (*in vitro* fertilization, artificial insemination, cloning)

[b] timing of XCI reported is influenced by limited accessibility of samples from developmental time points in some species, as well as assay used to measure inactivation

Chapter 2: Human *cis*-acting elements regulating escape from XCI function in mouse

2.1 Introduction

Over 1.5 Mb of primarily autosomal human transgenic DNA integrated at the 5' end of the *Hprt* locus on the mouse X chromosome has been assessed for evidence of escape from XCI, with only one truncated autosomal gene, *PHB*, being identified as expressed from the Xi (Yang et al, 2012). While finding that one transgene escapes from XCI demonstrates that expression is possible from this location on the Xi, the rarity of escape observed for BACs integrated at *Hprt* may reflect a resilience of the docking site to Xi expression. Alternatively, the BAC carrying *PHB* could contain some combination of escape elements and boundary factors to block silencing, which is otherwise capable of spreading across large constructs even if they lack additional waystations. As the majority of the BACs tested were of autosomal origin, this system was revisited to determine if the *Hprt* site recognizes intrinsic escape elements of X-linked genes. To first validate the *Hprt* docking site as permissive to X-linked escape, one of the BACs carrying *Kdm5c* previously shown to escape from XCI (Li and Carrel, 2008) was selected for integration into the locus, and analysis of escape ability. A gene is generally called as escaping if the proportion of gene expression from the Xi relative to the Xa is greater than 10%, often measured by known expressed polymorphisms in clonal female cells. In cases where there is no allelic information available, the differences between females (Xa/Xi) and males (who have a single Xa) have been used to predict the X-

inactivation status of the gene. Studies using DNAm to predict XCI status also take advantage of this sex difference, as genes escaping from XCI tend to have low levels of promoter DNAm similar to the Xa copy in males, while genes subject to inactivation typically have significantly greater DNAm at promoters (Cotton et al, 2011; Cotton et al, 2014). As this transgenic *Kdm5c* lacked informative polymorphisms, and male to female expression ratios would be complicated by endogenous alleles, wild-type mice of each sex were included for comparisons to the knock-in mice.

Known species differences between escape gene number and distribution on the X chromosome suggests possible regulatory differences between human and mouse, and so the conservation of escape status when human escape genes are integrated into mouse was addressed. Hypothesizing that a BAC containing genes in a subject-escape-subject orientation would provide the best chance of capturing elements and boundaries necessary for escape, a BAC carrying subject gene *ERCC6L*, escape gene *RPS4X*, discordant gene *CITED1*, and the 3' end of subject gene *HDAC8* was chosen. *RPS4X* encodes for ribosomal protein small subunit 4 and is found on autosomes in all vertebrates except mammals, where it lies on the X chromosome in the vicinity of the *XIC* (Hamvas et al, 1992; Brown et al, 1993). *RPS4X* escapes from XCI in a primate-specific manner (Fisher et al, 1990; Jegalian and Page, 1998), and is therefore subject to XCI in mouse (Zinn et al, 1994). The escape of *RPS4X* has been linked to the existence of a functional Y-linked copy, which has been lost in other mammals including mice, likely as a result of the degeneration of the Y-chromosome during evolution

(Omoe and Endo, 1996; Jegalian and Page, 1998; Bellot et al, 2014). The human Y homologue, *RPS4Y*, has been found to be functionally interchangeable with *RPS4X in vitro* by rescuing a mutant hamster cell line phenotype (Watanabe et al, 1993). *CITED1* is labelled as discordant as it has been called both subject (in hybrids) and escape (by promoter DNAm) in different studies, while *ERCC6L* has been found to be consistently subject to XCI (Balaton et al, 2015). In an attempt to delineate the differences between the abilities of human and mouse *RPS4X*/*Rps4x* to escape from XCI, and to narrow down interesting regions in human *RPS4X* BAC for further investigation, potential elements from our DNA elements model were examined (waystations, boundaries and escape elements) between the regions using available data in the UCSC (University of California, Santa Cruz) browser.

2.2 Methods

2.2.1 Construct generation

RP23–391D18 (*Kdm5c*) and RP11–1145H7 (*RPS4X*) BACs (CHORI, BACPAC Resources Center) were each retrofitted using the lambda recombination system (Yu et al, 2000) allowing the addition of *Hprt* homologous recombination targeting arms, which support the integration of constructs into the 5' end of the *Hprt* gene on the mouse X chromosome (Bronson et al, 1996; Yang et al, 2009). PCRs spanning retrofit junctions as well as pulsed-field gel electrophoresis confirmed proper retrofit of construct (Supplementary Table 2.1 for primer information).

2.2.2 Generation of mouse strains

Generation of mouse strains was performed with support from the Simpson Lab and Mouse Animal Production Services. BAC DNA was purified using the Nucleobond XTRA BAC kit (Macherey-Nagel) and linearized with *I-Sce*I (NEB). The BAC constructs were electroporated into male C57BL/6NTac (Taconic, Hudson, NY) ESCs (mEMS6131) carrying the *Hprt*$^{b\text{-}m3}$ deletion (N11 backcrossing from C57BL/6J [The Jackson Laboratory [JAX], Bar Harbor, ME, Stock 002171]) and homozygous for the *Aw-j* agouti allele (N10 backcrossing from B6.129 [JAX, Stock 00051]); with a BTX ECM 630 Electro cell manipulator (BTX). ESC clones were selected in HAT (Gibco) media for reconstitution of the *HPRT*/*Hprt* locus, isolated, and DNA purified. qPCR and PCR with primers spanning approximately every 10 kb along the construct to test intactness were performed for RP23–391D18 and RP11–1145H7 BACs, respectively. Number of integrations was tested using copy-number qPCR assays of BAC backbone regions common to both libraries (Supplementary Table 2.1). Approximately 100 ng of DNA was added to a master mix containing 0.16 μl Maxima Hot Start Taq (Fermentas) with 2 μl 10X buffer and 2 μl 25 mM MgCl2, 1 μl EvaGreen dye (Biotium), 0.16 μl 25 mM dNTPs, 0.2 μl of each 25 μM forward and reverse primers and sterile dH_2O to 20 μl. qPCR was performed in triplicate for each sample using a StepOnePlus Real-Time PCR System (Applied Biosystems) with conditions as follows for all primer sets: 95°C for 5 min; followed by 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 1 min; and a melt curve stage of 95°C for 15 s, 60°C for 1 min and an increase of 0.3°C until 95°C. Testing for multiple Tm peaks for primer specificity, as well as removal of outliers in

triplicate samples were performed using the StepOne software v2.1. Intactness and copy number assays were performed in ESCs prior to microinjection, as well as in N1 and N2 mouse generations. Copy number was analysed using the comparative CT method, normalized to *Hbb-bs* control assay and then to male wild-type controls. ESC derivation and culture was conducted as described previously (Yang et al, 2009). Targeted ESC clones were microinjected into C57BL/6J (JAX, Stock 000664) blastocysts to generate chimeras that were subsequently bred to C57BL/6J females to obtain female germline offspring carrying the BAC insert. The female germline offspring were then bred to C57BL/6J males and backcrossing to C57BL/6J (B6) continued such that mice used in this study were N3 or higher.

The floxed *Xist* strain 129-*Xist*^{tm2Jae} [Mutant Mouse Regional Resource Centre, Chapel Hill, NC, Stock 029172-UNC, Csankovszki et al, 1999] was crossed to the cre-deleter strain 129-*ACTB*^{Cre} (N7 backcrossing from C57BL/6J [JAX, Stock 003376]) to generate females carrying the *Xist* deletion (129-*Xist*^{1lox}/X). The 129-*Xist*^{1lox} strain was maintained by backcrossing to strain 129S1/SvImJ (129) (JAX, Stock 002448, Simpson et al, 1997). Females with the *Xist* deletion were then crossed to males with the BAC construct integrated at the *Hprt* locus (B6-*Hprt*^{BAC}/Y) to generate F1 129-*Xist*^{1lox}/B6-*Hprt*^{BAC} and F1 129-*Xist*^{WT}/B6-*Hprt*^{BAC} females. This *Xist* knockout has been shown to render the X chromosome carrying it unable to inactivate (Gribnau et al, 2005; Yang et al, 2012), thereby resulting in the knock-in X chromosome with an intact *Xist* becoming the Xi. As controls, females with the BAC construct (B6-*Hprt*^{BAC}/X) were crossed to 129 males to

produce F1 B6-*Hprt*^{BAC}/129-Y males. Chi-square tests were performed to assess breeding outcomes of the experimental mice.

2.2.3 Tissue collection and DNA and RNA extraction

Tissue collection was performed by the Simpson Lab. Adult mouse (8-10 weeks) livers, spleens, and brains were macrodissected and flash frozen with liquid nitrogen, then stored at 80°C for no more than six months before processing. For mouse embryonic fibroblasts (MEFs), 13.5 days post-coital embryos were isolated, the head and red organs removed, the remaining embryo individually minced with suction and expulsion using an 18-gauge needle in feeder medium (10% fetal bovine serum in D-MEM) and plated into a T75 flask. Two days following collection, cultures were rinsed with PBS, trypinized and re-plated in their original flasks to achieve maximal cell dispersal and to rid the cultures of debris. Confluency was typically achieved two days after replating and at this point cells from individual embryos were frozen for future expansion. DNA and RNA extraction was performed using TRIzol Reagent (Invitrogen), according to the manufacturer's protocol. A total of 50–100 mg samples of each liver and spleen were used, while an entire sagittal half of brain was homogenized to control for cellular heterogeneity in this tissue. Nucleic acids were quantified by UV spectrophotometry (Ultraspec 2000, Pharmacia Biotech). RNA extractions were diluted to concentrations of 1 μg/μl and treated with 1 μl DNase I with 10 μl buffer (Roche) and 1 μl Ribolock (Thermo Fisher Scientific) in a volume of 50 μl at 37°C for 1 h followed by heat inactivation at 75°C for 10 min.

2.2.4 Expression analysis

For analysis of transcription, 2 µg of DNased RNA extracted from tissues was converted to cDNA using standard reverse transcription conditions with Random Hexamer Primers (Thermo Fisher Scientific) and 200 U M-MLV Reverse Transcriptase (Invitrogen). Reactions were carried out at 42°C for 2 h followed by 5 min incubation at 95°C. RT-qPCR was used to determine relative transcription levels of transgenes compared to stable housekeeping gene *Pgk1* (Boda et al, 2009) in mice carrying the BAC constructs (Supplementary Table 2.1 for primer information). Samples were run in triplicate using a StepOnePlus Real-Time PCR System (Applied Biosystems) with conditions as follows for all primer sets: 95°C for 5 min; followed by 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 1 min; and a melt curve stage of 95°C for 15 s, 60°C for 1 min and an increase of 0.3°C until 95°C. Testing for multiple Tm peaks for primer specificity, as well as removal of outliers from triplicate samples were performed using the StepOne software v2.1. Negative controls of RNA without reverse transcriptase were also run to ensure that the samples contained no DNA contamination. Expression levels were quantified using the comparative CT method and tested for significant differences between groups using the unpaired *t*-test with Welch's correction in GraphPad Prism 5.

2.2.5 DNAm and SNP analyses

Using the EZ DNA Methylation-Gold Kit (Zymo Research), 500 ng of DNA was bisulphite converted following the manufacturer's instructions. Internal bisulphite conversion controls were included in the pyrosequencing assays to monitor complete

conversion of DNA. Each 25 μl pyrosequencing PCR was performed with 10 μl PCR buffer (Qiagen), 0.2 mM dNTPs, 0.125 μl Hot Start Taq DNA polymerase (Qiagen), 0.25 mM forward primer, 0.25 mM reverse primer and 12–35 ng bisulfite-converted DNA. Conditions for PCR were 95°C for 15 min, 50 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min and finally 72°C for 10 min. One forward or reverse primer was biotinylated, depending on which strand contained the target region to be sequenced, to subsequently isolate the strand of interest for pyrosequencing. Template preparation for pyrosequencing was done according to the manufacturer's protocol, using 10–15 μl of PCR products. Capillary dispensing tips were used to dispense the nucleotides for pyrosequencing, using the PyroMark MD machine (Qiagen). Each human promoter assay was tested in at least one mouse sample without the target transgene to ensure the specificity of the human primers. At least three CpGs in an island were evaluated and averaged per assay. Significance was tested using the Mann–Whitney t-test in GraphPad Prism 5. SNP pyrosequencing was performed as above (with annealing conditions of 58.3°C) using primers that amplify a single-nucleotide polymorphism of the *Flna* locus from cDNA of knock-in females without the *Xist* deletion to determine level of skewing by relative expression of the B6 and 129 alleles (Supplementary Table 2.1 for primer information).

2.3 Results

2.3.1 Generation of transgenic mice with BAC knock-ins on the Xi

To ensure that the *Hprt* locus is permissive for escape, BAC RP23–391D18 was chosen as it contains the mouse gene *Kdm5c*, previously shown to escape from XCI at four integration sites in a mouse cell line (Li and Carrel, 2008). This BAC also contains subject (in mouse) genes *Tspyl2*, *Kantr* and *Gpr173*. The choice of human BAC was driven by several criteria as outlined in Figure 2.1. The vector used to retrofit the BACs with homology arms for integration into *Hprt* was designed to work with BACs from the RPCI-11 library (Schmouth et al, 2012). In order to ensure that putative elements were present in the construct, the BAC should contain an escape gene flanked by subject genes, ideally all with broad expression levels and/or CpG island promoters to monitor DNAm. The restriction to small domains biases assessment of escape elements to regions that more closely resemble the single escape genes in mice, as capturing larger human escape domains in a single BAC is not feasible.

Human BAC RP11–1145H7 was selected as it contains multiple human genes with different XCI status calls. The primate specific escape gene *RPS4X* and discordant gene *CITED1* are in the center bounded by subject genes *ERCC6L* and *HDAC8* (truncated), giving confidence that human escape and boundary elements for regulating expression from the Xi were present within the BAC, and could be tested to see if they were recognized by mouse. Selected BACs were targeted by homologous recombination to the $Hprt^{b\text{-}m3}$ deletion on the mouse X chromosome in C57BL/6 (B6)

male ESCs (Figure 2.1). Proper integration was selected for with hypoxanthine aminopterin thymidine (HAT) media, as the BAC constructs contained a complementary sequence that rescues HPRT activity through creation of a chimeric locus consisting of the human *HPRT* promoter and exon 1 and mouse *Hprt* exons 2–9 (Bronson et al, 1996).

Validation of the intactness of the knock-ins utilized assays approximately every 10 kb along the BACs (Supplementary Table 2.1). As the mouse *Kdm5c* BAC is from B6 and was integrated into a B6 background, qPCR assays of genomic DNA were performed and showed single copy integration with no major deletions, both in ESCs prior to blastocyst microinjection, as well as in N2 males (Supplementary Figure 2.1). Human-specific PCR assays approximately every 10 kb along the BAC RP11–1145H7 integration confirmed that the *RPS4X* BAC was intact with no major deletions, both in ESCs prior to blastocyst microinjection as well as in N1 female offspring of chimeras. Copy number qPCR of the BAC backbone vector suggested that two copies of the human BAC had integrated (Supplementary Figure 2.1). Negative PCR assays for the BAC and backbone vector in eight male offspring of chimeras indicated that there was no autosomal transmission of the BAC, and thus both copies were likely linked on the X chromosome. To examine expression of the transgenes from the Xi only, B6 N2 male mice carrying the BAC at *Hprt* were crossed with 129S1/SvImJ (129) females carrying a deletion at the *Xist* gene responsible for initiation of XCI, resulting in experimental female offspring that always carried the BAC on their Xi (129-$Xist^{1lox}$/B6-$Hprt^{BAC}$).

2.3.2 An X-linked mouse gene escapes inactivation at *Hprt*

To determine if an X-linked gene can escape from XCI when integrated at *Hprt*, female mice with BAC RP23–391D18, carrying mouse escape gene *Kdm5c* (129-*Xist*^{1lox}/B6-*Hprt*^{Kdm5c}) on the Xi were analyzed (Figure 2.2A). Knock-in male mice carrying a copy of the BAC on their single Xa (B6-*Hprt*^{Kdm5c}/129-Y), as well as knock-in females with random XCI (129-*Xist*^{WT}/B6-*Hprt*^{Kdm5c}) were also generated and assessed. As the genes on the BAC have endogenous copies, the use of a cross with informative polymorphisms would have improved the sensitivity to detect expression from the integrated allele; however, the purpose was to test the mouse strains designed for integration of human BACs (Schmouth et al, 2012).Therefore wild-type 129/B6 male and female controls were included as a baseline for expression and DNAm. Genes on BAC RP23–391D18 were examined by RT-qPCR (Figure 2.2B) to assess expression in brain, liver and spleen for six mice of each genotype, as escape has been described to vary between individuals and tissues.

Increased expression in knock-in males carrying an additional copy of both *Kdm5c* and *Tspyl2*, compared to wild-type males, was significant in two tissues and provides evidence that the integrated BAC has retained all necessary elements for functional transcription from the X when it is active. To examine escape from XCI, expression from the Xi was analysed in 129-*Xist*^{1lox}/B6-*Hprt*^{Kdm5c} females with two endogenous copies of *Kdm5c* (one Xa and one Xi) plus the transgenic copy on the Xi, relative to wild-type females (one endogenous Xa and Xi copy). Expression of *Kdm5c* from experimentally

skewed transgenic females was significantly higher than wild-type females in liver and spleen only, with a trend toward higher expression in brain. *Tspyl2* has been previously reported to be subject upon random integration into the X (Li and Carrel, 2008) consistent with results in brain and liver; however, there was a significant increase in expression in spleen. Measuring DNAm of genes with CpG islands has been a robust indirect approach to examine XCI status as backup to, or in lieu of, expression analysis, therefore assays were established for the CpG-island promoters of *Kdm5c* and *Tspyl2*. DNAm was also examined at the promoters of the integrated human *HPRT* and the closest mouse endogenous gene to the integration site, *Phf6*, to see if there were any upstream or downstream regulatory effects of the BAC integration. DNAm is shown as an average of at least three CpGs for both endogenous and transgenic alleles where applicable (Figure 2.2C). The transgenic males mirrored the wild-type males in all assays typically showing hypomethylation of the analysed gene, as expected given they were on an Xa. As with expression, transgenic females with the knock-in on the Xi were compared to wild-type females to observe if the additional Xi copy raised (indicating transgene is silenced) or lowered (indicating transgene is expressed) DNAm levels. The *Kdm5c* promoter DNAm in the transgenic females remained low like the wild-type females and both males, reinforcing that all three copies of the gene are capable of expression. High promoter DNAm of neighbouring BAC gene *Tspyl2* in the experimentally skewed knock-in females suggests that the gene is subject to inactivation. *Tspyl2* also generally shows a slight increase in DNAm from wild-type females, which is expected if there is now a second inactive copy raising methylation. *HPRT* and *Phf6* both showed hypermethylation in females similar to gene averages

previously seen at these sites (Yang et al, 2012) suggesting that the elements permitting escape from XCI at the transgenic *Kdm5c* locus are not affecting the surrounding environment. *Kantr* and *Gpr173* were not assessed as they do not have promoter CpG islands and neither has been previously suggested to escape from XCI.

The integration of extra genes could be detrimental to the mouse, and selection against them being expressed on the Xa could lead to non-random XCI in females. Thus, expression of an X-linked SNP was used to assess knock-in females without the *Xist* deletion for deviations from random XCI. There was no consistent skewing of inactivation toward one X being silenced more often than the other (Figure 2.3A). This is in agreement with both expression and DNAm of non-escaping BAC genes *Tspyl2* and *HPRT* where they are now on an Xa approximately half of the time, thereby contributing to generally higher expression levels and lower DNAm compared to the experimentally skewed females. Lack of a negative influence of the transgene was also supported by normal breeding and genotype ratios of the experimental mice (Supplementary Table 2.2). As the expression differences between transgenic and wild-type females were not dramatic, an additional set of four 129-$Xist^{1lox}$/B6-$Hprt^{Kdm5c}$ females were tested separately; however, these were not significantly different in *Kdm5c* expression from wild-type females, although they continued to show hypomethylation at the promoter (Supplementary Figure 2.2), supporting that *Kdm5c* escapes from XCI. *Tspyl2* remained hypermethylated in the additional females, and lost significance in the expression difference from wild-type in spleen.

2.3.3 Human X-linked genes escape inactivation at *Hprt*

Next, the transgenic mice carrying the human BAC RP11–1145H7 with *RPS4X* (Figure 2.4A) were analysed to determine if the BAC carried the necessary elements for escape, and if mouse cells could recognize them. Male mice carrying the knock-in on their Xa (B6-*Hprt*^{RPS4X}/129-Y), knock-in females with random XCI (129-*Xist*^{WT}/B6-*Hprt*^{RPS4X}) and experimentally skewed females carrying the knock-in on their Xi (129-*Xist*^{1lox}/B6-*Hprt*^{RPS4X}) were assessed for gene expression (Figure 2.4B) and DNAm (Figure 2.4C) in brain, liver, and spleen for six mice of each genotype. In all tissues, knock-in males expressed *RPS4X*, demonstrating that the transgene is capable of expression from the Xa. *CITED1* expression was detected from the male Xa in brain, but was not detected at significant levels in liver or spleen. Significant expression of *ERCC6L* was not detected in any of the three tissues initially examined for any genotype; however, expression was detected in MEFs derived from two knock-in males (Figure 2.5A). These expression patterns are consistent with data in the corresponding tissues in human (Supplementary Figure 2.3).

In knock-in females where expression is measured only from the Xi, *RPS4X* and *CITED1* are expressed at approximately half the level of the male Xa in brain, with *RPS4X* also showing similar escape levels in liver and spleen. Knock-in female mice with random XCI displayed higher expression levels than the experimentally skewed female mice (Figure 2.4B). No *ERCC6L* expression was detected from MEFs derived from two experimentally skewed knock-in females (Figure 2.5A). Promoter CpG islands

associated with BAC genes *RPS4X*, *CITED1* and *ERCC6L* were examined, with all three being hypomethylated in males (Figure 2.4C and Figure 2.5B). Xi knock-in females have ongoing hypomethylation of the *RPS4X* promoter in all tissues, which supports that the gene escapes from XCI. *CITED1* DNAm is low in brain where the gene is expressed from the Xi; however, is slightly increased in other tissues. The *ERCC6L* promoter is hypomethylated in males suggesting the transgene is capable of expression as seen in the male MEFs, despite expression not being detectable in the other tissues examined. *ERCC6L* is hypermethylated in Xi knock-in females implying that it is subject to inactivation, concordant with the lack of expression seen in knock-in female MEFs. All mice show *Phf6* and *HPRT* DNAm averages similar to those previously seen at these sites (Yang et al, 2012), demonstrating that escape of *RPS4X* and *CITED1* was not spreading into neighbouring genes on the Xi (Figure 2.5C). The lack of DNAm at *RPS4X*, and consistency in expression levels between experimentally skewed female knock-in mice after multiple generations of breeding, further demonstrates that both copies of the human BAC are on the X chromosome, and both copies of *RPS4X* are escaping inactivation.

Expressed X-linked SNP analysis of knock-in females with random XCI showed no consistent skewing of inactivation toward one X being silenced more often than the other (Figure 2.3B). This is in agreement with both expression and DNAm of non-escaping BAC genes *ERCC6L* and *HPRT* where contribution of an Xa in addition to an Xi in these females presents itself in generally higher expression levels and lower

DNAm compared to the experimentally skewed females. Lack of a detrimental influence of the transgene was also supported by normal breeding and genotype ratios of the experimental mice (Supplementary Table 2.2).

2.3.4 Elements regulating escape of *RPS4X* and *CITED1* from XCI

A broad array of DNA elements is proposed to play a role in the ability of a gene to escape from XCI (depletion of waystations, enrichment of boundary elements between active and inactive domains, and presence of escape elements), as well as factors involved in chromatin ultrastructure. Considering this, available datasets were used to try and demarcate regions that could support the escape and silenced profiles on the human *RPS4X* BAC by profiling repetitive content of the transgenes, patterns of transcription factor binding and contact domain boundaries (Figure 2.6A). Colocalization of boundary factor CTCF, cohesion components SMC3 and RAD21, as well as YY1 potentially mark a boundary between subject gene *ERCC6L* and escape gene *RPS4X*, however, similar sites exist in the corresponding region in mouse (Figure 2.6B), which does not escape from XCI. Additionally, kilobase-resolution Hi-C data of the BAC region in its endogenous location in GM12878 reveals a contact domain boundary between *ERCC6L* and *RPS4X*, although in several other female cell lines examined at 5 kb resolution the end of the domain sometimes shifts from downstream of *RPS4X* to end between *CITED1* and *RPS4X* so that the two genes are not always in the same domain (Rao et al, 2014). It is possible that such a shift in boundaries is responsible for the

variability seen in *CITED1* DNAm and previously recorded discordance in escape status between studies.

2.4 Discussion

Previous studies have demonstrated that *Kdm5c* harbours an intrinsic escape element (Li and Carrel, 2008; Horvath et al, 2013). As those studies examined random integrations on the Xi, they could not be recapitulated for this assessment of DNA elements. Instead the *Hprt* docking site was used, previously suggested to support expression from the Xi (Yang et al, 2012). Using the *Kdm5c* BAC examined by others (Li and Carrel, 2008; Horvath et al, 2013), and a combination of direct expression and indirect DNAm assays, it was established that the integration site at the 5' end of *Hprt* can support escape gene expression. The *Kdm5c* transgene showed a range of expression from the Xi across three examined tissues, which is similar to what has been previously reported. *Kdm5c* expression levels have been shown to vary widely between mouse strains, tissue type, cell type and developmental stage, with Xi expression relative to the Xa ranging from 20–70% (*in vivo*: Carrel et al, 1996; Sheardown et al, 1996; Berletch et al, 2015; *in vitro*: reviewed in Marks et al, 2015). Interestingly, a study has shown that there is a significant sex difference in *Kdm5c* expression in adult mouse brain biased toward females that is not compensated for by the Y-linked paralogue *Kdm5d* (Xu et al, 2008). Authors of this study did not see this sex differences in adult liver but there was noticeable variation among females, possibly caused by individual differences between mice in transcription of the escape allele. The expression patterns

in the wild-type male and female controls in Figure 2.2B brain and liver agree with the described data for these two tissues in mice. As *KDM5C* has been linked to intellectual disability disorders in humans (Jensen et al, 2004; Tzschach et al, 2006) and mouse demonstrates there may be sexually dimorphic regulation of *Kdm5c* gene expression in brain (Xu et al, 2008), the third copy of the *Kdm5c* gene the transgenic female brains may also be more tightly regulated as this is where the lowest (and non-significant) level of additional Xi expression was seen.

Having confirmed the utility of the docking site with the *Kdm5c* BAC, the *Hprt* site was further used to examine a human escape domain containing *RPS4X* and *CITED1*. *RPS4X* is a broadly expressed (Supplementary Figure 2.3) primate-specific escape gene with Y homology retained in human but not mouse (Bellot et al, 2014), which may have led to the loss of a drive for extra gene dosage in mouse and subsequent lack of conservation of the DNA elements necessary for escape (Park et al, 2010). However, the mouse is able to recognize the intrinsic element(s) present at the human gene, demonstrating likely conservation of the element(s), or at a minimum, conservation of the machinery recognizing the element(s). Examination of NPC clones suggests that *Rps4x*, as well as neighbouring genes *Ercc6l* and *Cited1*, are occasionally capable of variable escape in some mouse cell types; however this could be due to epigenetic instability in cell culture leading to lack of maintenance of their initial inactivation (Marks et al 2015; Giorgetti et al, 2016). *ERCC6L* has been well-characterized as subject in humans (Balaton et al, 2015), in agreement with the lack of expression in

experimentally skewed knock-in female MEFs and promoter hypermethylation across all tissues. *ERCC6L* is not well expressed in brain, liver or spleen in humans (Supplementary Figure 2.3) which may be why expression was unable to be detected in the core dataset of tissues. *CITED1* has previously been called escape by DNAm analysis (Cotton et al, 2014) but subject by expression in Xi somatic cell hybrids (Carrel and Willard, 2005). *CITED1* shows tissue-specific expression in humans in testis, hypothalamus, and pituitary (Supplementary Figure 2.3), which agrees with the ability to detect expression in mouse brain only. Expression of *CITED1* from the Xi along with a hypomethylated promoter in brain gives confidence that it escapes in this tissue. Interestingly, *CITED1* DNAm increases in tissues where it is not expressed; however, it is still in the range where we call escape, although there is not a clear threshold of DNAm at which genes become subject to XCI. A previous X-chromosome-wide DNAm study generally called genes as escape if, in females with random XCI, they had 0–20% DNAm and a difference of less than 10% from males, yet left an uncallable zone (20-30%) between these gene and the subject genes that generally had >30% methylation (Cotton et al, 2014). By that metric *CITED1* would be called an escape gene in liver and uncallable in spleen. It is, however, important to note that gene silencing of *HPRT* has been observed with DNAm as low as 13% at its promoter in female mice with random XCI (Yang et al, 2012).

The integration and recapitulation of escape gene expression from the human BAC in mouse demonstrates that escape elements have a conserved recognition and

mechanism across species, and putative escape elements must lie within ~112 kb for *RPS4X* from the subject *ERCC6L* promoter to the end of the BAC (Figure 2.6). Both BACs must also carry boundaries to contain the open state, as spread of hypomethylation was not seen into *Phf6* or *HPRT*, previously detected when an autosomal transgene escaped from XCI (Yang et al, 2012). CTCF binding has been hypothesized to be involved in setting boundaries around escapees, with differential binding responsible for creating larger domains, and shifting the binding site of CTCF in different cell lines potentially adjusting which genes in a region are escaping. In both humans and mice the Xi forms a distinctive superdomain structure rather than the topologically associated domains (TADs) that are observed across the Xa and the autosomes (Deng et al, 2015; Giorgetti et al, 2016), yet evidence suggests that continued transcription of genes escaping from XCI along with binding of factors such as CTCF may enable the maintenance or re-creation of TAD structures around regions containing escape genes (Giorgetti et al, 2016). The clustering of human escape genes is suggestive of domain-based regulation and so *RPS4X* and *CITED1* may be under control of the same element; however, there is growing evidence for promoter-proximal elements being involved in escape (Calabrese et al, 2012; Mugford et al, 2014; Giorgetti et al, 2016). *In silico* analysis of the TSSs of escape genes in humans has found significant over-representations of YY1 transcription factor binding motif and ChIP-seq peaks, and similar to CTCF, YY1 occupancy is significantly biased toward the Xi at loci that are frequent contacts of Xi-specific superloops (Chen et al, 2016). Additionally, components of the cohesin complex (RAD21, SMC3) have been found to co-localize with CTCF at these sites in both human and mouse female cells (Yang et al, 2015),

again highlighting that there is likely a structural component to escape. Previous studies on the X and X; autosome translocations have found correlations between repetitive content and whether genes are subject to or escaping from XCI. Using large windows (50–100 kb) around the TSSs or full transcribed regions of genes (Wang et al, 2006; Tannan et al, 2013) as well as more promoter-centric methods (Cotton et al, 2013), LINEs are seen to be significantly enriched in regions surrounding genes that are subject to inactivation, while Alu repetitive elements and short motifs were significantly enriched in those that escape inactivation. Given that the escape genes on the BAC are small and located in close proximity to subject genes, a more promoter-centric approach was undertaken to determine if the Alu enrichment, LINE-depletion pattern holds for the escape genes *RPS4X* and *CITED1* compared to subject gene *ERCC6L*. Little difference was found in content between the three genes that would support using these sequences alone for predicting escape within the BAC (Supplementary Table 2.3).

Importantly, despite a lack of conservation in number and distribution of genes that escape from XCI, this work has demonstrated the ability of mouse to recognize human elements regulating escape from XCI, at least for this subject-escape-subject region, thereby providing a model system for the exploration of these elements.

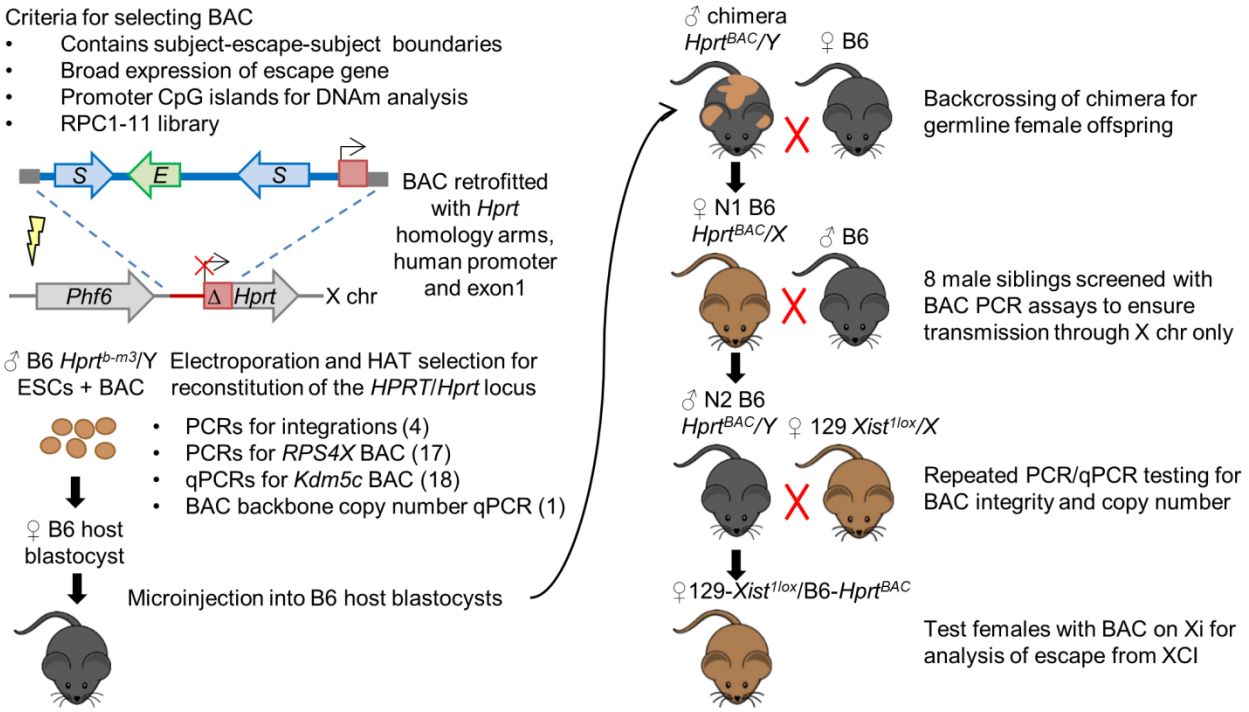


Figure 2.1 Generation of transgenic mice with BAC knock-ins on the Xi

Flow diagram showing criteria for selecting BACs containing genes that escape from XCI, followed by the breeding scheme after electroporation of the BAC into male mouse ESCs. Clones with successful integration at the *Hprt* locus are selected for with HAT media, followed by PCR and qPCR screening. Positive clones are microinjected into host blastocysts. ESCs have the *Aw–J* allele to allow us to follow coat colour and choose appropriate chimera offspring where the ESCs containing the BAC have gone germline. Only N1 females carry the BAC as it is transmitted on the X chromosome. N1 male siblings were screened by PCR assays for the BAC to ensure there was no autosomal transmission and did not detect any bands. BACs were again tested for intactness and copy number at the N1 and N2 stages, before breeding N2 male mice with the BAC on their X to 129-*Xist*^{1lox} females. Females from this cross could inherit either X from their mother, generating our test females with the BAC always on the Xi as well as knock-in females with random XCI.

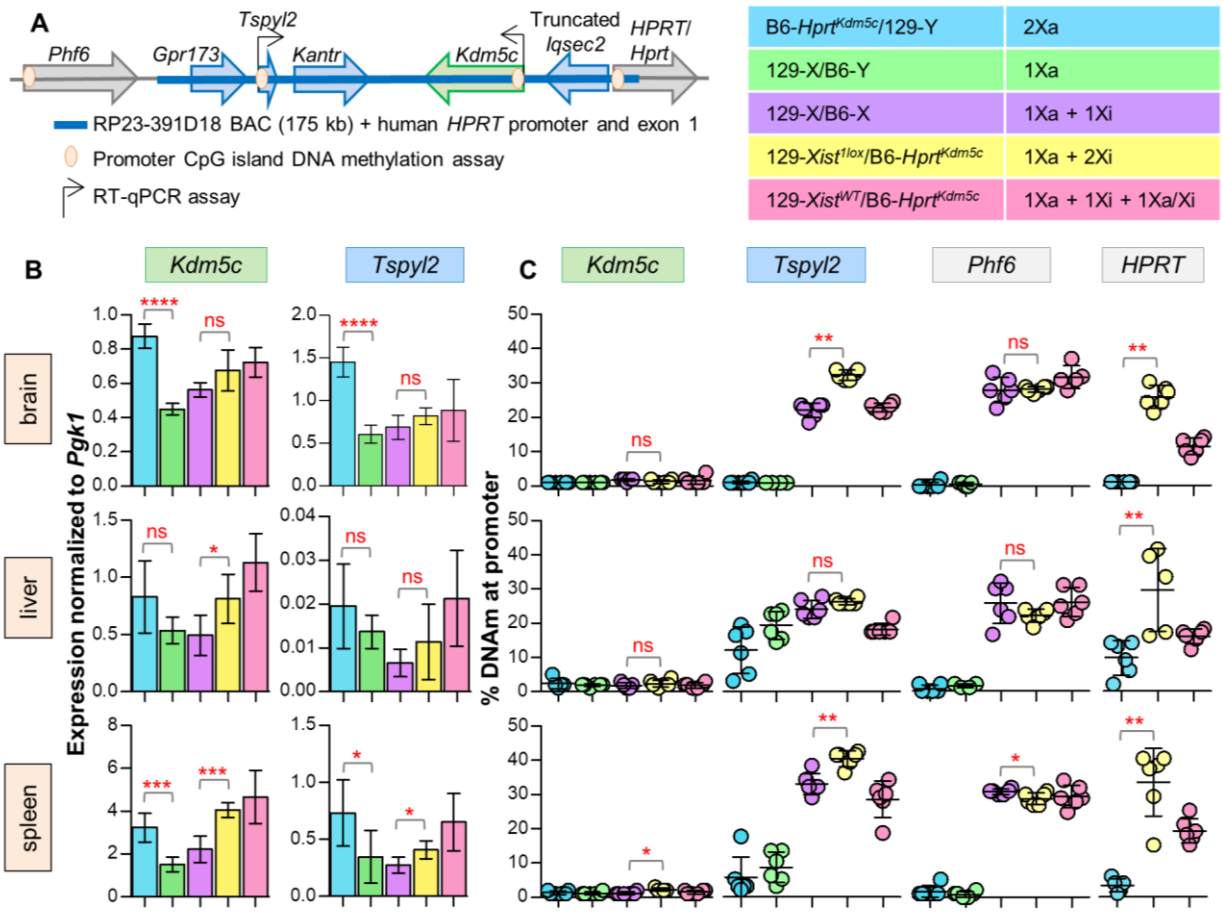


Figure 2.2 Analysis of BAC RP23–391D18 shows *Kdm5c* escapes XCI at *Hprt*

A) Integration of *Kdm5c* BAC at *Hprt*; genes on the BAC expected to escape from XCI in green, subject in blue and genes at integration site (both known to be subject to XCI) in grey. Genes with RT-qPCR and DNAm assays are indicated. **B)** Normalized to *Pgk1*, RT-qPCR of both *Kdm5c* and *Tspyl2* expression in knock-in males (blue) shows significantly more expression than wild-type males (green) in brain and spleen demonstrating that our transgene is expressed on an Xa. *Kdm5c* expression from the non-random Xi in knock-in females (yellow) is higher than wild-type females (purple) suggesting escape of the transgene, but does not reach significance in brain (unpaired t-test). Expression of knock-in females with random XCI (pink) is generally higher than non-random females (yellow). *Tspyl2* expression from the Xi was significant in spleen only but not supported by DNAm. **C)** Average DNAm of *Kdm5c* shows promoter hypomethylation in knock-in females (yellow), which supports the expression trend of

escape from the Xi. *Tspyl2*, *HPRT* and *Phf6* show hypermethylated promoters in knock-in females suggesting they are subject to XCI. Knock-in females are compared to wild-type females for all assays except *HPRT* as wild-type females do not carry the human gene (n=6 mice for each genotype, Mann–Whitney t-test, significance denoted by asterisks; P-value <0.001***, 0.001–0.01**, 0.01–0.05*, >0.05 ns).

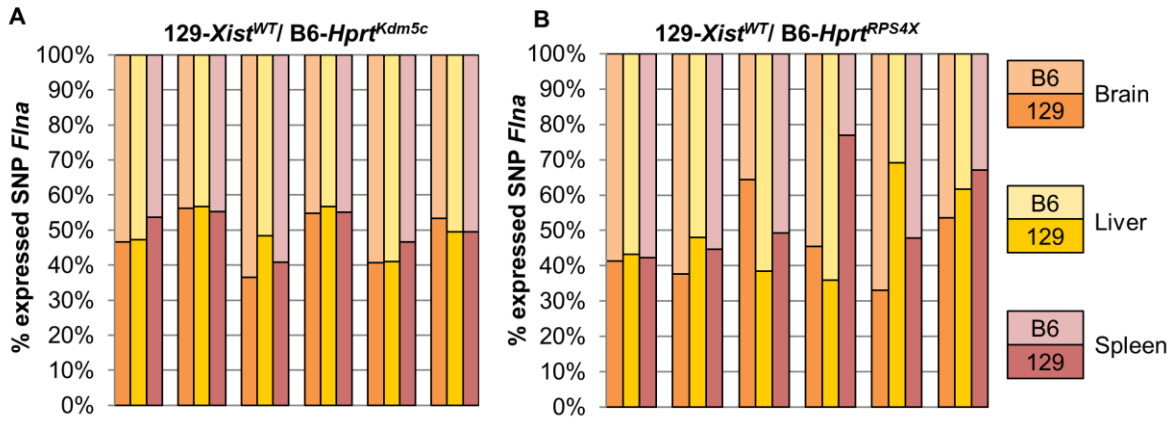


Figure 2.3 *Xist*^{WT} females do not have consistently skewed XCI

Percent allelic expression of X-linked gene *Flna* (subject to inactivation) in **A)** six female mice carrying *Kdm5c* transgene and **B)** six female mice carrying *RPS4X* transgene, all with wild-type *Xist* and random XCI. All three tissues are shown for each mouse with the percent expression indicating how often the B6 or 129 X is the Xa.

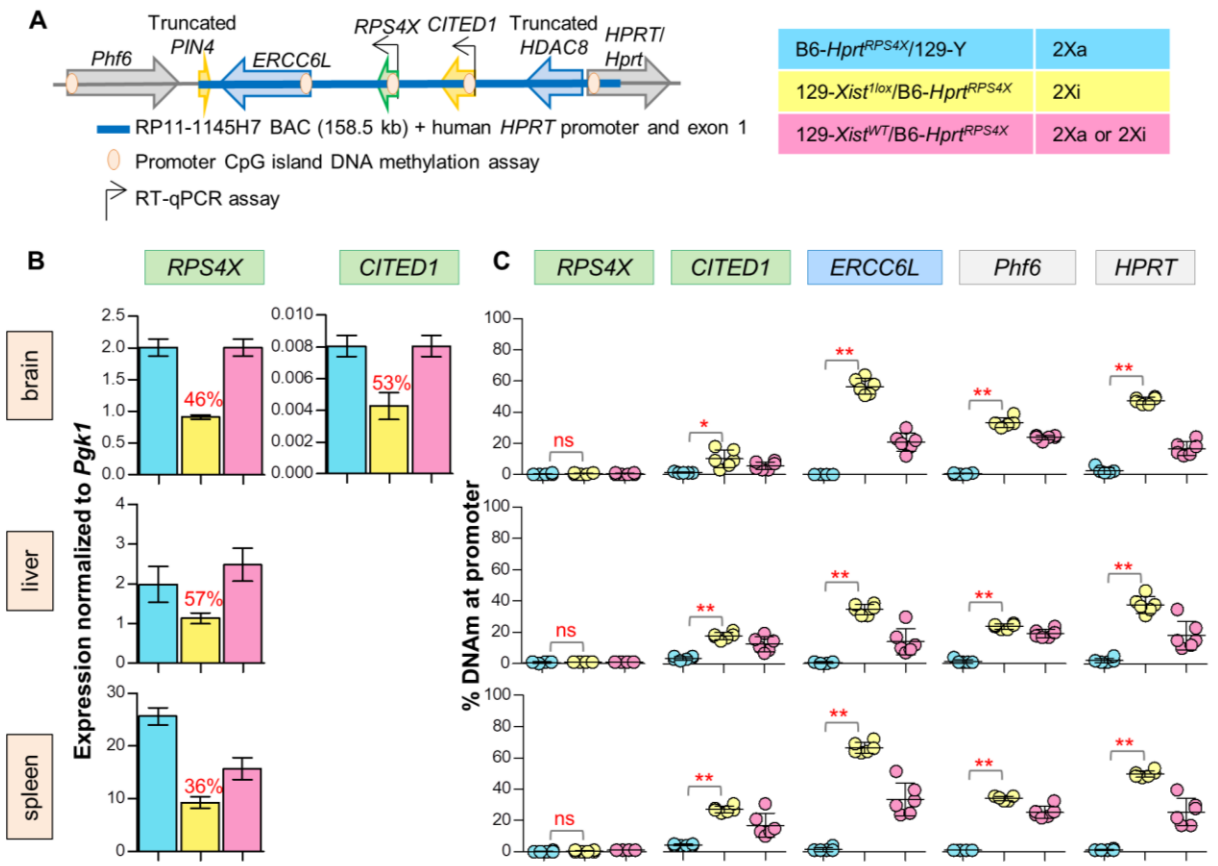


Figure 2.4 Analysis of BAC RP11–1145H7 shows *RPS4X* and *CITED1* escape XCI at *Hprt*

A) Integration of the *RPS4X* BAC at *Hprt*; genes on the BAC expected to escape from XCI in green, subject in blue, variable and discordant in yellow, and genes at integration site (both known to be subject to XCI) in grey. Genes with RT-qPCR and DNAm assays are indicated. **B)** Normalized to *Pgk1*, RT-qPCR of *RPS4X* expression in brain shows that the transgene is active on a male X, and escapes inactivation at ~50% when on the Xi in brain and liver, with a slightly lower level in spleen. An adjacent discordant gene *CITED1* also escapes from the Xi in females at similar levels in brain. *CITED1* expression was not detected in liver and spleen. Expression from female Xi is shown as percentage of the male X (red text). **C)** Average DNAm of skewed knock-in females shows a hypomethylated *RPS4X* promoter in all tissues, *CITED1* is low but different between tissues, and *ERCC6L*, *HPRT* and *Phf6* are significantly hypermethylated compared to knock-in males in all tissues (n=6 mice for each genotype, Mann–Whitney

t-test, significance from hypomethylated males denoted by asterisks; P-value <0.001***, 0.001–0.01**, 0.01–0.05*, >0.05 ns).

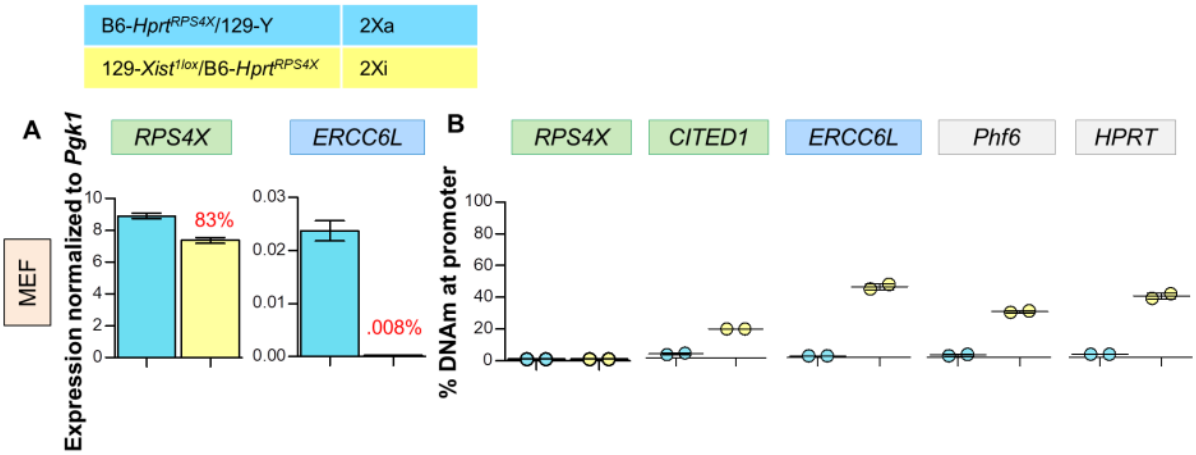


Figure 2.5 *ERCC6L* is subject to XCI in MEFs

A) Normalized to *Pgk1*, RT-qPCR of *RPS4X* and *ERCC6L* expression shows that the transgenes are both active on a male X in MEFs. When on the Xi in MEFs, *RPS4X* escapes at high levels while *ERCC6L* lacks expression in these females indicating it is subject to XCI. Expression from female Xi is shown as percentage of the male X (red text). **B)** Average DNAm of skewed knock-in females shows a hypomethylated *RPS4X* promoter, *CITED1* in the uncallable range, and *ERCC6L, HPRT* and *Phf6* hypermethylated compared to knock-in males. While not statistically significant due to low sample size (n=2 for each genotype), the data is similar to other tissues examined and supports *ERCC6L* being subject to XCI in mouse.

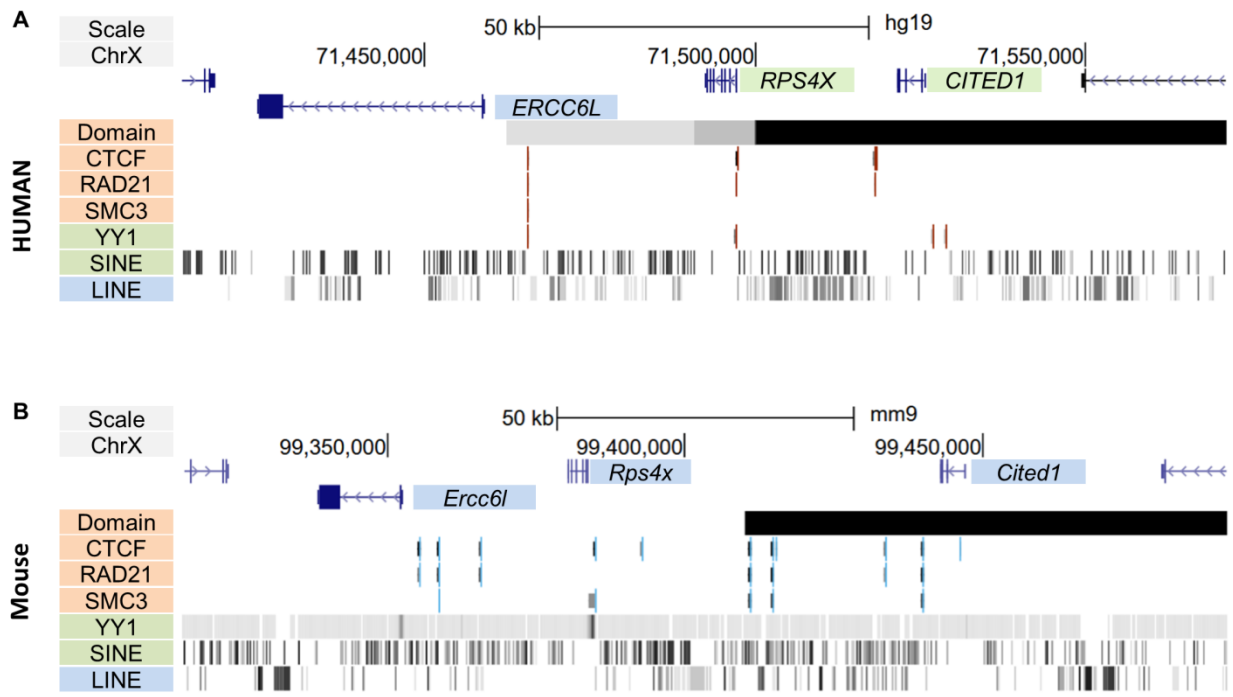


Figure 2.6 The *RPS4X* BAC contains elements that may contribute to escape from XCI

A) Using available datasets, potential elements were examined that may aid in setting up an escape domain around human *RPS4X* and *CITED1* (shaded green for escape) displayed in the UCSC browser (hg19 assembly). Factors associated with escape from XCI as well as components of structural complexes to form boundaries line up between silenced gene *ERCC6L* and escape gene *RPS4X* in female cell line GM12878. In GM12878 as well as several other female cell lines, a contact domain boundary is located within the region contained on the BAC, although the end shifts depending on cell line and resolution to either include *RPS4X* and *CITED1* in the same domain, or end between the two (grey to black, Rao et al, 2014). **B)** The corresponding region in mouse is subject to inactivation yet retains similar elements to the human region. Transcription factor binding sites as well as contact domain (black, Rao et al, 2014) are from CH12 cells in the UCSC browser (mm9 assembly), with the exception of YY1 binding which was done in ESCs (Sigova et al, 2015). Truncated genes *HDAC8* and *PIN4* not labelled.

Chapter 3: XCI status of human transgenes at *Hprt* is stable across development

3.1 Introduction

RPS4X is a primate-specific escape gene with a functional Y homologue (Watanabe et al, 1993; Jegalian and Page, 1998), although *RPS4Y* shows lower expression and protein contribution than the X-linked homologue (Zinn et al, 1994). Mouse lacks both escape from inactivation for *Rps4x* and a functional *Rps4y*, similar to many eutherian (Jegalian and Page, 1998). With the integration of a human BAC into the mouse X-linked *Hprt* gene, it was demonstrated that transgenes *RPS4X* and *CITED1* continued to be expressed on the Xi in 8 –10 week old mice (Peeters et al, 2018), despite the mouse *Rps4x* gene normally being subject to XCI. Thus, this region contains the necessary elements to be recognized as an escapee across species. The *RPS4X* model also provides the opportunity to address two long-standing questions of escapee biology across development. This has been challenging to address in humans for logistical and ethical reasons, leading to most surveys that identify escape genes having done so using adult somatic cells.

First, is escape from XCI actually the result of reactivation of the gene early in development? Evidence supports that escapees could include both genes that are resistant to initial silencing (and then silenced in tissue-specific situations later on), or genes that only reactivate in tissues where they escape following an initial global

inactivation. An early study on *Kdm5c* in mouse suggested that the gene was silenced during the preimplantation stage of development before it was subsequently expressed from the Xi later in development (Lingenfelter et al, 1998). More recent RNA-sequencing experiments from imprinted (paternal Xi) inactivation *in vivo* (embryonic day (E)1.5 to E3.5) have allowed the examination of additional mouse escapees and suggested the majority of constitutive escape genes, including *Kdm5c*, are continuously expressed from the Xi (Patrat et al, 2009; Borensztein et al, 2017). Constitutive escape genes also appear to remain active throughout NPC differentiation *in vitro* while facultative escapees have been reported to both be silenced and then reactivate, or to not inactivate at all during time points analysed, which could have missed periods of reactivation (Marks et al, 2015; Giorgetti et al, 2016). Differing results in studies performed *in vitro* are impacted by cellular heterogeneity during XCI, which can be influenced by the method of differentiation which often varies between studies and correlates to different time points in embryogenesis.

The second developmentally relevant question that can be addressed in the *RPS4X* model is if inactivation is a source of variability with aging, due to either reactivation of X-linked genes, or silencing of genes that normally escape (or variably escape) from XCI, either of which could have phenotypic consequences. Limited evidence can be found in human somatic cell lines for a few genes that may reactivate (Anderson and Brown, 2005; Kucera et al, 2011); however, epigenetic instability in culture may preclude these results from translating to humans. Early studies of select human genes

failed to identify evidence of aging-related reactivation (Migeon et al, 1998). In mouse, the *Otc* gene reactivates with age (Wareham et al, 1987), and drug treatment of somatic cells to induce reactivation shows that some genes seem to be more prone to reactivation than others (Minajigi et al, 2015). Reactivation is also observed in some human cancers (Chaligné et al, 2015), and has been suggested to be a predisposition of female lymphocytes leading them to overexpress immunity-related genes (Wang et al, 2016). Genes that are prone to reactivate with age could be contributing to variable escape genes observed between tissues, the stochastic cellular heterogeneity observed in single-cell RNA-seq, and the increased frequency of escape genes in cultured cells. Overall evidence suggests that a majority of escape genes seem to avoid the inactivation process rather than fail to maintain silencing, but as this topic clearly warrants further exploration, the stability of XCI statuses of the human genes on the *RPS4X* BAC were analyzed as early as E9.5 up to one year of age. After seeing stability of *RPS4X* over-expression throughout development but no obvious phenotype, the potential of transcriptional or posttranscriptional regulation was investigated as altered protein levels of RPS4X and CITED1 have been reported in cancer (Prasad et al, 2004; Tsofack et al, 2013; Xia et al, 2018).

3.2 Methods

3.2.1 Transgenic mouse model generation

Creation and breeding of the *RPS4X* knock-in mouse as described in 2.2.2.

3.2.2 Tissue collection and DNA and RNA extraction

Tissue collection was performed by Simpson Lab. E9.5 embryos and one year old livers, spleens, and brains were macrodissected and flash frozen with liquid nitrogen, and then stored at −80°C for up to nine months before processing. For MEFs, E13.5 embryos were isolated, the head and red organs removed, the remaining embryo individually minced with suction and expulsion using an 18-gauge needle in feeder medium (10% fetal bovine serum in D-MEM) and plated into a T75 flask. Two days following collection, cultures were rinsed with PBS, trypsinized and re-plated in their original flasks to achieve maximal cell dispersal and to rid the cultures of debris. Confluency was typically achieved two days after re-plating and at this point cells from individual embryos were frozen for future expansion. DNA and RNA extraction was performed as described in 2.2.3.

3.2.3 Expression, DNAm and SNP analyses

As described in 2.2.4 and 2.2.5, additional primer information in Supplementary Table 3.1.

3.2.4 Western blotting

Performed by S Baldry, Brown Lab. RIPA protein extracts (Abcam RIPA protocol done with an additional sonication step 3x [30 s on, 30 s off] on ice after cells resuspended in RIPA buffer) from cell pellets were quantitated by Bradford assay (Fermentas).

Specified amounts were run on 12% MiniProtean TGX precast protein gels (BioRad) at 200 V for 50 min and then transferred to Immobilon PVDF (Immobilon-FL, IPFL10100) using ethanolamine transfer, 30 V for 75 min at 4°C. Blots were stained with REVERT total protein stain (Licor), imaged, and staining reversed as per Licor protocol. Blots were blocked for 60 min at room temperature with agitation in 1% BSA in PBS-Tween, and then incubated with primary antibodies in PBS-Tween for 60 min at room temperature with agitation. RPS4X antibody (Thermo Fisher Scientific) is shown; Proteintech 14799-1-AD gave comparable results both used at 1/750 dilution, Pierce anti-Beta Actin loading control monoclonal antibody (Thermo Fisher Scientific) used at 1/1000 dilution. Blots were washed then incubated for 60 min at room temperature with agitation with fluorescently labelled secondary antibodies (Licor IRDye 680RD Goat anti-Rabbit IgG (H+L) and IRDye 800CW Goat anti-Mouse IgG (H+L)) diluted 1/20 000 in PBS-Tween+0.01% SDS. Blots were washed twice in PBS-Tween, then once in PBS and imaged on a Licor Odyssey imager. Resulting images were analysed with the Odyssey Application software to quantitate fluorescent signal with median all sides background setting. For the REVERT signal, a box was drawn around the total protein signal from ~55 kDa to ~25 kDa using Fermentas PageRuler prestained protein ladder as a size reference and median right/left background setting was used.

3.3 Results

3.3.1 Early *RPS4X* expression in mice

The BAC contained three genes that had been shown to recapitulate their human XCI statuses in an 8-10 week mouse knock-in (Peeters et al, 2018) with *RPS4X* and *CITED1* escaping XCI and *ERRC6L* inactivating (Figure 3.1A). To address whether the escape from XCI that was observed reflects an early reactivation of the transgene, transgenic mouse embryos were dissected at E9.5. *RPS4X* showed clear gene expression from the Xi at this early embryonic stage, with levels detected by RT-qPCR approximately 40% of the level seen from the Xa in male embryos (Figure 3.1B). No expression was observed for the tissue-specific *CITED1* and *ERRC6L* genes, as at E9.5 individual tissues could not be dissected. DNAm was assessed using pyrosequencing to examine an average of at least three CpG sites in the CpG island promoters for *RPS4X*, *CITED1* and *ERRC6L* as well as the flanking *HPRT* human/mouse chimaeric gene and the mouse *Phf6* gene (Figure 3.1C). *RPS4X* showed almost no DNAm, consistent with escape from XCI. The *ERRC6L* promoter showed significant DNAm on the Xi relative to the low DNAm on the Xa, consistent with this gene being subject to XCI. For *CITED1*, DNAm was significantly greater in the female than the male, and similar to what was observed previously for the adult liver and spleen. The presence of DNAm is consistent with *CITED1* being a tissue-specific escape gene, and suggests that the gene is initially silenced, at least in the majority of the embryonic cells.

3.3.2 Maintenance of inactivation status at one year timepoint

The BAC knock-in mice also provided an opportunity to address stability of silencing and escape with aging. The mice were aged for one year and tissues were isolated for the DNAm and expression assays described above to address whether the genes were maintaining their XCI statuses (Figure 3.1D, E). *RPS4X* continued to be expressed from both the Xa and Xi, and was never observed to acquire DNAm, consistent with ongoing escape from XCI. *CITED1* was only expressed in brain, and similar to what was observed in 8-10 week old adult mice, the one year old mice showed ongoing expression from the Xi. In the 8-10 week old mice the *CITED1* promoter showed approximately 10% DNAm, still within the range of DNAm seen for genes that escape XCI (Cotton et al, 2014). Interestingly, despite ongoing expression in the one year brain, the DNAm had risen to ~20%, comparable with the liver where expression was not detected. In the liver and spleen where *CITED1* was not expressed, elevated DNAm was observed in female, but not male. The *ERCC6L* gene was only expressed in MEFs (Peeters et al, 2018), and thus not surprisingly, no expression was seen at one year in the tissues analyzed, suggesting no tissues lost silencing of this gene. The *ERCC6L* DNAm was consistently high across tissues and, along with the flanking *HPRT* and *Phf6* genes, showed no evidence for loss in DNAm over the year. In fact, methylation was generally higher than observed for the E9.5 embryo. The transgene is solely on the Xi in the 129-*Xist*^{1lox}/B6-*Hprt*^{RPS4X} female mice, whereas in the 129-*Xist*^{wt}/B6-*Hprt*^{RPS4X} female mice, natural skewing of XCI could contribute to altered levels of DNAm. Therefore, expression of a gene that is polymorphic between the B6 and 129 X chromosomes was examined to determine levels of skewing (Figure 3.2). There was a trend towards

inactivating the B6 (transgenic) X chromosome in brain and spleen, with less skewing observed in liver. Overall, no change in XCI status was observed for the genes with age.

3.3.3 Consequences of *RPS4X* expression in mice

Despite transcription of the *RPS4X* gene from both the Xa (in males and females) and the Xi (in females), no overt phenotype was detectable in the mice, nor was there any transmission distortion observed in the initial analysis (Peeters et al, 2018). The mice were aged for one year to determine if there were any long-term consequences to *RPS4X* expression, and again observed no obvious phenotype in males or females. Furthermore, the mice have been bred to homozygosity and remain viable. Using MEFs, the transcription of both the mouse *Rps4x* and the human (transgenic) *RPS4X* genes were examined (Figure 3.3A). There was no observable reduction in the amount of mouse *Rps4x* transcription due to the presence of the human *RPS4X* transcription. Post-transcriptional compensation, however, might regulate the dosage of protein.

To examine the level of protein in the presence of the human BAC knock-in, western blotting was used with antibodies to RPS4X in male lines with or without the transgene (Figure 3.3B). To control for protein amounts, a total protein stain (REVERT) was used. Similar results were observed comparing to ACTB, or using a different RPS4X antibody. Inter-gel variability precluded comparisons across all samples; however, within a gel,

quadruplicate loadings of extracts from MEFs derived from a B6129F1-*Hprt*^{RPS4X} male knock-in mouse (KI) or a B6129F1 line without the BAC knock-in (WT) were compared. No significant difference (unpaired *t*-test) was observed for two independent pairings of WT male versus KI male.

3.4 Discussion

It has been demonstrated that the sequences instructive for escape from XCI are recognized across species (Peeters et al, 2018) despite differences in timing of inactivation between humans and mice (Marks et al, 2015; Chen et al, 2016). The same BAC knock-in model was used here to demonstrate that for *RPS4X,* escape from XCI occurs early (by E9.5), so is unlikely to be reactivation, and is stable across lifespan of the mouse up to one year. Global studies of inactivation timing *in vivo* suggesting that constitutive escape genes remain active have been performed in trophoblast cells with imprinted XCI, which takes occurs around E1.5 and reactivates at the early blastocyst stage (by E4.5), after which random XCI takes place (by E6.5). As E9.5 was measured here, it is possible that *RPS4X* was transiently inactivated at E6.5 and a reactivation window was missed, but as the timing of the second wave of random XCI is mainly studied by *in vitro* differentiation of female ESCs subject to variability in protocols, it is challenging to draw conclusions across studies.

For the adjacent *CITED1* gene, DNAm is detectable by E9.5. However, expression is observed later in life from the Xi in brain, suggesting tissue-specific reactivation may occur despite the presence of significantly more DNAm on the Xi than was observed on the Xa. While *CITED1* escape in one year brain is still greater than 10% of the Xa expression, it is less than what was seen in adult brain previously analysed, and correspondingly its promoter DNAm had also increased. While DNAm has been successfully used to identify genes escaping from XCI, an uncallable zone does exist between the most robust escape and subject genes where expression information is necessary to make confident calls (Cotton et al, 2014). Additionally, location and number of CpGs evaluated within the promoter island likely cause variability in DNAm measured for some genes.

Many of the genes retaining X/Y homology are escapees, and they have been argued to escape from XCI because they have extreme dosage-sensitivity and thus were selected to maintain Y and X expression (Naqvi et al, 2018). Such dosage-sensitivity has been attributed to critical gene regulatory roles (Bellot et al, 2014) or to genes that participate in multi-protein complexes (Pessia et al, 2012). The lack of increased protein, despite augmented RNA levels resulting from the human knock-in as well as the endogenous mouse *Rps4x* gene, is consistent with members of multi-protein complexes being more likely to undergo post-translational dosage compensation (Ishikawa et al, 2017). This compensation may underlie absence of a phenotype from the augmented *RPS4X* transcription in this mouse model.

Overall, having demonstrated that escape from XCI of human-specific escapees can be recapitulated in the mouse, allowing utilization of mouse to examine escapee biology, this mouse model has provided access to developmental timepoints confirming that *RPS4X* constitutively escapes from XCI across the lifespan.

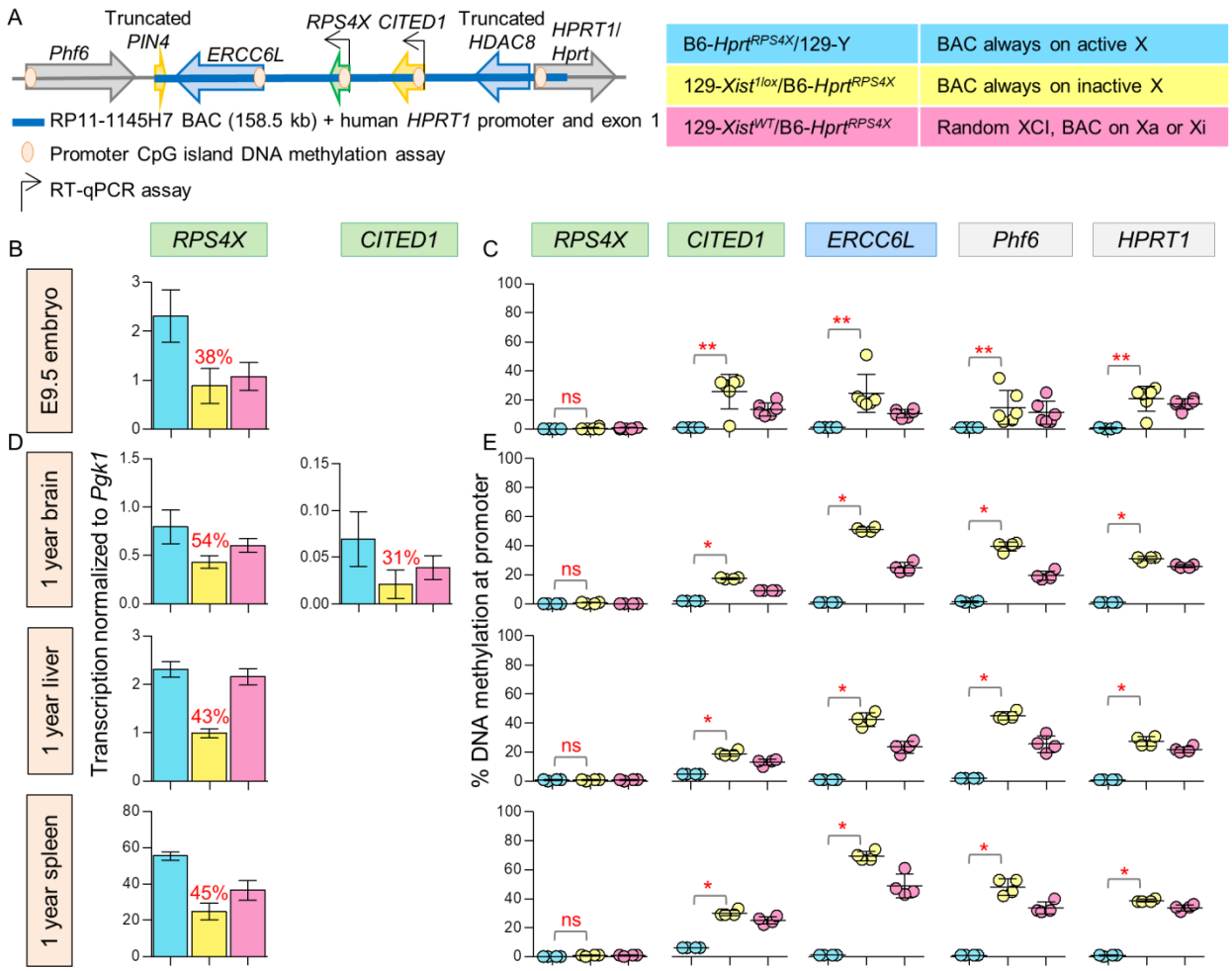


Figure 3.1 Analysis of BAC RP11–1145H7 shows *RPS4X* (and *CITED1* in brain) escape XCI at *Hprt* across lifespan

A) Integration of the *RPS4X* BAC at *Hprt*; genes on the BAC expected to escape from XCI in green, subject in blue, variable in yellow, and genes at integration site (both known to be subject to XCI) in grey. Genes with RT-qPCR and DNAm assays are indicated. Six mice for each genotype (transgenic male, transgenic female with completely skewed XCI, and transgenic female with random XCI) were analyzed at E9.5, and four mice for each genotype were analyzed at one year. **B)** Normalized to *Pgk1*, RT-qPCR of *RPS4X* expression at E9.5 shows that the transgene is active on a male X, and escapes inactivation at ~38% when on the Xi in females. Expression for an adjacent variable gene *CITED1* was not detected. Expression from female Xi is shown as percentage of the male X (red text). **C)** Average DNAm of skewed knock-in females

shows a hypomethylated *RPS4X* promoter, with *CITED1*, *ERCC6L, HPRT* and *Phf6* all significantly hypermethylated compared to knock-in males (Mann–Whitney *t*-test, significance from hypomethylated males denoted by asterisks; *P*-value <0.001***, 0.001–0.01**, 0.01–0.05*, >0.05 ns). **D**) Normalized to *Pgk1*, RT-qPCR of *RPS4X* expression at one year shows that the transgene is active on a male X, and escapes inactivation at ~50% when on the Xi in brain and liver and spleen. An adjacent discordant gene *CITED1* also escapes from the Xi in females at ~30% in brain only. *CITED1* expression was not detected in liver and spleen. **E)** Average DNAm of skewed knock-in females shows a hypomethylated *RPS4X* promoter in all tissues, with *CITED1* having low DNAm yet still significantly hypermethylated along with *ERCC6L, HPRT* and *Phf6* compared to knock-in males in all tissues.

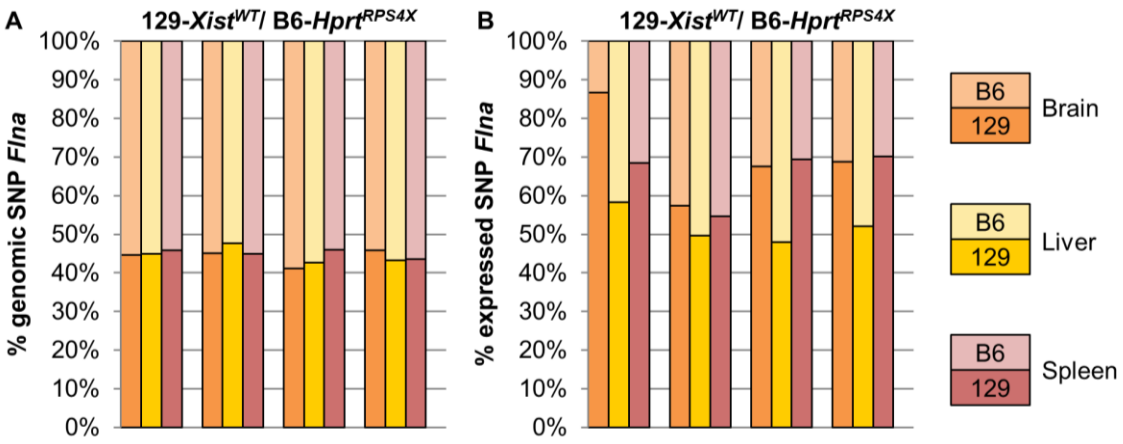


Figure 3.2 *Xist*^{WT} females have variable skewing of XCI in brain and spleen

Four one year female mice carrying the *RPS4X* BAC with wild-type *Xist* and random XCI were assessed for skewing of XCI with an X-linked gene subject to inactivation, *Flna*. **A)** Percent detection of each allele in genomic DNA showed close to 50% amplification. **B)** Percent allelic expression in cDNA showed variable extents of skewing, tending to favour expression of the 129 allele with the knock-in being on the Xi more often in brain and spleen.

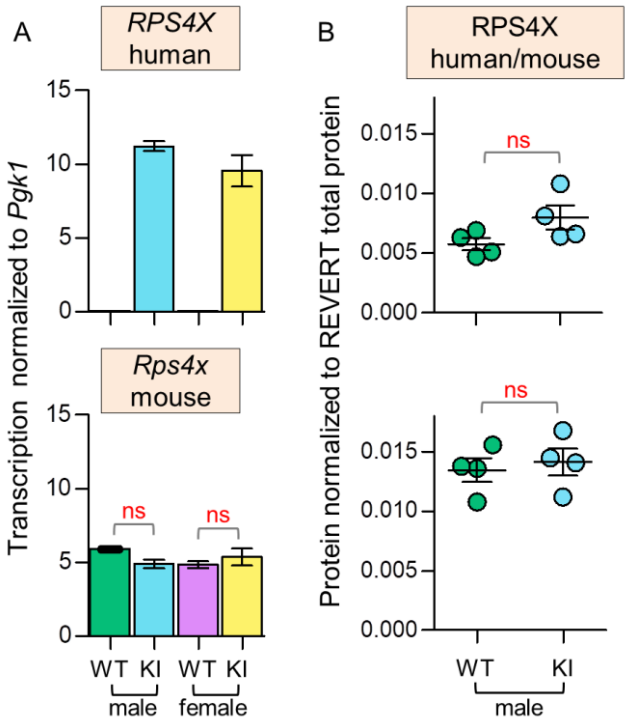


Figure 3.3 *RPS4X* transcription and protein levels in MEFs

A) Transcription of human *RPS4X* (top) and mouse *Rps4x* (bottom) by RT-qPCR relative to mouse *Pgk1* for (left to right): wild-type (WT) male (green), knock-in (KI) male (blue), WT female (purple) and KI female (yellow) mice; n=2 for each genotype. No significant difference (ns) in *Rps4x* gene expression detected between KI and WT mice for each sex despite additional human *RPS4X* transcription in KIs (unpaired *t*-test). **B)** Combined mouse and human protein levels of RPS4X normalized to a total protein stain (REVERT) for two sets of KI versus WT male mice. Due to inter-gel variability, comparisons are between quadruplicate samples on the same gels and are not significant (unpaired *t*-test).

Chapter 4: Expanding the use of mouse *Hprt* docking site to explore mechanisms of escape

4.1 Introduction

As the mouse X-linked *Hprt* docking site is permissive of escape from XCI for both mouse and human escape transgenes, it offers a genomic location to consistently screen more human constructs for evidence of escape. Similar to previous work in the lab (Yang et al, 2012), materials were used from an existing project called CanEuCre, which is a Canadian partnership with the European Commission for the development of new cre-driver resources for the study of brain and eye genes and their involvement in disease and use in gene therapy (Korecki et al, 2019; E.M. Simpson unpublished data). This project integrated human transgenes into the docking site at the 5' end of *Hprt* in mice, allowing further examination of the XCI status of different human DNA sequences in the same genomic environment.

Importantly, a BAC containing human *KDM5C* was part of the CanEuCre project and transgenic mice were successfully created. Crossing these to the *Xist* deletion strain would allow more in-depth study of females who always had the BAC on the Xi. *KDM5C* was hypothesized to be the best candidate from CanEuCre to escape from XCI in mouse as it escapes at the endogenous location in humans. Further, the mouse homologue *Kdm5c* had been shown to reproducibly escape from XCI at several X-linked locations (Li and Carrel, 2008) including *Hprt* (Peeters et al, 2018). *KDM5C* is a

widely expressed histone lysine demethylase, and plays a key role in transcriptional repression that is conserved across almost all eutherian species. Interestingly, *KDM5C* has also been implicated in sexually dimorphic diseases including cancer, as an escape from X-inactivation tumor suppressor gene, as well as X-linked intellectual disability where phenotypes in males are more severe with haploinsufficiency (reviewed in Snell and Turner, 2018). Data generated from mice is consistent with these roles, as *Kdm5c* is expressed in XX mice more highly than XY mice, independent of their gonadal types, and is not compensated by expression from the Y-linked paralogue *Kdm5d*, suggesting that sex-specific expression of *Kdm5c* may contribute to sex differences in brain function (Xu et al, 2008). As human *KDM5C* lies in an escape domain with several other escapees, regulation of its escape could differ from the mouse gene, which is a single escape gene like the previously examined human *RPS4X* (in most tissues). However, given that *KDM5C* is the conserved escape gene in its domain and studies have demonstrated the importance of a second expressed allele, it was hypothesized that the presence of an escape element would also be conserved and *KDM5C* would escape from XCI in mouse as has been shown for *Kdm5c*.

Of note, all of the genes of interest in the CanEuCre project had a reporter cassette integrated into the gene at the start codon (in exon 1 in *KDM5C*), and several (including *KDM5C*) contained an SV40 early Poly(A) signal. Thus, the resulting construct was designed to have a reporter gene under the influence of the human regulatory regions without full transcription of *KDM5C*. To validate findings from the transgenic mouse *in*

vitro, CRISPR-Cas9 technology was utilized in a female mouse ESC line to delete a region of interest near the endogenous mouse *Kdm5c* gene. Additionally, to further investigate this escape region as well other transgenes of interest, an *Hprt* deletion was generated in the female ESC line for the selectable integration of more constructs 5' of the end of *Hprt* in a higher throughput manner without the need for mouse generation and breeding time.

4.2 Methods

4.2.1 Construct generation

CanEuCre BAC constructs were generated essentially as described in 2.2.1 with the additional integration of an iCre/ERT2 fusion gene reporter using either an endogenous (Korecki et al, 2019) or artificial Poly(A) signal (E.M. Simpson, unpublished data). Most of the constructs originated from human autosomal regions, with only two X-linked BACs (carrying *KDM5C* and *SOX3*) being assessed (list of BACs and genes in Supplementary Table 4.1).

4.2.2 Generation of mouse strains

Mouse strains were generated essentially as described in 2.2.2 and (Korecki et al, 2019). Only the *KDM5C* transgenic mouse was bred to the $Xist^{1lox}$ strain for study of the BAC on the Xi. Number of integrations of *KDM5C* BAC in transgenic mice was tested using copy-number qPCR assays of a BAC backbone region (in *camR* gene) common

to the previously established single-copy *Kdm5c* BAC integration (described in 2.2.2). Copy number was analysed using the comparative CT method, normalized to *Hbb-bs* control assay and then to a *Kdm5c* control sample (Supplementary Table 4.2 for primer information).

4.2.3 Tissue collection and DNA and RNA extraction

Tissue collection performed by the Simpson Lab. For DNAm screening of CanEuCre constructs, an ear notch of ~1 mm in diameter was taken from each mouse postweaning (~4 weeks old) and digested with 200 μl of mouse homogenization buffer [50 mM KCL, 10 mM Tris- HCL, pH 8.3, 2 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.45% IGEPAL CA-630 (Sigma-Aldrich), 0.45% Tween 20 (Sigma-Aldrich), and 24 mg of Proteinase K (Sigma-Aldrich)] overnight at 55°. The digested samples were then heat inactivated at 95° for 10 min and stored at 4° until use. Tissue collection and extraction for brain, liver and spleen from 8-10 week old *KDM*5C mice as described in 2.2.3.

4.2.4 Expression, DNAm and SNP analyses

As described in 2.2.4 and 2.2.5, see Supplementary Table 4.2 for additional primer information.

4.2.5 Tissue culture

The polymorphic *Mus musculus*/*Mus castaneus* female F1 2-1 ES cell line (129/Cast) was gifted from Joost Gribnau. ESCs were cultured without feeders on 0.1% gelatin (Fisher Chemical) coated plates in the presence of FBS-ESC media [DMEM (Gibco) with 2 mM l-glutamine (Invitrogen), 0.1 mM MEM nonessential amino acid solution (Invitrogen), 1000 U/ml LIF (Millipore/Chemicon), 15% fetal bovine serum (FBS, Wisent), and 0.01% β-mercaptoethanol (Sigma-Aldrich)]. Cells were continuously sampled for retention of two X chromosomes by testing gDNA by pyrosequencing for X-linked allelic ratios of *Zfx* and *Taf1* genes (Supplementary Table 4.2 for primer information).

4.2.6 CRISPR-Cas9 design and transfections

Guide RNA (gRNA) target sequences were designed with the E-CRISP online tool (Heigwer et al, 2014). gRNAs between 19 and 22 bp were included and off-target analysis was carried out using Bowtie2 against the *Mus musculus* GRCm38 genome and the puromycin resistance gene in the nuclease plasmid to ensure the guides would not interfere with selection.

4.2.6.1 *Iqsec2* deletion

Two gRNAs targeting sequences in the 129-*Iqsec2* region of interest were individually cloned into the *Bbs*I restriction site in pSPgRNA vectors (gifted from Charles Gersbach,

Addgene plasmid #47108, Perez-Pinera et al, 2013). The two guide plasmids along with nuclease pSpCas9(BB)-2A-Puro(PX459) (gifted from Feng Zhang, Addgene plasmid #48139, Ran et al, 2013) were transfected into ESCs using Lipofectamine 3000 Transfection Reagent (Invitrogen) according to manufacturer's protocol. 24 hours after transfection cells were treated with 1.5 μg/ml puromycin for 48 hours to select for cells that had successfully taken up plasmid DNA. Resistant colonies were picked and tested by PCR amplification and sequencing of products spanning the deletion. Additionally, pyrosequencing was performed on gDNA testing for allelic ratios of a SNP in the deleted region to ensure a heterozygous deletion with loss of 129 and retention of Cast (Supplementary Table 4.2 for primer information).

4.2.6.2 *Hprt* deletion

GeneArt Platinum Cas9 Nuclease (Invitrogen) and custom synthetic IVT gRNA (Invitrogen) were transfected into ESCs using Lipofectamine CRISPRMAX Cas9 Transfection Reagent (Invitrogen) according to manufacturer's protocol. Cells were selected for double knock-out of *Hprt* function using 6-thioguanine (6-TG, Sigma-Aldrich) which is toxic to cells still producing the HPRT protein. Colonies resistant to 6-TG were picked and tested by PCR amplification and sequencing of amplicons spanning the deletion as well as those internal to the deletion to ensure homozygosity. Additionally, clones were tested for lack of survival with HAT (Gibco) media to ensure they could be used for selection of proper integration and reconstitution of the locus with transgenic constructs.

All gRNA sequences and PCR and pyrosequencing assays are listed in Supplementary Table 4.2.

4.2.7 BAC electroporations at 5' end of *Hprt*

BAC DNA was purified using the Nucleobond XTRA BAC kit (Macherey-Nagel) and linearized with *I-Sce*I (NEB). The BAC constructs were electroporated into the 129-*Hprt*^{del}/Cast-*Hprt*^{del} line with a BTX ECM 630 Electro cell manipulator. ESC clones were selected in HAT media for reconstitution of the *HPRT*/*Hprt* locus, isolated, and DNA purified. PCR with primers spanning approximately every 10 kb along the construct to test intactness were performed for RP11-236P24 (primer information in Supplementary Table 4.2).

4.3 Results

4.3.1 DNAm reflects inactivation of CanEuCre constructs at 5' end of *Hprt*

Previous examination of over 1.5 Mb of mainly autosomal genes integrated at the *Hprt* docking site in mouse revealed only one (truncated) gene out of 47 that escaped from inactivation, suggesting that elements permitting ongoing expression from the Xi are rare in the human genome, and that waystations can act over large sequences (Yang et al, 2012). To further analyze the spread of epigenetic marks and gene silencing into different human DNA sequences in the identical genomic environment, samples from a set of knock-ins directed to the 5' end of the *Hprt* locus as part of the CanEuCre project

were analysed. As the female mice generated in this project were not crossed to the *Xist*^{1lox} strain used to experimentally skew inactivation, expression could not be directly assessed to call escape as the BACs would reside on the Xa approximately 50% of the time. Instead, DNAm assays were designed for 11 genes that had promoter CpG islands (Supplementary Table 4.1) as a surrogate measure of inactivation, since promoter DNA hypermethylation in females can be attributed to DNAm of the Xi and thus reflects inactivation of the gene. There was a general trend in DNAm at female promoters being higher than male and above the ~30% cutoff used to call inactivated genes (Cotton et al, 2014), suggesting these constructs were subject to inactivation (Figure 4.1A). Bisulphite-converted DNA samples were derived from limited numbers of ear notches for each construct and therefore precluded testing for significance. Additionally, as not all genes had promoter CpG islands, neighbouring promoters *Phf6* and *HPRT* were tested for indications that the transgene had influenced their DNAm, as the previous integration of the BAC with the truncated escape gene resulted in lower DNAm of the flanking *HPRT* promoter (Yang et al, 2012). In addition to indicating escape, as *HPRT* is always located on the same X as the transgene, low DNAm could also suggest that the heterozygous females are skewed toward the transgene being more frequently on the Xa. Average female DNAm for *Phf6* and *HPRT* was 34% and 36% respectively (Figure 4.2B), not indicative of any constructs escaping and influencing the nearby heterochromatic environment or of the mice being skewed. Overall, DNAm data suggested that these constructs were inactivated and therefore were not pursued for further study.

4.3.2 Human *KDM5C* does not escape inactivation at *Hprt*

As mouse *Kdm5c* escaped from XCI at the 5' end of *Hprt*, and *RPS4X* (and *CITED1*) demonstrated escape was possible for a human gene in mouse, it was hypothesized that human *KDM5C* would also escape when integrated at the *Hprt* docking site. Transgenic mice carrying human BAC RP11–236P24 (Figure 4.2A) were analysed to determine if the BAC carried the necessary elements for escape, and if they could be recognized in mouse. Male mice carrying the knock-in on their Xa (B6-$Hprt^{KDM5C}$/129-Y), knock-in females with random XCI (129-$Xist^{WT}$/B6-$Hprt^{KDM5C}$) and females carrying the knock-in on their Xi (129-$Xist^{1lox}$/B6-$Hprt^{KDM5C}$) were assessed for gene expression and promoter DNAm in brain, liver, and spleen for six mice of each female genotype and four of the male. In all tissues, knock-in males expressed *KDM5C*, demonstrating that the transgene is capable of expression from the Xa (Figure 4.2B). Expression of an adjacent lncRNA *KANTR* was also detected from all three tissues in males. Knock-in female mice with random XCI also showed *KDM5C* and *KANTR* expression although expectedly less than a male as the transgene spends half as much time on an Xa. Unexpectedly, in knock-in females where expression is measured only from the Xi, *KDM5C* is not expressed at >10% of an Xa in any of the tissues tested. *KANTR* expression was also not detected although this was to be expected as it is normally subject to XCI in humans.

In agreement with expression data, DNAm at the CpG island promoter of *KDM5C* shows hypomethylation in males, while experimentally skewed knock-in females are

significantly more hypermethylated in all tissues (Figure 4.2C). DNAm at the promoters of the integrated human *HPRT* and the closest mouse endogenous gene to the integration site, *Phf6*, was examined to see if there were any upstream or downstream regulatory effects of the integration, but both showed methylation patterns similar to averages previously seen at these sites (Yang et al, 2012; Peeters et al, 2018).

To exclude the possibility of a second *KDM5C* BAC integration influencing results, copy number was analysed by qPCR comparing a region common between the *KDM5C* BAC and *Kdm5c* and *RPS4X* BACs, the latter two having previously been established as one and two copy respectively (Supplementary Figure 4.1, Peeters et al, 2018). *KDM5C* appears to be single-copy at *Hprt*; however, this is only based on one location in the BAC. Expressed X-linked SNP analysis of knock-in females with random XCI showed slight bias to skew the X carrying the BAC to inactivate more often but skewing was not substantial in all females (Figure 4.3). A lack of extreme skewing in the females with random XCI is in agreement with both expression and DNAm of the non-escaping BAC genes where contribution of an Xa in addition to an Xi in these females presents itself in generally higher expression levels and lower DNAm compared to the experimentally skewed females.

4.3.3 Investigating lack of *KDM5C* escape from *Hprt in vitro*

There are several possible reasons why human *KDM5C* was unable to escape from the *Hprt* site, including that it could have retained the necessary regulatory elements needed for escape in humans, but they are different from the ones recognized by mouse in the human *RPS4X* integration. It is also possible that the escape element is conserved between human and mouse *KDM5C*, but was missing on the human BAC integration. Evidence in mouse suggests that the escape element for *Kdm5c* is within a 112 kb region from the gene extending 5' into the end of *Iqsec2* (Horvath et al, 2013). Overlapping regions in human and mouse reveals that the human BAC does not capture the ~41 kb at the most 5' end of the mouse BAC (Figure 4.4), therefore this region was tested for a candidate escape element by deleting the region at the endogenous location in mouse ESCs.

The *Mus musculus*/*Mus castaneus* female F1 2-1 ES cell line (129Sv-Cast/EiJ) is especially useful for investigating endogenous genes as it is extremely polymorphic and relatively stable at retaining both X chromosomes in culture as ESCs (Jonkers et al, 2008). As the *Xce* of 129 is weaker than in Cast, the 129 locus was targeted to increase the chance of capturing more clones with the deletion on the Xi upon differentiation. Additionally, as a NPC protocol was recommended for differentiation (Gendrel et al, 2014), and *Kdm5c* has been shown to be important for neuronal development (Scandaglia et al, 2017), a homozygous deletion of the *Iqsec2* region should be avoided in case it negatively affects *Kdm5c* expression on both alleles and impacts

differentiation potential. Using CRISPR-Cas9 technology with guides specific for a PAM sequence in 129, three unique ESC clones with deletions from intron 2 to intron 6 in *Iqsec2* (Figure 4.5A) were isolated and confirmed to be on the 129 chromosome only by checking allelic ratios of an *Iqsec2* SNP within the deletion (Figure 4.5B). To determine if the deletion had any impact on *Kdm5c* expression from the Xa in ESCs, an expressed SNP in *Kdm5c* was measured and compared to the wild-type parent line (Figure 4.5C, D). The *Iqsec2* deletion did not alter *Kdm5c* expression on the Xa in three unique ESC clones, although repeated analysis after differentiation of cells and XCI has the potential to reveal an escape-specific effect on expression.

4.3.4 Generation of an *Hprt* docking site ESC model, and integration of the *KDM5C* BAC

An alternative hypothesis for the silencing of *KDM5C* at *Hprt* is that escape from inactivation is tied to full-length transcription of the gene, and that the reporter and transcription stop integrated into exon 1 of *KDM5C* interfered with the mechanism for continued expression on the Xi. To investigate if the inactivation status holds for a BAC containing *KDM5C* without the reporter, a homozygous deletion at *Hprt* was generated in the 129/Cast ESCs described above to recreate a docking site compatible with our current homologous recombination and selection system (Figure 4.6). Using CRISPR-Cas9 technology with guides recognizing target sequences in both alleles, ~5 kb from intron 1 to exon 2 was deleted in *Hprt*, likely causing a frameshift mutation and eliminating functional protein production. Lack of a functional HPRT protein from both

alleles was confirmed by selecting with 6-TG which is toxic to cells still producing the protein. Resistant colonies to 6-TG were picked and tested by PCR amplification and sequencing of products spanning the deletion, as well as products internal to the deletion to ensure homozygosity. Additionally, cells were tested for lack of survival with addition of HAT media to ensure they could be used for selection of proper integration and reconstitution of the locus with transgenic constructs containing complementary *HPRT* sequences. As female ESCs can lose an X chromosome in culture, they were continuously monitored for retention of both X chromosomes throughout all experiments by pyrosequencing of X-linked SNPs.

The wild-type *KDM5C* BAC was retrofitted with the *Hprt* homology arms and *HPRT* complementation sequence and targeted to the locus by electroporation. Two successful clones integrated and passed PCR assays approximately every 10 kb on the BAC suggesting it was intact in both. Pyrosequencing for a SNP in the region in *Hprt* replaced by the BAC detected the Cast allele only, confirming that the BAC had integrated on 129 which was to be expected as the homology arms are designed for that strain. The two clones are ready to be tested for escape from XCI upon differentiation into NPCs and subsequent expression and DNAm analysis.

4.4 Discussion

While most transgenes targeted to *Hprt* are silenced on the Xi, they demonstrate differential capacities to accumulate DNAm at their promoters. Since they are integrated into an identical location, this suggests additional *cis*-acting sequence effects, although the location of the DNAm assays within the CpG island, and the limited number of CpG sites that are measured in each one, could account for some of the variability seen.

Surprisingly, the human escape gene *KDM5C* does not escape from XCI in mouse at the *Hprt* site despite having a mouse homologue that does so. The most likely hypothesis is that the construct was missing a crucial escape element or boundary, thus demonstrating that not all escape elements are promoter-proximal. To follow-up on this hypothesis and attempt to identify elements, a region found on the *Kdm5c* escapee BAC but not on the human BAC was deleted. The deletion in *Iqsec2* did not skew *Kdm5c* expression on the Xa in ESCs compared to wild-type, therefore differentiation of those cells will be pursued to conclude if the deleted region has an escape-specific effect on expression. Alternatively, the escape element or boundary could still exist outside of the *KDM5C* region tested, as the escape domain is larger in humans and extends past *IQSEC2*. Using available data from the UCSC browser to look at potential elements that differ between human and mouse could pinpoint possible regions of interest in the escape domain, and follow-up studies integrating more BACs from this region may find that the escape element or boundary is closer to another gene (Figure 4.7). A promoter-proximal repeat analysis on *KDM5C* and *KANTR* was not overtly

predictive of their XCI statuses as *KDM5C* generally had both lower Alus (predictive of escape) and lower LINE elements (predictive of subject) than *KANTR* (Supplementary Table 4.3). There are several other BACs in the *KDM5C* escape region covering the entirety of *IQSEC2* as well as genes *SMC1A* and *RIBC1*, therefore it will be interesting to see if any of these other human escape genes are capable of escape in mouse, or if elements and boundaries potentially span the entirety of the domain and no single gene within it can escape alone.

Additionally, as the *KDM5C* analysed was designed to not be fully transcribed in the transgenic mice, the wild-type *KDM5C* could have a different inactivation status at *Hprt*. To examine this, and the possibility of examining more BACs in this region (and others), the *Hprt* docking site was recapitulated in a 129/Cast female mouse ESC model capable of differentiating (and going through XCI) to test escape potential of integrated constructs.

While two clones had successful *KDM5C* BAC integrations, recombination efficiencies at the newly generated *Hprt* site were low compared to electroporations into the original $Hprt^{b\text{-}m3}$ deletion (Supplementary Table 4.4) which precluded investigation of other constructs at *Hprt* in a timely manner. As the new deletion was about 30 kb smaller than the original, the reduction in efficiency could have been due to the increase in genomic distance between the homology arms and therefore was more challenging for the

construct to recombine. To test this, the deletion was re-created to more closely match *Hprt*$^{b\text{-}m3}$, but the cell line needs testing to demonstrate if it can provide the desired increase in efficiency.

Overall, the assessment of multiple human constructs from the CanEuCre project uncovered many more genes becoming subject to XCI in the same location, including the human escape gene *KDM5C*. While surprising, this construct demonstrated that not all human escape genes are capable of escape despite >150 kb of the endogenous environment being integrated with it. This work has set the stage to uncover why *KDM5C* was inactivated by generating a candidate region knockout and a new ESC system for integrating new constructs including the BAC without the transcription stop, and additional BACs in the region.

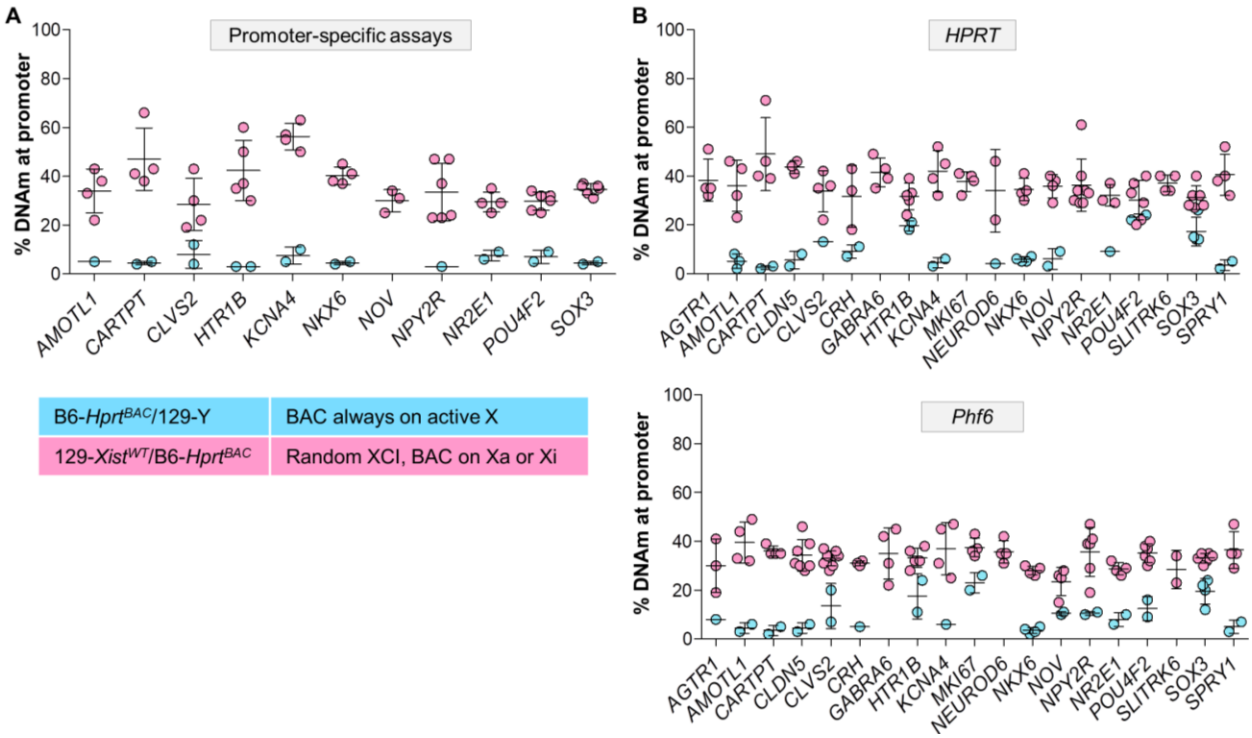


Figure 4.1 DNAm screening of CanEuCre constructs suggests that all BACs are subject to XCI despite their size or native location in humans

A) Eleven BAC constructs had assayable CpG islands in their gene promoters, and DNAm levels were tested from knock-in male (blue) and knock-in females with random XCI (pink) to see if any indicated escape. While transgene promoters had variable methylation, none of the females were as hypomethylated as males. **B)** Promoter DNAm of neighbouring genes *HPRT* and *Phf6* for the original eleven BACs plus an additional eight BAC integrations without CpG island promoters in genes of interest generally show hypermethylation in females.

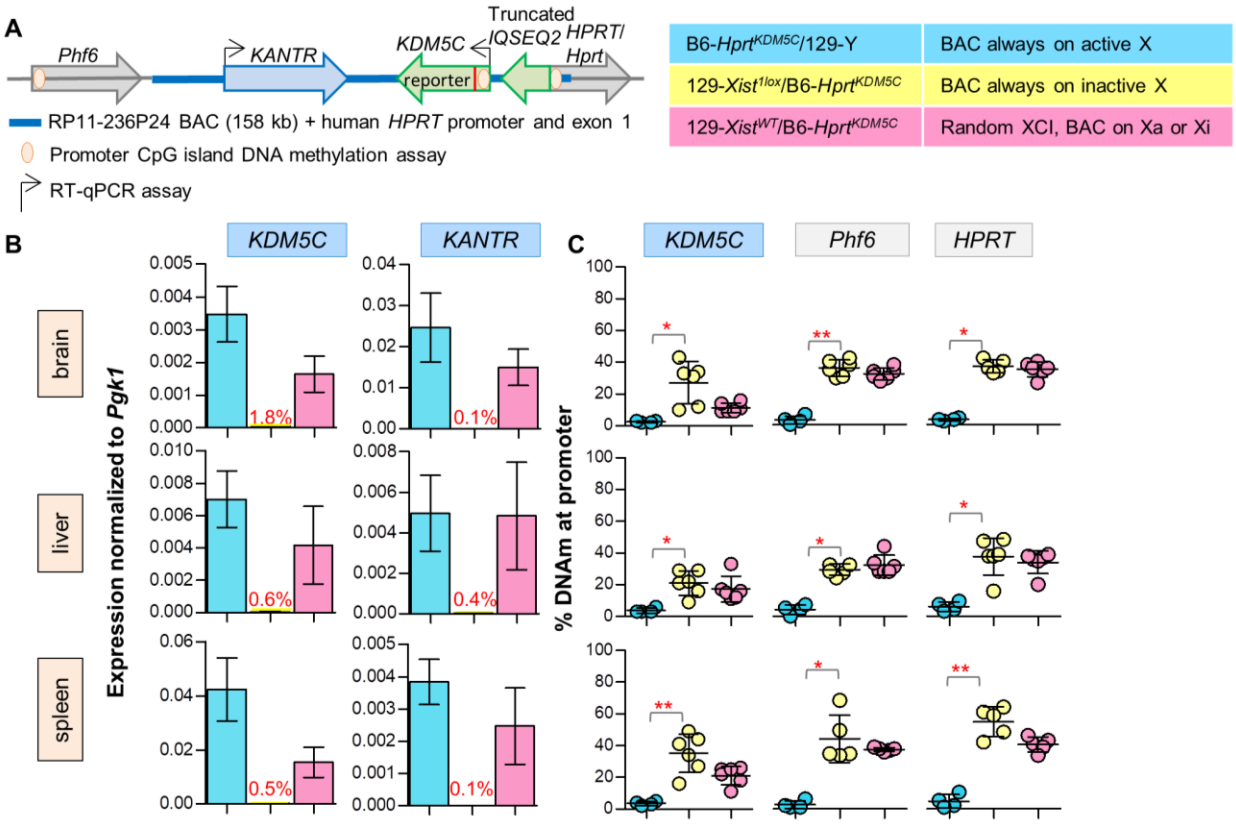


Figure 4.2 Analysis of BAC RP11–236P24 shows *KDM5C* does not escape from XCI at *Hprt*

A) Integration of the *KDM5C* BAC at *Hprt*; genes on the BAC expected to escape from XCI in green, subject in blue, and genes at integration site (both known to be subject to XCI) in grey. Genes with RT-qPCR and DNAm assays are indicated as well as a red line in *KDM5C* to indicate reporter gene and lack of full *KDM5C* transcription (RT-qPCR assay measuring *KDM5C* expression targets exon 1 prior to the SV40 polyA signal. Description of genotypes shown, six mice for each female, four mice for male. **B)** Normalized to *Pgk1*, RT-qPCR of *KDM5C* expression shows that the transgene is active on a male X, but surprisingly subject to inactivation on the female Xi in all three tissues tested. An adjacent lncRNA *KANTR* is subject to XCI as expected. Expression from the female Xi is shown as percentage of the male X (red text). **C)** Average DNAm of *KDM5C* as well as *HPRT* and *Phf6* shows hypermethylated promoters on the Xi in females (Mann-Whitney t-test, significance from hypomethylated males denoted by

asterisks, p-value <0.001 ***, 0.001 to 0.01 **, 0.01 to 0.05 *, >0.05 ns), supporting the lack of expression and demonstrating that integration of the BAC had no effect on neighbouring genes.

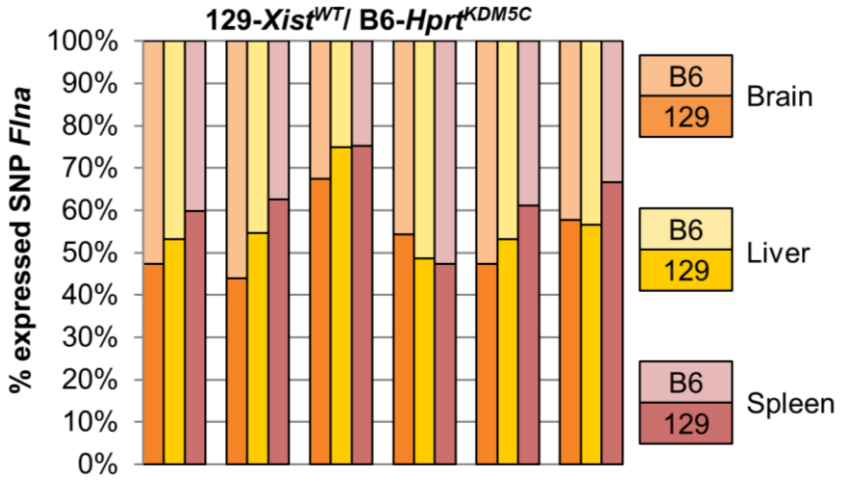


Figure 4.3 *Xist*^{WT} females do not have consistently skewed XCI

Percent allelic expression of X-linked gene *Flna* (subject to inactivation) in six female mice carrying *KDM5C* transgene all with wild-type *Xist* and random XCI. All three tissues are shown for each mouse with the percent expression indicating how often the B6 or 129 X is the Xa.

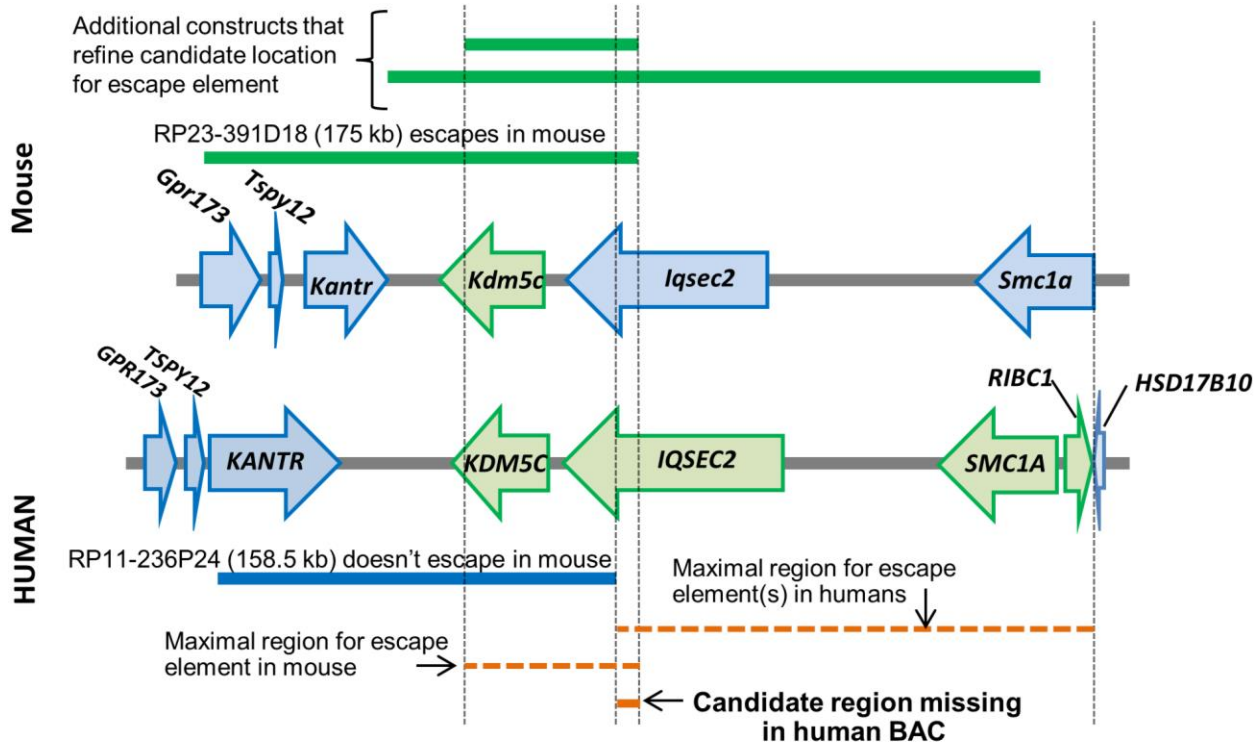


Figure 4.4 Comparison of the human and mouse regions containing *KDM5C*/*Kdm5c* offers new areas to interrogate for involvement in escape

Human *KDM5C* resides in an escape domain with three additional escape genes that are subject to XCI in mouse. BAC transgenes tested at *Hprt* are labelled (Peeters et al, 2018) with other informative *Kdm5c* constructs shown that have escaped from XCI and demarcate the minimal region necessary to escape (Li and Carrel, 2008; Horvath et al, 2013). Overlapping regions in human and mouse reveals a candidate region in the *Iqsec2* gene located on the mouse BAC but missing from the human integration.

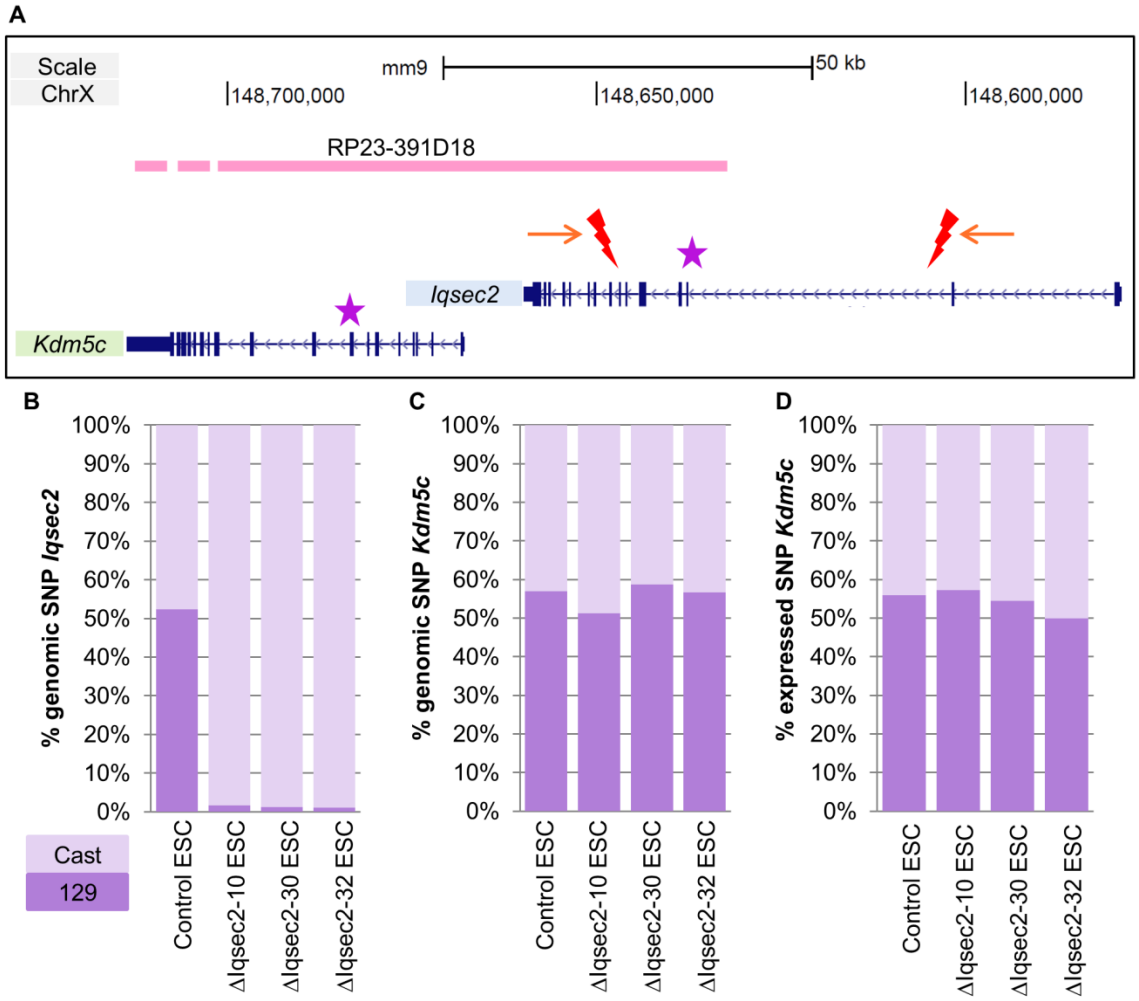


Figure 4.5 Deletion of the candidate region in *Iqsec2* does not alter allelic expression of *Kdm5c* in ESCs

A) Representation of the *Iqsec2* deletion and assay positions in the UCSC browser. Guide RNAs (red bolts) targeting *Iqsec2* take out a larger region past where the mouse BAC ends (pink bar) in order to use a 129-specific SNP in the PAM sequence. SNPs used for pyrosequencing shown as purple stars. Orange arrows denote PCR assay to amplify and sequence deletion. **B)** Pyrosequencing of gDNA from three deletion clones shows they have lost the 129 *Iqsec2* SNP confirming a heterozygous deletion. **C)** *Kdm5c* SNP analysis of gDNA shows that all three clones retain both X chromosomes in culture. **D)** SNP analysis of cDNA shows allelic expression of *Kdm5c* in three ESC XaXa *Iqsec2* deletion clones is not affected.

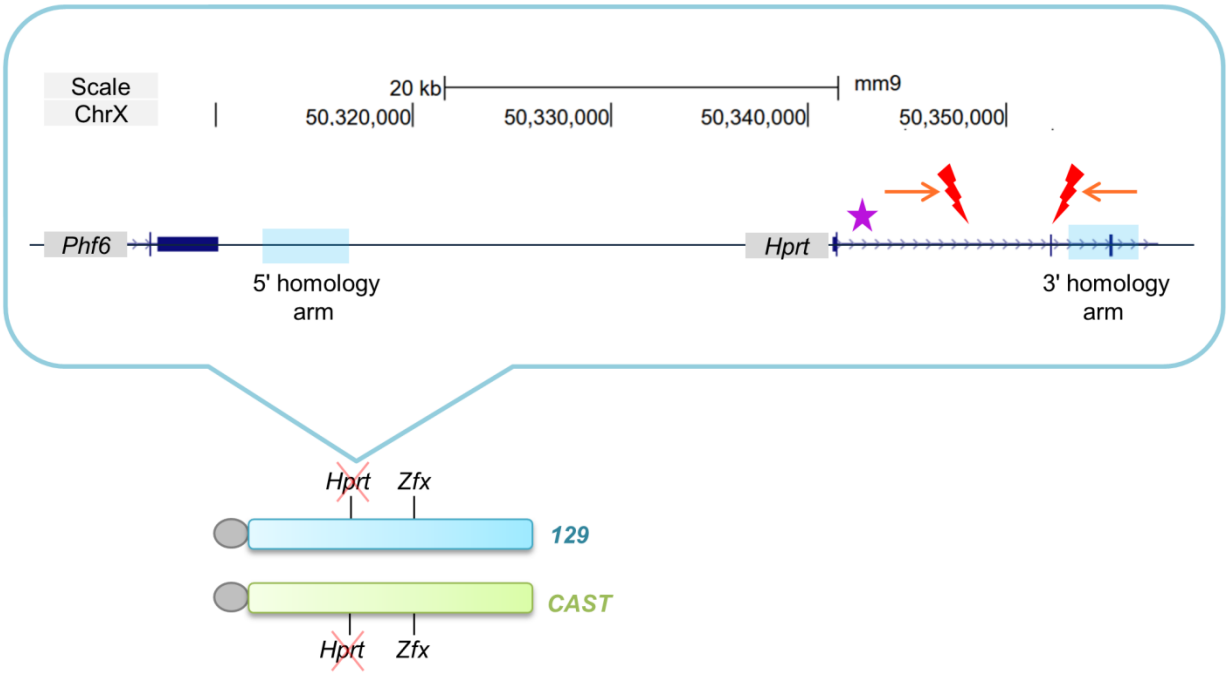


Figure 4.6 Generation of *Hprt* docking site in female ESCs

Hprt function was homozygously knocked out in a 129/Cast female ESC line using CRISPR-Cas9 and gRNAs targeted to a small region spanning intron 1 to exon 2 (red bolts). Clones surviving 6-TG selection for lack of a functional HPRT protein were amplified around the deletion (orange arrows) and sequenced. This deletion is smaller than the original $Hprt^{b\text{-}m3}$ deletion which extends closer to the 5' homology arms (as shown in Figure 2.1). Successful integration of a construct outfitted with homology arms and the *HPRT* complementary sequence can recapitulate *Hprt* function and can be selected for with HAT media. A 129/Cast SNP in *Hprt* (purple star) is checked by pyrosequencing to determine which allele the BAC has integrated into and “replaced” while a SNP in another X-linked gene, *Zfx* is routinely checked to ensure both a 129 and Cast copy are present.

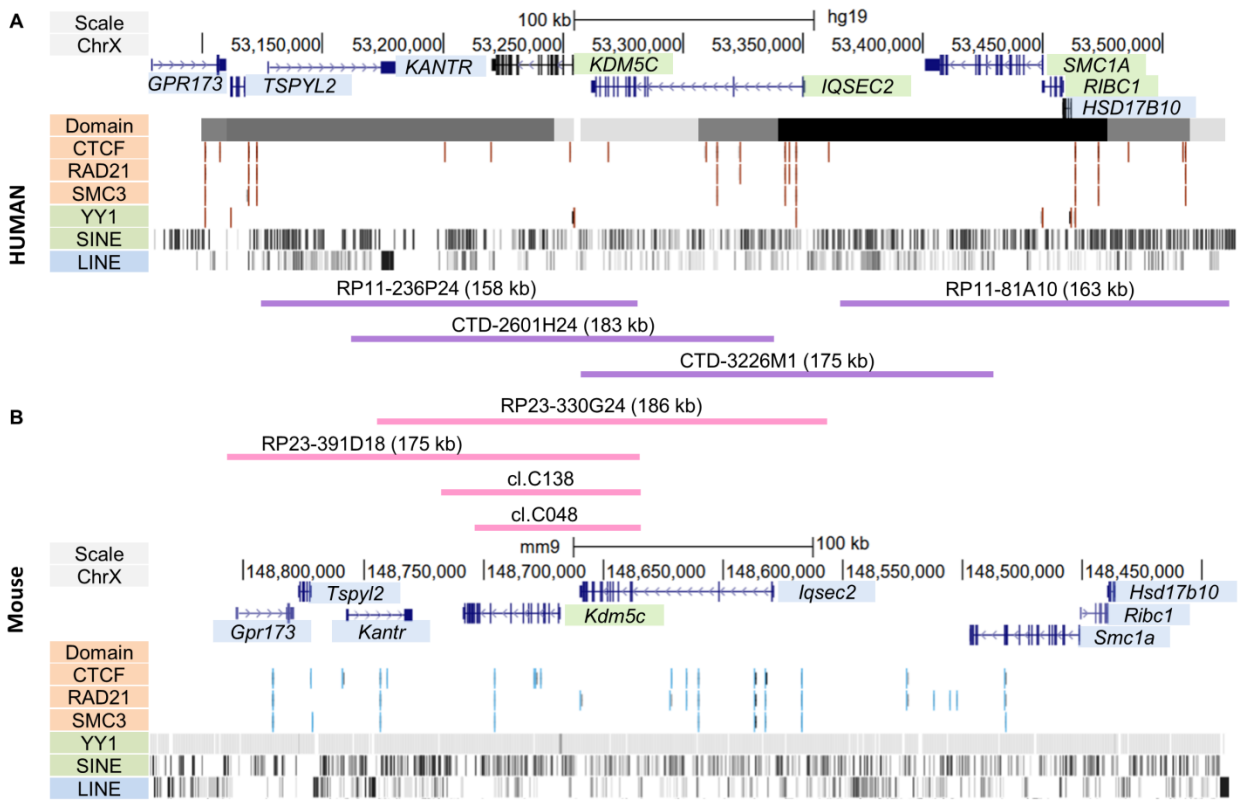


Figure 4.7 Comparing potential DNA elements between mouse and human *KDM5C*/*Kdm5c* escape regions

Using available datasets, potential elements were examined that may aid in setting up an escape domain around the **A)** human and **B)** mouse genes in their endogenous locations (shaded for escape) displayed in the UCSC browser (hg19 and mm9 assemblies). Factors associated with escape from XCI as well as components of structural complexes to form boundaries are displayed in humans from female cell line GM12878. In mouse, transcription factor binding sites are from CH12 cells, with the exception of YY1 binding which was done in ESCs (Sigova et al, 2015). Contact domains are shown in human but were not informative for this region in mouse (Rao et al, 2014). Potential BAC integrations are displayed for human (purple, Peeters et al, 2018) and mouse (pink, Li and Carrel, 2008; Horvath et al, 2013).

Chapter 5: Discussion

5.1 Importance of investigating escape from XCI

The goals of this thesis were to determine if the elements regulating escape from inactivation were conserved between mice and human, and if mice are an informative model for studying human escape at both an organismal and *in vitro* level. XCI achieves dosage compensation for a majority of genes on the X, yet more than 12% of genes continue to show expression from the Xi across cells, tissues, and individuals, while another 15% exhibit variable expression (Balaton et al, 2015). This Xi expression has a significant impact on sexually dimorphic traits throughout development, and human health. Breeding schemes controlling for gonadal type in mouse have pinpointed a role for the Xi affecting adiposity and metabolic disease, cardiovascular injury, and behavior (Arnold and Chen, 2008). Phenotypic impacts are likely to be even more pronounced in humans due to the larger number of escape genes. In human females, the differential expression of escapees can offer a protective effect against *de novo* and inherited X-linked mutations, but they could also contribute to the over-representation of females for some complex traits, such as autoimmune disorders, and contribute to the features seen in X chromosome aneuploidies (reviewed in Carrel and Brown, 2017). As many linkage and association studies historically have not included the X chromosome because of the complexity of analysis, more connections are likely to be found as new X-specific tools are developed (Broman et al, 2006; Gao et al, 2015). Characterizing the regulatory elements governing genes that escape from XCI will contribute to understanding sexual dimorphisms, spread of gene activity in *cis*, and will yield

elements whose ability to protect transgenes from silencing will be useful biotechnology tools. It has been shown that *XIST* can be used for “chromosome therapy” (Jiang et al, 2013), and the utility of regulating heterochromatin will be improved by understanding the nature of the elements and boundaries that influence its spread. In an attempt to harness the potential of the second allele to protect from X-linked disease in females, recent studies have explored shRNA and pharmacological reactivation of the Xi, but have determined that such approaches tend to reactivate a substantial portion of silenced genes (for example, Lessing et al, 2016). Therefore, understanding of how expression from the Xi is limited to only a subset of genes on the X might provide insights into utilization of such therapeutic approaches.

5.2 Major findings of this thesis

Through the generation and analysis of the *RPS4X* transgenic mice, work in this thesis established that signals driving escape in humans are recognized in mouse, and are stable across development. *CITED1* was confirmed as an escape gene with brain-specific Xi expression in the set of tissues analyzed, and added evidence supporting *cis*-acting elements regulating escape from XCI regardless of genomic location, which previously had only begun to be functionally characterized for the mouse gene *Kdm5c* (Li and Carrel, 2008; Horvath et al, 2013). For *RPS4X*, escape from XCI occurs early by E9.5, so is unlikely to be reactivation, and continues across the lifespan of the mouse up to one year with no obvious detrimental effects from overexpression of the gene, likely due to post-transcriptional regulation. There is a significant increase in gene

expression from *Kdm5c* in transgenic mice with a second (males) or third (females) copy, but whether this translates into additional protein is unknown, and perhaps unlikely in tissues such as the brain where deviations in protein level could have phenotypic consequences. More work pairing gene expression data with protein quantification and phenotypic outcomes will be required to elucidate what amount of escape at the RNA level is necessary to translate into biologically relevant sexual dimorphisms at the organismal level.

While mouse cells were able to recognize and respond appropriately to the human elements regulating *RPS4X*, it remains unclear if the *cis*-acting elements themselves are conserved between species. The endogenous mouse *Rps4x* gene does not escape XCI; however, no notable difference in described elements is apparent between humans and mice, strongly suggesting that there are more elements to be identified. Where an escape element was hypothesized to be conserved between species for the *KDM5C* gene, transgenic copies of the mouse and human genes behave in a different manner on the Xi. The *KDM5C* region that was targeted to the *Hprt* site had insufficient elements to escape from XCI, while the mouse homologue escaped inactivation at *Hprt*, in addition to four previously described random integration sites (Li and Carrel, 2008). As human *KDM5C* lies in an escape region in humans, it could retain some of the same elements as mouse but be missing a critical boundary or combination of elements that lie closer to other genes in its domain. *Kdm5c* is a single escape gene bounded by subject genes suggesting it has retained its regulatory elements in closer proximity.

Reasons for why *KDM5C* is subject to XCI in mice have begun to be addressed by testing the deletion of a possible escape element contained on the mouse BAC, but not on the human BAC. Removing this region near the endogenous *Kdm5c* in a 129/Cast mouse ESC model demonstrated that it is not essential for *Kdm5c* expression on the Xa in ESCs. *Kdm5c* expression from the Xi will be monitored by DNAm and allelic expression upon differentiation to see if the region has an element that acts in an escape-specific manner only.

Additionally, as transcription through the gene body could be necessary for escape, and the human *KDM5C* had a reporter and early transcription stop signal, testing of the fully expressed gene in the ESC line was initiated. To accomplish this, a docking site at the 5' end of *Hprt* was recapitulated to provide the capacity to screen more rapidly for Xi-elements without mouse generation. While both the *KDM5C* BAC and smaller control plasmids have been successfully integrated into this cell line, this site demonstrated low efficiency compared to the original *Hprt* deletion. gRNAs were redesigned to more closely match the original deletion, but these new knockout cells require electroporation and recombination testing to determine if efficiency has improved by bringing the homology arm targets closer together.

5.3 Future directions

Improving the *in vitro* model is a critical step to moving forward with high-throughput analysis at *Hprt.* If the larger deletion (which therefore has a shorter region between homology arms) does indeed increase efficiency of recombination, the docking site can also be used to integrate other BACs in the ~354 kb *KDM5C* escape region as there is still the possibility that the intrinsic escape element is contained in, or spread across the other escape genes. Additionally, it will be interesting to investigate the escape potential of other escape gene BACs from across the Xi, including those from PAR1 as it is the largest escape domain in humans.

While the ESCs can be differentiated through a NPC protocol, the method is time-consuming with the additional step of single-cell cloning for a line that has retained both X chromosomes, as loss of an X chromosome is common (Gendrel et al, 2014, Choi et al, 2017). In addition, clones that retain both X chromosomes ideally need to include ones where the construct resides on the Xi (to test for escape) as well as on the Xa (as a control for expression). Moving forward, culturing cells prior to differentiation in a 2i (or only a MEK inhibitor) media has been reported to improve retention of two Xs (Choi et al, 2017). Additionally, inducing *Xist* expression can decrease the total differentiation time and also lead to less X chromosome loss (Żylicz et al, 2019).

Investigating the elements controlling *RPS4X* escape could also be revisited in an improved *in vitro* system. Overall, *RPS4X* is a good candidate for follow-up studies as its stability of escape across individual mice, tissues, and age, lends confidence that changes in XCI status due to manipulations of the genomic environment (such as CRISPR-Cas9 deletions) will not be related to these types of variability. One approach would be to test a reporter construct (Supplementary Figure 5.1) under regulatory control of the *RPS4X* promoter and proximal elements that could elucidate the minimal region necessary to drive escape. Constructs could be integrated into the female ESC model with the *Hprt* docking site and tested for expression of the reporter gene after differentiation and XCI. An alternative approach would be to interrogate candidate regions contained within the BAC by making targeted deletions with CRISPR-Cas9 either in ESCs with the BAC allowing analysis during initiation of XCI, or human female somatic cells to analyze effects on maintenance of the *RPS4X* escape status. If the *RPS4X* promoter is incapable of driving escape of the reporter, or no influence of the directed deletions is observed, then studies could be extended to a series of tiling deletions between the subject and escape genes on the BAC.

5.4 Outstanding questions about how genes escape from XCI

It remains to be elucidated why some genes escape from chromosome-wide XCI, and there are many outstanding questions about the types of elements involved including: whether there are different mechanisms for different genes; whether they are gene-specific or can one element control escape within a domain; and are escape elements

the same regulatory controls as used on the Xa. Transgenes like human *KDM5C* that retain expression on an Xa in males and females with random XCI, but do not escape from the Xi, would argue that the element missing is uniquely important for Xi expression, such as a boundary to block the spread of heterochromatin. Common regulatory regions utilized by both the Xa and Xi could be identified by interrogating transgenes that did escape from the Xi at *Hprt*. Knowing that the minimum region needed for escape is contained within BACs such as *RPS4X* narrows down the testable elements for escape; however, identification of a region abolishing expression on both the Xa and Xi does not definitively mean it is the only escape element as it could be combinatorial. Testing small promoter candidates with a reporter gene would provide more conclusive data for a single escape element. Experiments such as these demonstrate the utility of a transgenic human-mouse model, as discerning whether an element is Xi-specific or used by both alleles would be complicated by not knowing the maximum area in which to look for said elements. Additionally, having a copy of the gene on both the Xa and Xi in the same cell in humans requires the use (and existence) of known expressed polymorphisms, or a strong enough RNA signal to confidently detect transcription and loss of transcription by fluorescent *in situ* hybridization, which may hinder the analysis of some escape genes (particularly when Xi expression is only a fraction of Xa expression).

The differences in number and organization of escape genes between human and mouse indicates that both promoter-proximal and domain regulation are at play. The

ability of transgenic mouse *Kdm5c* to escape while *KDM5C* was silenced provides evidence for this, suggesting *KDM5C* may require an element or boundary located elsewhere in its endogenous escape domain. Some of the difference between escape number in human and mouse may be accounted for by differential positioning of escape genes relative to the *XIC* and other loci contributing to the structure of the Xi. In humans, *KDM5C* resides on the p arm which is enriched in escape genes and on the other side of the centromere from *XIST*, unlike in mouse where the X chromosome is acrocentric. It is possible that the centromere may act as a barrier to the spread of inactivation from the q arm to the p arm in humans, but to date Xist RNA exclusion from constitutive heterochromatin has only been demonstrated in mouse and is likely impacted by cell cycle dynamics at the time of analysis (Duthie et al, 1999; Clemson et al, 2006). The *XIST*/*Xist* loci are mid-chromosome in both species, however, other loci contributing to the Xi structure are differentially positioned along the mouse and human X chromosome (reviewed in Balaton and Brown, 2016). Distance from the *Xic* and *Xist* entry sites has been correlated with both escape and timing of inactivation in mouse, as genes that inactivate latest tend to be located more distally (Marks et al 2015; Borensztein et al, 2017). As genes in the Xq PAR2 (Ciccodicola et al, 2000), and X material translocated beyond the Xp terminus (Becroft et al,1977) can still be inactivated, distance alone from the *XIC* is not the sole determining factor for escape ability. Indeed, in a series of *Xist* translocations to several different loci along the mouse X chromosome, a group of genes consistently escaped from XCI regardless of the location of *Xist*, demonstrating an intrinsic ability to consistently resist inactivation (Loda et al, 2017).

In a study of human X-autosome translocations the authors noted it was a di-centric chromosome that showed the lowest degree of inactivation (Cotton et al, 2013). The same X-autosome translocation also showed the lowest LINE and the highest Alu content of all the translocations studied. Alu elements have been consistently enriched in regions surrounding the transcription starts of genes that escape inactivation, especially 50 kb and 100 kb windows, located both upstream and downstream of the TSS (Wang et al, 2006) suggesting that the larger genomic environment, and not simply the promoter context, may be most relevant for determining X-inactivation status. This is only true for humans as Alu repeats are primate-specific (reviewed in Batzer and Deininger, 2002), and the relatively low proportion of genes escaping from XCI in mouse may be consistent with the much lower B1 element frequency (the Alu equivalent) in the mouse X-chromosome (10% in humans to 2% in mouse, Wang et al, 2006). While Alu elements may have a potential role in human escape, they are not informative for predicting escape for all genes as they too have a variable distribution between evolutionary strata on the X (Wang et al, 2006). Although Alus were informative for the X-autosome translocations discussed above, they were not able to discern between human autosomal BAC transgenes as being subject to or escaping from XCI at *Hprt* (Yang et al, 2012) or distinguish between subject and escape genes on X-linked BACs tested here. Therefore, only a subset of human genes may make use of independently evolved regulatory elements in their escape mechanism, which in turn could affect their capability to escape from XCI in mouse.

5.5 Conclusion

It is possible that all genes are capable of escape, but need the correct context to do so, which is suspected to be a combination of multiple factors such as interactions of DNA sequence, local chromatin environment, and three-dimensional chromosome ultrastructure (Figure 5.1). This combination may differ between escape genes, and be influenced by their evolutionary history on the X chromosome. The nature of these elements or definitive evidence that any one particular feature is necessary or sufficient for a gene to escape XCI is still elusive. While bioinformatic studies have indicated a number of potential features, a predictive model has not yet had the power to classify the majority of escape genes correctly, again suggesting that different combinations of features are likely, and some features have not yet been identified, demonstrating a need for the type of functional studies pioneered by this thesis on individual genes for validation. Importantly, despite a lack of conservation in number and distribution of genes that escape from XCI, this thesis has demonstrated the ability of mouse to recognize human elements regulating escape from XCI and stably express escapees throughout development, thereby providing a model system for the exploration of these elements *in vivo.* It has also described an *in vitro* system for screening of escapee-containing BACs, and refinement of candidate *cis*-acting elements involved in escape.

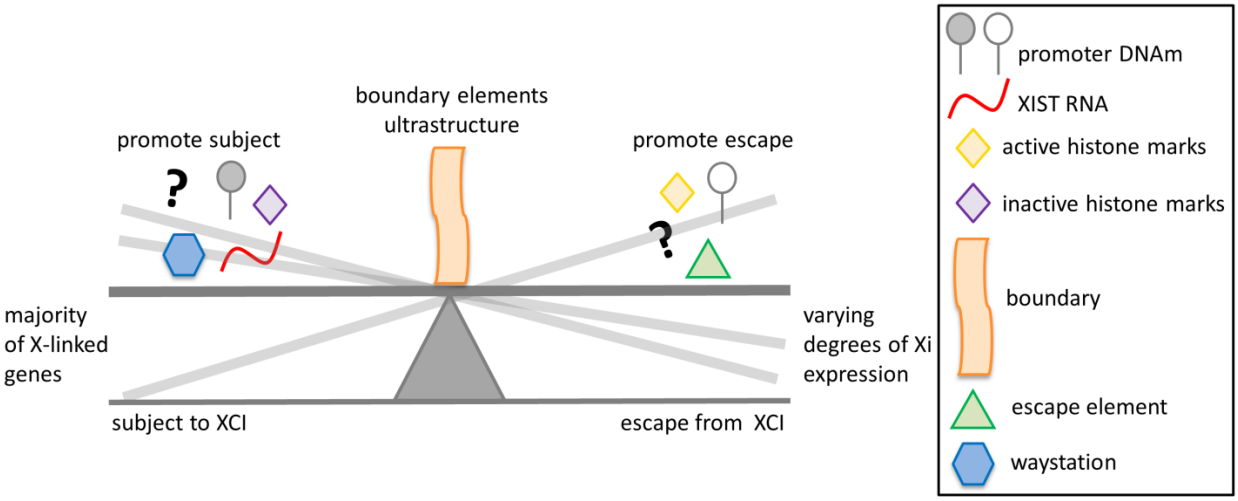


Figure 5.1 A combination of features contribute to the ability of a gene to escape from XCI

There does not appear to be a definitive set of features that cause a gene to escape from XCI, rather, it is likely a combination of multiple features including interactions of DNA sequence, local chromatin environment, and three-dimensional chromosome ultrastructure. Escape genes can be further broken down by varying degrees of expression from the Xi in different cells, tissues, and females, which could be due to additional influence of tissue-specific regulators, genetic differences in the element binding sites, stochastic variation, and measurement sensitivity affected by cellular heterogeneity and transcriptional dynamics.

References

Agulnik AI, Mitchell MJ, Mattei MG, Borsani G, Avner PA, Lerner JL, Bishop CE. A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. 1994; Hum Mol Genet, 3(6):879–884.

Anderson CL, Brown CJ. Epigenetic predisposition to expression of TIMP1 from the human inactive X chromosome. 2005; BMC Genet, 6:48.

Arnold AP, Chen X. What does the “four core genotypes” mouse model tell us about sex differences in the brain and other tissues? 2008; Front Neuroendocrinol, 30(1):1–9.

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. 2013; Nat Rev Genet, 14(2):113–124.

Bailey JA, Carrel L, Chakravarti A, Eichler EE. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. 2000; Proc Natl Acad Sci USA, 97(12):6634–6639.

Balaton BP, Brown CJ. Escape Artists of the X Chromosome. 2016; Trends Genet, 32(6):348–359.

Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. 2015; Biol Sex Differ, 6:35.

Balaton BP, Dixon-McDougall T, Peeters SB, Brown CJ. The eXceptional nature of the X chromosome. 2018; Hum Mol Genet, 27(R2):R242–R249.

Batzer MA, Deininger PL. Alu repeats and human genomic diversity. 2002; Nat Rev Genet, 3(5):370–379.

Becroft DM, Costello JM, Shaw RL. 46, X,X-X terminal rearrangement/45, X mosaicism in a child with short stature. 1977; Clin Genet, 11(2):122–127.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T, Rock S, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. 2014; Nature, 508(7497):494–499.

Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteche CM, Deng X. Escape from X inactivation varies in mouse tissues. 2015; PLoS Genet, 11(3):e1005079.

Boda E, Pini A, Hoxha E, Parolisi R, Tempia F. Selection of reference genes for quantitative real-time RT-PCR studies in mouse brain. 2009; J Mol Neurosci, 37(3):238–253.

Bonora G, Deng X, Fang H, Ramani V, Qiu R, Berletch JB, Filippova GN, Duan Z, Shendure J, Noble WS, et al. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. 2018; Nat Commun, 9(1):1445.

Borensztein M, Syx L, Ancelin K, Diabangouaya P, Picard C, Liu T, Liang J-B, Vassilev I, Galupa R, Servant N, et al. Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. 2017; Nat Struct Mol Biol, 24(3):226-233.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, et al. Characterization of a murine gene expressed from the inactive X chromosome. 1991; Nature, 351(6324):325–329.

Boyle AL, Ballard SG, Ward DC. Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. 1990; Proc Natl Acad Sci USA, 87(19):7757–7761.

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. 1991; Nature, 351(6324):329–31.

Broman KW, Sen S, Owens SE, Manichaikul A, Southard-Smith EM, Churchill GA. The X chromosome in quantitative trait locus mapping. 2006; Genetics, 174(4):2151–2158.

Bronson SK, Plaehn EG, Kluckman KD, Hagaman JR, Maeda N, Smithies O. Single-copy transgenic mice with chosen-site integration. 1996; Genesis, 93(17):9067–9072.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. 1991; Nature, 349(6304):38–44.

Brown CJ, Greally JM. A stain upon the silence: genes escaping X inactivation. 2003; Trends Genet, 19(8):432–438.

Brown CJ, Rider S, Chelly J, Taillon-Miller P, Chinault AC, Monaco AP, Willard HF. 2.6 Mb YAC contig of the human X inactivation center region in Xq13: physical linkage of the RPS4X, PHKA1, XIST and DXS128E genes. 1993; Hum Mol Genet, 2(8):1105–1115.

Calabrese JM, Sun W, Song L, Mugford JW, Williams L, Yee Della, Starmer J, Mieczkowski P, Crawford GE, Magnuson T. Site-specific silencing of regulatory elements as a mechanism of X inactivation. 2012; Cell, 151(5):951–963.

Calaway JD, Lenarcic AB, Didion JP, Wang JR, Searle JB, McMillan L, Valdar W, de Villena FP-M. Genetic architecture of skewed X inactivation in the laboratory mouse. PLoS Genet, 2013;9(10):e1003853.

Cameron-Pimblett A, La Rosa C, King TFJ, Davies MC, Conway GS. The Turner syndrome life course project: Karyotype-phenotype analyses across the lifespan. 2017; Clin Endocrinol (Oxf), 87(5):532–538.

Cantrell MA, Carstens BC, Wichman HA. X chromosome inactivation and Xist evolution in a rodent lacking LINE-1 activity. 2009; PLoS One, 4(7):e6252.

Carrel L, Brown CJ. When the Lyon(ized chromosome) roars: ongoing expression from an inactive X chromosome. 2017; Philos Trans R Soc Lond B Biol Sci, 372(1733).

Carrel L, Hunt PA, Willard HF. Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. 1996; Hum Mol Genet, 5(9):1361–1366.

Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD. Genomic environment predicts expression patterns on the human inactive X chromosome. 2006; PLoS Genet, 2(9):e151.

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. 2005; Nature, 434(7031):400–404.

Chaligné R, Popova T, Mendoza-Parra M-A, Saleem M-AM, Gentien D, Ban K, Piolot T, Leroy O, Mariani O, Gronemeyer H, Vincent-Salomon A, Stern MH, Heard E. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. 2015; Genome Res, 25(4):488–503.

Chen C-Y, Shi W, Balaton BP, Matthews AM, Li Y, Arenillas DJ, Mathelier A, Itoh M, Kawaji H, Lassmann T, et al. YY1 binding association with sex-biased transcription revealed through X-linked transcript levels and allelic binding analyses. 2016; Sci Rep, 6:37324.

Chen G, Schell JP, Benitez JA, Petropoulos S, Yilmaz M, Reinius B, Alekseenko Z, Shi L, Hedlund E, Lanner F, et al. Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. 2016; Genome Res, 26:1342–1354.

Choi J, Huebner AJ, Clement K, Walsh RM, Savol A, Lin K, Gu H, Di Stefano B, Brumbaugh J, Kim S-Y, et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. 2017; Nature, 548(7666):219–223.

Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E, et al. LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. 2010; Cell, 141(6):956–969.

Ciavatta D, Kalantry S, Magnuson T, Smithies O. A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. 2006; Proc Natl Acad Sci USA, 103(26):9958–9963.

Ciccodicola A, D'Esposito M, Esposito T, Gianfrancesco F, Migliaccio C, Miano MG, Matarazzo MR, Vacca M, Franzè A, Cuccurese M, et al. Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. 2000; Hum Mol Genet, 9(3):395–401.

Clemson CM, Hall LL, Byron M, McNeil J, Lawrence JB. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. 2006; Proc Natl Acad Sci USA, 103(20):7688–7693.

Cooper DW, VandeBerg JL, Sharman GB, Poole WE. Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. 1971; Nat New Biol, 230(13):155–7.

Cotton AM, Chen C-Y, Lam LL, Wasserman WW, Kobor MS, Brown CJ. Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains. 2013; Hum Mol Genet, 23(5):1211–1223.

Cotton AM, Lam L, Affleck JG, Wilson IM, Peñaherrera MS, McFadden DE, Kobor MS, Lam WL, Robinson WP, Brown CJ. Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. 2011; Hum Genet, 130(2):187–201.

Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. 2014; Hum Mol Genet, 24(6):1528–1539.

Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, Wang JR, Morgan AP, Calaway JD, Aylor DL, et al. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. 2015; Nat Genet, 47(4):353–60.

Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. 1999; Nat Genet, 22(4):323–324.

Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang S-C, Sanborn AL, Machol I, Shamim M, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. 2016; Proc Natl Acad Sci USA, 113(31):E4504–12.

Deng X, Ma W, Ramani V, Hill A, Yang F, Ay F, Berletch JB, Blau CA, Shendure J, Duan Z, et al. Bipartite structure of the inactive mouse X chromosome. 2015; Genome Biol, 16:152.

Dixon-McDougall T, Brown C. The making of a Barr body: the mosaic of factors that eXIST on the mammalian inactive X chromosome. 2015; Biochem Cell Biol, 94(1):56–70.

Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, Sullivan TJ, Hess JM, Gimelbrant AA, Beroukhim R, et al. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. 2016; Nat Genet, 49(1):10–16.

Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM, Brockdorff N. Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. 1999; Hum Mol Genet, 8(2):195–204.

Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. 2013; Science, 341(6147):1237973.

Ferreira AR, Machado GM, Diesel TO, Carvalho JO, Rumpf R, Melo EO, Dode MAN, Franco MM. Allele-specific expression of the MAOA gene and X chromosome inactivation in in vitro produced bovine embryos. 2010; Mol Reprod Dev, 77(7):615–621.

Filippova GN, Cheng MK, Moore JM, Truong J-P, Hu YJ, Tsuchiya KD, Disteche CM. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. 2005; Dev Cell, 8(1):31–42.

Fisher EM, Beer-Romero P, Brown LG, Ridley A, McNeil JA, Lawrence JB, Willard HF, Bieber FR, Page DC. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. 1990; Cell, 63(6):1205–1218.

Galupa R, Heard E. X-chromosome inactivation: new insights into cis and trans regulation. 2015; Curr Opin Genet Dev, 31:57–66.

Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z, Keinan A. XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome. 2015; J Hered, 106(5):666–671.

Gartler SM, Riggs AD. Mammalian X-chromosome inactivation. 1983; Annu Rev Genet, 17(1):155–190.

Geens M, De Sousa Lopes SMC. X chromosome inactivation in human pluripotent stem cells as a model for human development: back to the drawing board? 2017; Hum Reprod Update, 23(5):520–532.

Gendrel A-V, Attia M, Chen C-J, Diabangouaya P, Servant N, Barillot E, Heard E. Developmental dynamics and disease potential of random monoallelic gene expression. 2014; Dev Cell, 28(4):366–380.

Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Kaplan N, Chang HY, Heard E, Dekker J. Structural organization of the inactive X chromosome in the mouse. 2016; Nature, 535(7613):575–579.

Goto Y, Kimura H. Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. 2009; Nucleic Acids Res, 37(22):7416–7428.

Gribnau J, Luikenhuis S, Hochedlinger K, Monkhorst K, Jaenisch R. X chromosome choice occurs independently of asynchronous replication timing. 2005; J Cell Biol, 168(3):365–373.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N, Goodfellow P, Lovell-Badge R. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. 1990; Nature, 346(6281):245–250.

Guo G, Meyenn von F, Rostovskaya M, Clarke J, Dietmann S, Baker D, Sahakyan A, Myers S, Bertone P, Reik W, et al. Epigenetic resetting of human pluripotency. 2017; Development, 144(15):2748–2763.

Hamvas RM, Zinn A, Keer JT, Fisher EM, Beer-Romero P, Brown SD, Page DC. Rps4 maps near the inactivation center on the mouse X chromosome. 1992; Genomics, 12(2):363–367.

Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. 2014; Nat Methods, 11(2):122–123.

Hickmott JW, Chen C-Y, Arenillas DJ, Korecki AJ, Lam SL, Molday LL, Bonaguro RJ, Zhou M, Chou AY, Mathelier A, et al. PAX6 MiniPromoters drive restricted expression from rAAV in the adult mouse retina. 2016; Mol Ther Methods Clin Dev, 3(10):16051.

Hook EB, Warburton D. Turner syndrome revisited: review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. 2014; Hum Genet, 133(4):417–424.

Horvath LM, Li N, Carrel L. Deletion of an X-inactivation boundary disrupts adjacent gene silencing. 2013; PLoS Genet, 9(11):e1003952.

Ishikawa K, Makanae K, Iwasaki S, Ingolia NT, Moriya H. Post-Translational Dosage Compensation Buffers Genetic Perturbations to Stoichiometry of Protein Complexes. 2017; PLoS Genet, 13:e1006554.

Jegalian K, Page DC. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. 1998; Nature, 394(6695):776–780.

Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A, Janecke AR, Tariverdian G, Chelly J, Fryns J-P, et al. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. 2004; Am J Hum Genet, 76(2):227–236.

Jiang J, Jing Y, Cost GJ, Chiang J-C, Kolpa HJ, Cotton AM, Carone DM, Carone BR, Shivak DA, Guschin DY, et al. Translating dosage compensation to trisomy 21. 2013; Nature, 500(7462):296–300.

Jonkers I, Monkhorst K, Rentmeester E, Grootegoed JA, Grosveld F, Gribnau J. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. 2008; Mol Cell Biol, 28(18):5583–5594.

Korecki AJ, Hickmott JW, Lam SL, Dreolini L, Mathelier A, Baker O, Kuehne C, Bonaguro RJ, Smith J, Tan C-V, et al. Twenty-Seven Tamoxifen-Inducible iCre-

Driver Mouse Strains for Eye and Brain; Including Seventeen Carrying a New Inducible-First Constitutive-Ready Allele. 2019; Genetics, 211(4):1155-1177.

Kucera KS, Reddy TE, Pauli F, Gertz J, Logan JE, Myers RM, Willard HF. Allele-specific distribution of RNA polymerase II on female X chromosomes. 2011; Hum Mol Genet, 20(20):3964–3973.

Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. 1999; Science, 286(5441):964–967.

de Leeuw CN, Dyka FM, Boye SL, Laprise S, Zhou M, Chou AY, Borretta L, McInerny SC, Banks KG, Portales-Casamar E, et al. Targeted CNS Delivery Using Human MiniPromoters and Demonstrated Compatibility with Adeno-Associated Viral Vectors. Mol Ther Methods Clin Dev, 2014;1:5.

Lessing D, Dial TO, Wei C, Payer B, Carrette LLG, Kesner B, Szanto A, Jadhav A, Maloney DJ, Simeonov A, et al. A high-throughput small molecule screen identifies synergism between DNA methylation and Aurora kinase pathways for X reactivation. 2016; Proc Natl Acad Sci USA, 113(50):14366–14371.

Li N, Carrel L. Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. 2008; Proc Natl Acad Sci USA, 105(44):17055–17060.

Lingenfelter PA, Adler DA, Poslinski D, Thomas S, Elliott RW, Chapman VM, Disteche CM. Escape from X inactivation of Smcx is preceded by silencing during mouse development. 1998; Nat Genet, 18(3):212–213.

Loda A, Brandsma JH, Vassilev I, Servant N, Loos F, Amirnasr A, Splinter E, Barillot E, Poot RA, Heard E, et al. Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. 2017; Nat Commun, 8(1):690.

Lyon MF. Epigenetic inheritance in mammals. 1993; Trends Genet, 9(4):123–8.

Lyon MF. X-chromosome inactivation: a repeat hypothesis. 1998; Cytogenet Cell Genet, 80(1-4):133–137.

Marks H, Kerstens HHD, Barakat TS, Splinter E, Dirks RAM, van Mierlo G, Joshi O, Wang S-Y, Babak T, Albers CA, et al. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. 2015; Genome Biol, 16(1):149.

de Mello JCM, de Araújo ESS, Stabellini R, Fraga AM, de Souza JES, Sumita DR, Camargo AA, Pereira LV. Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. 2010; PLoS One, 5(6):e10947.

Migeon BR, Axelman J, Beggs AH. Effect of ageing on reactivation of the human X-linked HPRT locus. 1988; Nature, 335(6185):93–96.

Min B, Park JS, Jeon K, Kang Y-K. Characterization of X-Chromosome gene expression in bovine blastocysts derived by in vitro fertilization and somatic cell nuclear transfer. 2017; Front Genet, 8:42.

Minajigi A, Froberg J, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W, et al. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. 2015; Science, 349(6245):aab2276.

Mugford JW, Starmer J, Williams RL, Calabrese JM, Mieczkowski P, Yee Della, Magnuson T. Evidence for local regulatory control of escape from imprinted X chromosome inactivation. 2014; Genetics, 197(2):715–723.

Nadaf Al S, Deakin JE, Gilbert C, Robinson TJ, Graves JAM, Waters PD. A cross-species comparison of escape from X inactivation in Eutheria: implications for evolution of X chromosome inactivation. 2011; Chromosoma, 121(1):71–78.

Naqvi S, Bellott DW, Lin KS, Page DC. Conserved microRNA targeting reveals preexisting gene dosage sensitivities that shaped amniote sex chromosome evolution. 2018; Genome Res, 28:474–483.

Okamoto I, Patrat C, Thépot D, Peynot N, Fauque P, Daniel N, Diabangouaya P, Wolf J-P, Renard J-P, Duranthon V, et al. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. 2011; Nature, 472(7343):370–4.

Omoe K, Endo A. Relationship between the monosomy X phenotype and Y-linked ribosomal protein S4 (Rps4) in several species of mammals: a molecular evolutionary analysis of Rps4 homologs. 1996; Genomics, 31(1):44–50.

Park C, Carrel L, Makova KD. Strong purifying selection at genes escaping X chromosome inactivation. 2010; Mol Biol Evol, 27(11):2446–2450.

Patrat C, Okamoto I, Diabangouaya P, Vialon V, Le Baccon P, Chow J, Heard E. Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. 2009; Proc Natl Acad Sci USA, 106(13):5198–5203.

Peeters SB, Cotton AM, Brown CJ. Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. 2014; Bioessays, 36(8):746–756.

Peeters SB, Korecki AJ, Baldry SEL, Yang C, Tosefsky K, Balaton BP, Simpson EM, Brown CJ. How do genes that escape from X-chromosome inactivation contribute to Turner syndrome? 2019; Am J Med Genet C Semin Med Genet, 181(1):28-35.

Peeters SB, Korecki AJ, Simpson EM, Brown CJ. Human cis-acting elements regulating escape from X-chromosome inactivation function in mouse. 2018; Hum Mol genet, 27:1252–1262.

Peeters SB, Yang C, Brown CJ. Have humans lost control: The elusive X-controlling element. 2016; Semin Cell Dev Biol, 56:71–77.

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. 2013; Nat methods, 10(10):973–976.

Perry J, Palmer S, Gabriel A, Ashworth A. A short pseudoautosomal region in laboratory mice. 2001; Genome Res, 11(11):1826–1832.

Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GAB. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. 2012; Proc Natl Acad Sci USA, 109(14):5346-51.

Portales-Casamar E, Swanson DJ, Liu L, de Leeuw CN, Banks KG, Sui SJH, Fulton DL, Ali J, Amirabbasi M, Arenillas DJ, et al. A regulatory toolbox of MiniPromoters to

drive selective expression in the brain. 2010; Proc Natl Acad Sci USA, 107(38):16589–16594.

Prasad ML, Pellegata NS, Kloos RT, Barbacioru C, Huang Y, la Chapelle de A. CITED1 protein expression suggests Papillary Thyroid Carcinoma in high throughput tissue microarray-based study. 2004; Thyroid, 14(3):169–175.

Ramos-Ibeas P, Sang F, Zhu Q, Tang WWC, Withey S, Klisch D, Wood L, Loose M, Surani MA, Alberio R. Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. 2019; Nat Commun, 10(1):500.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. 2013; Nat Protoc, 8(11):2281–2308.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. 2014; Cell, 159(7):1665–1680.

Raudsepp T, Das PJ, Avila F, Chowdhary BP. The pseudoautosomal region and sex chromosome aneuploidies in domestic species. 2011; Sex Dev, 6(1-3):72–83.

Reinius B, Shi C, Hengshuo L, Sandhu KS, Radomska KJ, Rosen GD, Lu L, Kullander K, Williams RW, Jazin E. Female-biased expression of long non-coding RNAs in domains that escape X-inactivation in mouse. 2010; BMC Genomics, 11(1):614.

Sado T. What makes the maternal X chromosome resistant to undergoing imprinted X inactivation? 2017; Philos Trans R Soc Lond B Biol Sci, 372(1733).

Sahakyan A, Kim R, Chronis C, Sabri S, Bonora G, Theunissen TW, Kuoy E, Langerman J, Clark AT, Jaenisch R, et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. 2016; Cell Stem Cell, 20(1):87–101.

Scandaglia M, Lopez-Atalaya JP, Medrano-Fernandez A, Lopez-Cascales MT, Del Blanco B, Lipinski M, Benito E, Olivares R, Iwase S, Shi Y, Barco A. Loss of Kdm5c Causes Spurious Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. 2017; Cell Rep, 21(1):47-59.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. 2012; Cell, 148(1-2):335–348.

Schmouth J-F, Bonaguro RJ, Corso-Diaz X, Simpson EM. Modelling human regulatory variation in mouse: finding the function in genome-wide association studies and whole-genome sequencing. 2012; PLoS Genet, 8(3):e1002544.

Schmouth J-F, Castellarin M, Laprise S, Banks KG, Bonaguro RJ, McInerny SC, Borretta L, Amirabbasi M, Korecki AJ, Portales-Casamar E, et al. Non-coding-regulatory regions of human brain genes delineated by bacterial artificial chromosome knock-in mice. 2013; BMC Biol, 11:106.

Schulz EG, Meisig J, Nakamura T, Okamoto I, Sieber A, Picard C, Borensztein M, Saitou M, Blüthgen N, Heard E. The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. 2014; Cell Stem Cell, 14(2):203–216.

Sharman GB. Late DNA replication in the paternally derived X chromosome of female kangaroos. 1971; Nature, 230(5291):231–2.

Sheardown S, Norris D, Fisher A, Brockdorff N. The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. 1996; Hum Mol Genet, 5(9):1355–1360.

Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. Transcription factor trapping by RNA in gene regulatory elements. 2015; Science, 350(6263):978–981.

Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. 2013; Nature, 504(7480):465–469.

Simpson EM, Linder CC, Sargent EE, Davisson MT, Mobraaten LE, Sharp JJ. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. 1997; Nat Genet, 16(1):19–27.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. 1990; Nature, 346(6281):240–244.

Snell DM, Turner JMA. Sex Chromosome Effects on Male-Female Differences in Mammals. 2018; Curr Biol, 28(22):R1313–R1324.

Tagaki N, Sasaki M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. 1975; Nature, 256(5519):640–642.

Tannan NB, Brahmachary M, Garg P, Borel C, Alnefaie R, Watson CT, Thomas NS, Sharp AJ. DNA methylation profiling in X;autosome translocations supports a role for L1 repeats in the spread of X chromosome inactivation. 2013; Hum Mol Genet, 23(5):1224–1236.

Tsofack SP, Meunier L, Sanchez L, Madore J, Provencher D, Mes-Masson A-M, Lebel M. Low expression of the X-linked ribosomal protein S4 in human serous epithelial ovarian cancer is associated with a poor prognosis. 2013; BMC Cancer, 13:303.

Tsuchiya KD, Greally JM, Yi Y, Noel KP, Truong J-P, Disteche CM. Comparative sequence and X-inactivation analyses of a domain of escape in human Xp11.2 and the conserved segment in mouse. 2004; Genome Res, 14(7):1275–1284.

Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, et al. Landscape of X chromosome inactivation across human tissues. 2017; Nature, 550(7675):244–248.

Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, Fryns J-P, Kleefstra T, Raynaud M, Turner G, Ropers H-H, et al. Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. 2006; Hum Mutat, 27(4):389.

Wake N, Takagi N, Sasaki M. Non-random inactivation of X chromosome in the rat yolk sac. 1976; Nature, 262(5569):580–1.

Wang J, Syrett CM, Kramer MC, Basu A, Atchison ML, Anguera MC. Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for

increased expression from the inactive X. 2016; Proc Natl Acad Sci USA, 113(14):E2029–38.

Wang X, Miller DC, Clark AG, Antczak DF. Random X inactivation in the mule and horse placenta. 2012; Genome Res, 22(10):1855–63.

Wang Z, Willard HF, Mukherjee S, Furey TS. Evidence of influence of genomic DNA sequence on human X chromosome inactivation. 2006; PLoS Comput Biol, 2(9):e113.

Wareham KA, Lyon MF, Glenister PH, Williams ED. Age related reactivation of an X-linked gene. 1987; Nature, 327(6124):725–727.

Watanabe M, Zinn AR, Page DC, Nishimoto T. Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. 1993; Nat Genet, 4(3):268–271.

Wu H, Luo J, Yu H, Rattner A, Mo A, Wang Y, Smallwood PM, Erlanger B, Wheelan SJ, Nathans J. Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. 2014; Neuron, 81(1):103–19.

Xia E, Wang Y, Bhandari A, Niu J, Yang F, Yao Z, Wang O. CITED1 gene promotes proliferation, migration and invasion in papillary thyroid cancer. 2018; Oncol Lett, 16(1):105–112.

Xu J, Deng X, Disteche CM. Sex-specific expression of the X-linked histone demethylase gene Jarid1c in brain. 2008; PLoS One, 3(7):e2553.

Xue F, Tian XC, Du F, Kubota C, Taneja M, Dinnyes A, Dai Y, Levine H, Pereira LV, Yang X. Aberrant patterns of X chromosome inactivation in bovine clones. 2002; Nat Genet, 31(2):216–20.

Yang C, Chapman AG, Kelsey AD, Minks J, Cotton AM, Brown CJ. X-chromosome inactivation: molecular mechanisms from the human perspective. 2011; Hum Genet, 130(2):175–185.

Yang C, McLeod AJ, Cotton AM, de Leeuw CN, Laprise S, Banks KG, Simpson EM, Brown CJ. Targeting of >1.5 Mb of human DNA into the mouse X chromosome reveals presence of cis-acting regulators of epigenetic silencing. 2012; Genetics, 192(4):1281–1293.

Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, Moore JM, Filippova GN, Xu J, Liu Y, et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. 2015; Genome Biol,16:52.

Yang GS, Banks KG, Bonaguro RJ, Wilson G, Dreolini L, de Leeuw CN, Liu L, Swanson DJ, Goldowitz D, Holt RA, et al. Next generation tools for high-throughput promoter and expression analysis employing single-copy knock-ins at the Hprt1 locus. 2009; Genomics, 93(3):196–204.

Yen ZC, Meyer IM, Karalic S, Brown CJ. A cross-species comparison of X-chromosome inactivation in Eutheria. 2007; Genomics, 90(4):453–463.

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. An efficient recombination system for chromosome engineering in Escherichia coli. 2000; Proc Natl Acad Sci USA, 97(11):5978–5983.

Zinn AR, Alagappan RK, Brown LG, Wool I, Page DC. Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. 1994; Mol Cell Biol, 14:2485–2492.

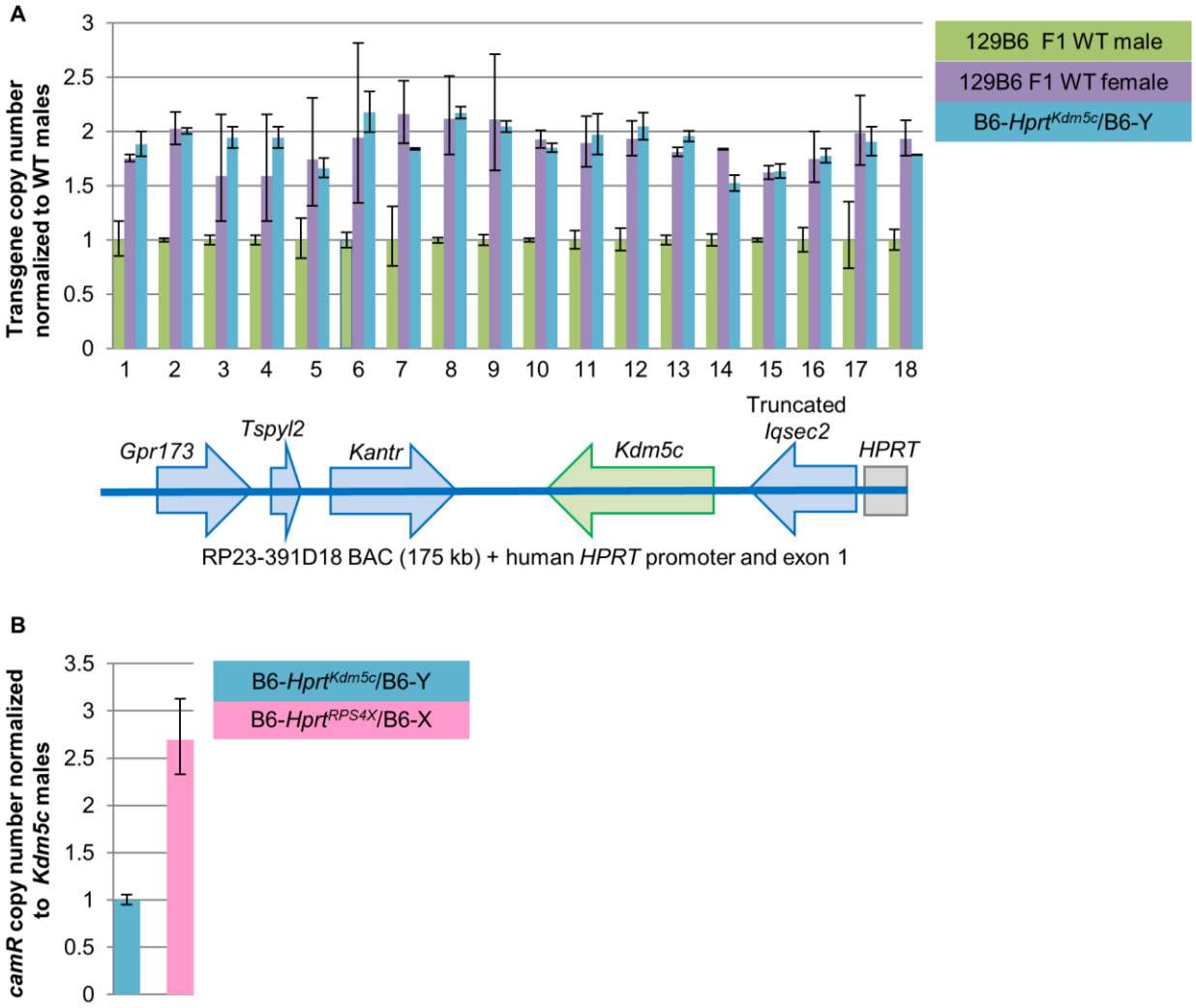
Zvetkova I, Apedaile A, Ramsahoye B, Mermoud JE, Crompton LA, John R, Feil R, Brockdorff N. Global hypomethylation of the genome in XX embryonic stem cells. 2005; Nat Genet, 37(11):1274–1279.

Żylicz JJ, Bousard A, Žumer K, Dossin F, Mohammad E, da Rocha ST, Schwalb B, Syx L, Dingli F, Loew D, et al. The Implication of Early Chromatin Changes in X Chromosome Inactivation. 2019; Cell, 176(1-2):182–197.e23.

Appendices

Appendix A Supplementary materials for Chapter 2

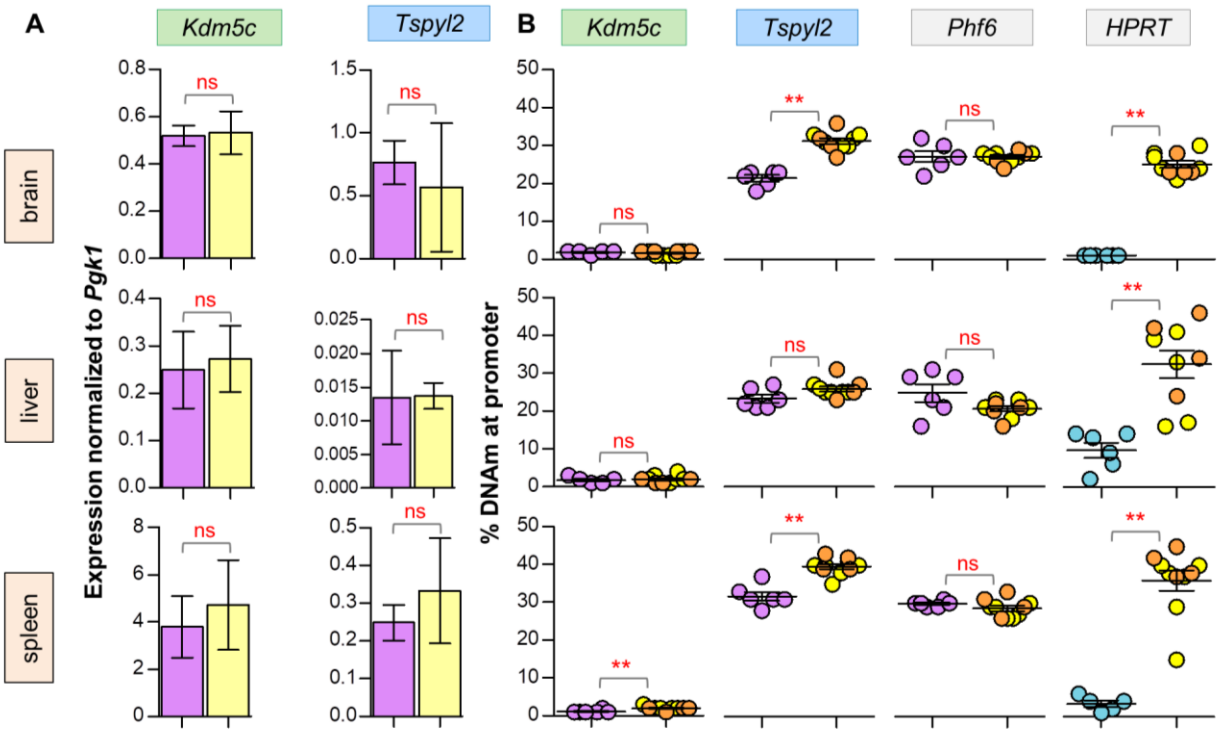
A.1 Supplementary Figures



Supplementary Figure 2.1 Copy number analyses suggest single-copy integration of RP23-391D18 and double integration of RP11-1145H7

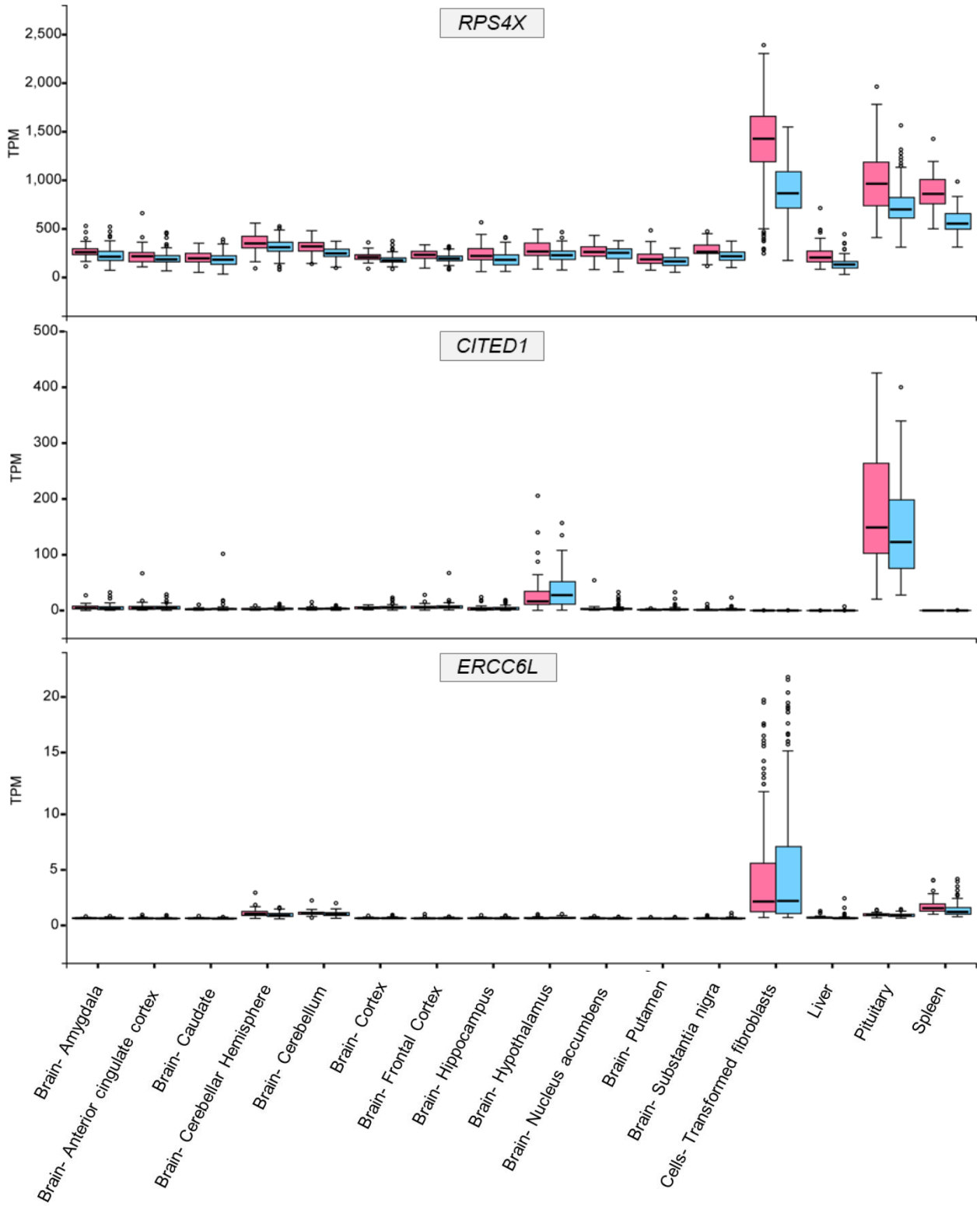
A) 18 qPCR assays approximately every 10 kb along the *Kdm5c*-containing BAC were generated to test for copy number of the integration. N2 transgenic males with an

endogenous and transgenic copy on their Xa were compared to wild-type males with one endogenous copy, and wild-type females with two endogenous copies. The N2 transgenic males more closely resembled wild-type females with two copies suggesting that indeed only one copy of the transgene had integrated. Each bar is an average of two mice per genotype. **B)** N1 females with the *RPS4X* transgene were tested for copy number by comparing the copy number of a gene, *camR*, in the construct that is common between both the *RPS4X* and *Kdm5c* BACs. As the *Kdm5c* transgene was determined to be single-copy in **A)**, six *RPS4X* females were normalized to *Kdm5c* males which indicated at least two copies of the *RPS4X* transgene had been integrated.



Supplementary Figure 2.2 Analysis of additional 129-*Xist*^{1lox}/B6-*Hprt*^{Kdm5c} mice supports escape of *Kdm5c* by DNAm

A) Normalized to *Pgk1*, RT-qPCR of *Kdm5c* and *Tspyl2* expression in average of six original wild-type females (purple) compared to average of six original Xi knock-in females (analyzed in Figure 2.2) plus four additional Xi knock-in females (yellow). **B)** Average DNAm of *Kdm5c* shows promoter hypomethylation in four additional knock-in females (orange) consistent original six Xi knock-in females (yellow) also plotted. *Tspyl2*, *HPRT* and *Phf6* also shown (Mann-Whitney t-test, significance denoted by asterisks; p-value <0.001 ***, 0.001 to 0.01 **, 0.01 to 0.05 *, >0.05 ns). Knock-in females are compared to wild-type females for all assays except *HPRT* (compared to knock-in males, blue) as wild-type females do not carry the human gene.



Supplementary Figure 2.3 RP11-1145H7 region gene expression and DNAm in humans

RPS4X is broadly expressed in all human tissues corresponding to the tissues examined in mouse. *CITED1* and *ERCC6L* are limited in their expression for the tissues examined. Males shown in blue, females with random XCI in pink. Expression values are shown in TPM (Transcripts Per Million), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range (Processed data from The Genotype-Tissue Expression Project, GTEx, gtexportal.org).

A.2 Supplementary Tables

Supplementary Table 2.1 Primer information

All PCR and qPCR primers were designed using the PrimerQuest Tool (IDT) and tested for specificity by In-Silico PCR (UCSC). All pyrosequencing primers were designed using PyroMark Assay Design software (Qiagen). Primers biotinylated at the 5' end are indicated with an asterisk (*).

Experiment	Assay	Sequence
BAC retrofit junction PCR	3'retrofit_F1	AAGCCTTCGCGAAAGAAAAT
	3'retrofit_R2	AACCTGGGTAACATGGTGAGA
	5'retrofit_F1	TCCTTAAGCCCCCACTAGGT
	5'retrofit_R1	TGTCCTTTGTTACAGGCCAGA
Hprt integration and BAC backbone vector	5'CamR oEMS4863	TGGCAATGAAAGACGGTGAGC
	5'CamR oEMS4864	GCATCAGCACCTTGTCGCCT
	Hprt Correction oEMS2267	TCAGGCGAACCTCTCGGCTT
	Hprt Correction oEMS2269	TGCTGGACATCCCTACTAACCCA
	Hprt WT oEMS2236	TCCAGGGAATCCATCACAAT
	Hprt WT oEMS2238	CCAGTGCTGACGTTACAAGC
	Hprt Null oEMS2236	TCCAGGGAATCCATCACAAT
	Hprt Null oEMS2240	GGCATCCAGTGCTCTTCACT
RP23-391D18 BAC copy-number qPCR	qRP23-391D18_F1	CTACTTCTGGGCAGGTGGTC
	qRP23-391D18_R1	GGAATGCCACTTCCTAGCCT
	qRP23-391D18_F2	CACCTGTCTTCACTTCCCCA
	qRP23-391D18_R2	CAGCCCATAAGCCAGGTGTA
	qRP23-391D18_F3	CAGACTCCCATGTGCAAGGA
	qRP23-391D18_R3	TGCTACACACATTCCTGACCA
	qRP23-391D18_F4	CGCTCCCTTTGTTGTCAGTG
	qRP23-391D18_R4	ACCGATGCCAATGAGAACCT
	qRP23-391D18_F5	TGGAGGTTAGTTGCCAGCAT
	qRP23-391D18_R5	CTGGAGGAGGAGTTACAGGC
	qRP23-391D18_F6	CTTTCACTGGTCTGGAGCGA
	qRP23-391D18_R6	CGGGGCTCAATAATGGCTTG
	qRP23-391D18_F7	CATAGCCAAACATGCCCCAG
	qRP23-391D18_R7	AGAGAGTACAAAGGCTGGCC
	qRP23-391D18_F8	CAGTCGGAGGAGTCTGTGTT
	qRP23-391D18_R8	CCCCATCCCAACCTGTTACA

	qRP23-391D18_F9	TCACCTCCTTCTTGAGCTCC
	qRP23-391D18_R9	CTGCCACTTTGCTGCTCA
	qRP23-391D18_F10	GTGTGAGTTTCCAGACCTGC
	qRP23-391D18_R10	ACTGTTCCATCATGTCCGCT
	qRP23-391D18_F11	GAGGACCAGGATGGCTTAGAA
	qRP23-391D18_R11	CCGGGAAGTCAGAGTAGAGAAA
	qRP23-391D18_F12	GATCCTGGCCACTTTCCTCA
	qRP23-391D18_R12	TCTGCAGGAAACGACCCAG
	qRP23-391D18_F13	AACTGTGGAGTATGGGGCTG
	qRP23-391D18_R13	GGAAACCGCTGCCAAATTCT
	qRP23-391D18_F14	GAGGAAGGAGCAGGGATGAG
	qRP23-391D18_R14	ACACCTTCCATCTGAACCCC
	qRP23-391D18_F15	AGCTGTACGGTGTTAGTGGT
	qRP23-391D18_R15	CAGATCCAACCTGCCTCTGT
	qRP23-391D18_F16	ACAGCTTCCGTCAGTCCTTT
	qRP23-391D18_R16	GCTAGGGTTCAAGAGGGGAC
	qRP23-391D18_F17	TCAGGGCCTATGTCTGAGGA
	qRP23-391D18_R17	TGTGGCTGAGCTGTCTTCTT
	qRP23-391D18_18F	GGGAGTGGGATACGAAGAGAA
	qRP23-391D18_18R	CCTTACTGTCCCTCCCTGAATA
	qHbb-bs_F (control)	CTGCTCACACAGGATAGAGAGGG
	qHbb-bs_R (control)	GCAAATGTGAGGAGCAACTGATC
	qCamR_F (BAC backbone)	TCCCAATGGCATCGTAAAGAA
	qCamR_R (BAC backbone)	CAGCTGAACGGTCTGGTTATAG
RP11-1145H7 BAC PCR	RP11-1145H7_F1	CAAACTCTTTCTGTAGCTTG
	RP11-1145H7_R1	TGAAGTCTCATTCTGTCATC
	RP11-1145H7_F2	CTTAACAATGGAGTTTGGAG
	RP11-1145H7_R2	CACAGTCTATCTTTGGATTG
	RP11-1145H7_F3	CTGTCCTCTCAACAAGAAC
	RP11-1145H7_R3	ACATCTGTTGTGTCTAATGC
	RP11-1145H7_F4	ATTGGGACGGTATCCAGTAAGA
	RP11-1145H7_R4	TGGGCTTGAGTCCCTGTAAT
	RP11-1145H7_F5	CAGATTGTAATCACTGAACC
	RP11-1145H7_R5	TTAGTATGGGGTTTCACC
	RP11-1145H7_F6	GCCTGTAATCTCAGCTACTC
	RP11-1145H7_R6	TGTCAGCTGCTTTCTATATC
	RP11-1145H7_F7	ATCATCTCTTTCCCTCATC
	RP11-1145H7_R7	TTTAAGACAGGGTTTCACTC

	RP11-1145H7_F8	ATAGAGTGGTGAACAAGATG
	RP11-1145H7_R8	CATCAACCTTCTGAGTAGC
	RP11-1145H7_F9	TCCTCAGACTAGAGAGAAGG
	RP11-1145H7_R9	AGTACATGTGAGATGGATTG
	RP11-1145H7_F10	CTGTAATCCCAGTTACTCAG
	RP11-1145H7_R10	ATAACAAGTGTTGGTGAGC
	RP11-1145H7_F11	CATATGATCCATCTTGGTC
	RP11-1145H7_R11	CTGTAATCTCAGCACTTTG
	RP11-1145H7_F12	TGAGTTAGAATCAAGACCAG
	RP11-1145H7_R12	TGAGTAGCTGAGACTACAGG
	RP11-1145H7_F13	CAAGAATTGGGTCTAGTTG
	RP11-1145H7_R13	TCTTTCCAGTCCTATATTCC
	RP11-1145H7_F14	CTCCTTGGCTAAGTTTATTC
	RP11-1145H7_R14	AACGATCTCTGTCTCAATG
	RP11-1145H7_F15	CATCTTATGAGTTGTGAAGC
	RP11-1145H7_R15	AATTTAACTGGAGAGTGAGG
	RP11-1145H7_F16	GGTCTTTGATATTGCTTGTC
	RP11-1145H7_R16	GTCCTGACTTTCTACTCTGC
	RP11-1145H7_F17	ACTCAAAGGTAGGAGAACTG
	RP11-1145H7_R17	AATCAGCTCTAAAGTGTTCC
DNA methylation PCR and pyrosequencing	DNAmHPRT_F	GGAATTAGGGAGTTTTTTGAATAGG
	DNAmHPRT_R	*CCTACCAATTTACAAACTCACTAAATA
	DNAmHPRT_S	GGGAGGGAAAGGGGT
	DNAmPhf6_F	*GTGGTTTTTTTTATTGTTAGGGATTTT
	DNAmPhf6_R	GAAATATTGGGATGGGGGTTTT
	DNAmPhf6_S	ATAGAGGTTGGYGATTT
	DNAmKdm5c_F	GTAAGGTTGGGAGTTGATGG
	DNAmKdm5c_R	*CCCATATTCTTCCCACACCTACTA
	DNAmKdm5c_S	GTAAGGTTGGGAGTTGA
	DNAmTspyl2_F	TGAGGGGTAGTTAGTTTGATGA
	DNAmTspyl2_R	*CTCAACCCCTACCTTCTCT
	DNAmTspyl2_S	GGGTAGTTAGTTTGATGATT
	DNAmRPS4X_F	ATTAGTAGATGGTAAGAAAGAGTT
	DNAmRPS4X_R	*CCCAACTCAACCCTTTACT
	DNAmRPS4X_S	AGATGGTAAGAAAGAGTTT
	DNAmERRC6L_F	GGGTAGTTTTTTTTATTTATAATGATGGTA
	DNAmERRC6L_R	*TTCAATCCAATTCAAACCATACTACA
	DNAmERRC6L_S	TTTTTTTTATTTATAATGATGGTAT
	DNAmCITED_F	AAGTGGAATTTATTGGGTAAGTT

	DNAmCITED_R	*CCTAACCAATACCCCACTTCTAAAACTATC
	DNAmCITED_S	GTGGAATTTATTGGGTAAGTTTA
SNP PCR and pyrosequencing	Fln_F	*CCAGCTTCCCTAGTCCAAATGC
	Fln_R	TGCATACAGTCAGTGTCAAGTACAAG
	Fln_S	CCTAGAGAGGGCTGAA
Expression RT-qPCR	qPgk1_F (control)	CGTCTGCCGCGCTGTT
	qPgk1_R (control)	AACACCGTGAGGTCGAAAGG
	qKdm5c_F	GCGACTGGGACTTAACTGTAG
	qKdm5c_R	TCCGTTTCTTCCACACCTTAC
	qTspyl2_F	GGATGACAAGGAGAGTGTGAGG
	qTspyl2_R	TCTGGATGAATTTGCGCTTGAG
	qRPS4X_F	CAAGGTCCGAACTGATATAACCTAC
	qRPS4X_R	GGAAATTCTCTCCCGTCTTGTC
	qCITED_F	GGCAGAATCACTCTCTCCTTCT
	qCITED_R	ACTGGGTCCGAATCGATGATAG
	qERCC6L_F	GGCTGAGGCTGTGGTTTATT
	qERCC6L_R	CCTGGCTAAGAGAACCTGTATTTC

Supplementary Table 2.2 Experimental crosses yield normal sex and genotype ratios

129-*Xist*^{1lox}/X females were crossed to B6-*Hprt*^{BAC}/Y males to generate F1 129-*Xist*^{1lox}/B6-*Hprt*^{BAC} and F1 129-*Xist*^{WT}/B6-*Hprt*^{BAC} experimental females. Neither male to female ratio nor female *Xist*^{WT} to *Xist*^{1lox} is significantly different from expected (chi-square test, 0.05 significance level) suggesting that the additional expression from the BAC genes in females, and specifically the *Xist*^{WT} mice, did not have a deleterious effect on breeding or significantly skew the ratios of offspring.

***Kdm5c* experimental cross**

	male	female	*Xist*^{WT}	*Xist*^{1lox}
Total mice	38	33	18	15
Chi-square		0.352		0.273
P-value		0.553		0.602

***RPS4X* experimental cross**

	male	female	*Xist*^{WT}	*Xist*^{1lox}
Total mice	71	54	26	25
Chi-square		2.312		0.02
P-value		0.128		0.889

Supplementary Table 2.3 Repeat analysis of genes on RP11-1145H7

Repeat analysis of *RPS4X*, *CITED1* and *ERCC6L* for LINE elements predicted to be involved in the spreading of XCI and Alu elements predicted to be involved in escape from XCI. Numbers are percentages of repetitive sequence base pairs out of total base pairs (column headings) analysed, calculated using RepeatMasker (http://www.repeatmasker.org, Institute for Systems Biology).

	DNA sequence features	**TSS +/-5kb**	**Length of transcription**	**TSS +15 kb upstream**	**Whole gene +/-50kb**
RPS4X 4689bp	Alu	11.43	1.04	27.36	23.66
	LINE1	6.53	0	2.77	15.82
	LINE2	7.98	0	6.19	10.11
CITED1 4277bp	Alu	1.29	3.02	15.85	18.46
	LINE1	5.62	0	24.35	16.32
	LINE2	0.49	1.15	6.27	9.75
ERCC6L 34352 bp	Alu	33.96	22.37	25.7	25.16
	LINE1	5.32	11.85	12.33	12.79
	LINE2	13.78	7.66	9.87	9.81

Appendix B Supplementary materials for Chapter 3

B.1 Supplementary Tables

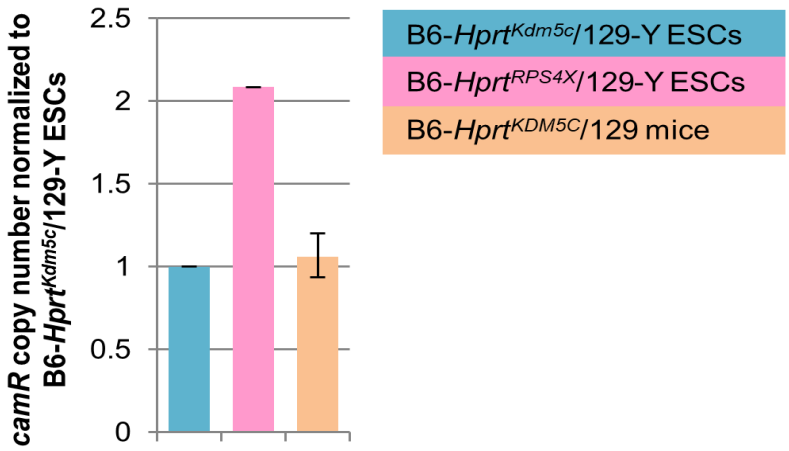
Supplementary Table 3.1 Primer information

All PCR and qPCR primers were designed using the PrimerQuest Tool (IDT) and tested for specificity by In-Silico PCR (UCSC).

Experiment	Assay	Sequence
Expression RT-qPCR	qRps4x_F	CAAATCAATCTGAATGGTGTCGTTC
	qRps4x_R	CTCGTACTATTCGGTACCCTGAT

Appendix C Supplementary materials for Chapter 4

C.1 Supplementary Figures



Supplementary Figure 4.1 Copy number analyses suggest single-copy integration of RP11-236P24

Transgenic mice carrying *Kdm5c* and *RPS4X* BACs were previously assessed for copy number by qPCR and determined to have one and two copies integrated at *Hprt* respectively (Peeters et al, 2018). Based on this information, five transgenic mice carrying *KDM5C* BACs were tested by comparing the copy number of a gene, *camR*, in the construct that is common between all three BACs. Using the ESC lines that the *Kdm5c* and *RPS4X* mice had originated from as controls, and normalizing to *Kdm5c* as single copy, data suggests that only a single copy of the *KDM5C* BAC has integrated at *Hprt*.

C.2 Supplementary Tables

Supplementary Table 4.1 CanEuCre constructs

List of BACs integrated at *Hprt* and chromosome location in humans, construct named by gene-promoter of interest. Constructs shaded grey were assessed by DNAm at promoter-specific assays. *KDM5C* is highlighted blue as it was the only gene pursued for further study.

Construct	**Chromosome**	**BAC #**
AGTR1	3	RP11-487M24
AMOTL1	11	RP11-936P10
CARTPT	5	RP11-88J2
CLDN5	22	RP11-1107K6
CLVS2/RLBP1L2	6	RP11-74B13
CRH	8	RP11-1006F7
GABRA6	5	RP11-631F18
HTR1B	6	RP11-990K4
KCNA4	11	RP11-63H13
KDM5C	X	RP11-236P24
MKI67	10	RP11-636J22
NEUROD6	7	RP11-1087H14
NKX6-1	4	RP11-876O24
NOV	8	RP11-840I14
NPY2R	4	RP11-937K11
NR2E1	6	RP11-144P8
POU4F2	4	RP11-72H2
SLITRK6	13	RP11-398A22
SOX3	X	RP11-1019M14
SPRY1	4	RP11-1126D8

Supplementary Table 4.2: Primer information

All PCR and qPCR primers were designed using the PrimerQuest Tool (IDT) and tested for specificity by In-Silico PCR (UCSC). All pyrosequencing primers were designed using PyroMark Assay Design software (Qiagen). Primers biotinylated at the 5' end are indicated with an asterisk (*).

Experiment	Assay	Sequence
RP11-236P24 BAC PCR	RP11-236P24_F1	CAGGCATATCACAGGTTCAGTTCC
	RP11-236P24_R1	CAGTCAGCAACCATCAACATCG
	RP11-236P24_F2	TGTTGTCACTTGGGCATTTGG
	RP11-236P24_R2	ACCCCTACTCTCACGCTATACACAA
	RP11-236P24_F3	TGATTGGGTATGTGAAGGTGGG
	RP11-236P24_R3	AGAATCCAGGCTCCACACCG
	RP11-236P24_F4	TTGGCTTTTGGCAGTAAGGTTC
	RP11-236P24_R4	TCCAAGAAACACCACGCACAG
	RP11-236P24_F5	GCATTGGGATTGGGGACTTTG
	RP11-236P24_R5	TTGCTCGGGCTGGAACTGA
	RP11-236P24_F6	CCCATTGTTTTACCTTCCCTCC
	RP11-236P24_R6	TTGAGTTGGAAGCGGTTTGGA
	RP11-236P24_F7	GTGGGAGGGATAGCATTAGGAGA
	RP11-236P24_R7	GCCGACTCCCTACAATGAATCTT
	RP11-236P24_F8	TCGCACGCACAGTTTGTCC
	RP11-236P24_R8	ATACTTCCGCCCTTACTCGCA
	RP11-236P24_F9	ACCTCATCCCAAAACATCCTCAC
	RP11-236P24_R9	TGACCTGGAATAGTCTGCTTGGC
	RP11-236P24_F10	GGAGAACATTTGCCTGTATTGGAC
	RP11-236P24_R10	AGTGAAAGAAGCTCCCCAGTGTC
	RP11-236P24_F11	AACTGTCTCATCAAGCCACGAAG
	RP11-236P24_R11	GCTATGTGACCAGGCAGTTGTTCT
	RP11-236P24_F12	CAGGACAGGAGCAAAGTGGGT
	RP11-236P24_R12	GCCCAGTCATTCCCTCTCTTGTC
	RP11-236P24_F13	AAATCGGTTCTCCAGTTCAGGG
	RP11-236P24_R13	AATACGGCATAGTGGCGTGTG
	RP11-236P24_F14	CAACACTACCTCAGCCCTCCATC
	RP11-236P24_R14	TGGTGGCAGAAGAACACAGGAT
	RP11-236P24_F15	GTGGAACTGGCTTTAGGGAACC
	RP11-236P24_R15	TGGTGGGGTTGGGAGAATGT
	RP11-236P24_F16	AGGTGCTGTTGGATGGTCTTCA

	(iCre)	
	RP11-236P24_R16 (iCre)	GACGAAGAGGCGGTAGCAGTAGA
	RP11-236P24_F17	ACGGACAGAAGGACTCAGGCA
	RP11-236P24_R17	CCAGTCTATGCCCTTTCTCCAGT
	RP11-236P24_F18	TATGGGCTATTGAGGTGGACAGG
	RP11-236P24_R18	GGCAAAGGCAAGGTAGCAGATAAG
SNP PCR and pyrosequencing	Zfx_F	*GGCAACGATAAGGCAGCATATT
	Zfx_R	TGTGGTGGAAGTCAGTTGTGAGTC
	Zfx_S	CTGAGCATACCTGATCTCAC
	Taf1_F	ACTCCTTTAATGGCATTTATTGCT
	Taf1_R	*CTTTGTTCCAGTCTGCCAATTA
	Taf1_S	TGTTGTATAAACATGCCA
	Iqsec2rs29293593_F	AACTCGCTTATACCAGAGACACCT
	Iqsec2rs29293593_R	*ACTCTCAGTACCCCAGCATAGG
	Iqsec2rs29293593_S	GTGGTCTGATCTAGGCCC
	Kdm5cExon8_F	*TTGCCTGCTGCCTCCTTTG
	Kdm5cExon8_R	CACACTTTGGACACCTCCAGACA
	Kdm5cExon8_S	TCCAGACACCTTTCG
	HprtIntron1_F	CCATGCTTGGTATGCTAATGAA
	HprtIntron1_R	*ACATATAATGCAGGCAAAACACTC
	HprtIntron1_S	AATAAGCATGTTTCACCA
Expression RT-qPCR	qKDM5C_F	GGTGTGACGCAACGTATACGA
	qKDM5C_R	CCACAACAAACCGGATCCTT
	qKANTR_F	GCTGCAGCCCGTGGAA
	qKANTR_R	AGAGGCCTACAGCAAGATCTGAA
DNA methylation PCR and pyrosequencing	DNAmKDM5C_F	AGGGGTTTTTTGTAATTAGGTTTTTAGA
	DNAmKDM5C_R	*ATATTCTCTATCCCTTCTCCCTAATCTT
	DNAmKDM5C_S	GGGTTTTTTGTAATTAGGT
	DNAmAMOTL_F	GGGATAAAGGAAGGGATGTTG
	DNAmAMOTL_R	*TCACTAAAACCCTACACTCCACC
	DNAmAMOTL_S	GGAGGGTGTTTGTAGA
	DNAmCARTPT_F	*AATGGTAGAGGGGTGGAAATTTAG
	DNAmCARTPT_R	AACCCCCAAACACCCACTACCATC
	DNAmCARTPT_S	CTCACCAACTCCTAACAATCTATTA
	DNAmCLVS2_F	TGGGGAGGTTGGAAAGTATG
	DNAm CLVS2_R	*CCCCACTCCTCAACAAACTACT
	DNAmCLVS2_S	GGGGAGGTTGGAAAG
	DNAmHTR1B_F	GGTGGGTTAGTTTTTAGTAATTTAGGTT
	DNAmHTR1B_R	*CCAAATTCACAACTAAAACTAAAAATC

	DNAmHTR1B_S	GGTGGGTTAGTTTTTAGTA
	DNAmKCNA4_F	AGGAGTAAATTTTGGAGAAGTGT
	DNAmKCNA4_R	*ACTACACCTCCCATTAAAACT
	DNAmKCNA4_S	GTAAATTTTGGAGAAGTGTT
	DNAmNKX6-1_F	TTAGGAAAAGTGAGGAAGAGAGA
	DNAmNKX6-1_R	*CAACAAAAATATCCAAACCCT
	DNAmNKX6-1_S	TGAGGAAGAGAGAATAGTTA
	DNAmNOV_F	TTGTAGTGTAGGGAGGAGG
	DNAmNOV_R	*ACACCCACAACCAATTACCATAA
	DNAmNOV_S	TGTAGGGAGGAGGGG
	DNAmNPY2R_F	GGTAGAGAGTAAAGGGAGAGA
	DNAmNPY2R_R	*CCTACTATACCACCCCCAAATTTAAT
	DNAmNPY2R_S	AGAGTAAAGGGAGAGAT
	DNAmNR2E1_F	*TTAGGAGTTGGGGGAAAAGTTAA
	DNAmNR2E1_R	AACTAAATCCCCTATAATATCTCCAAAA
	DNAmNR2E1_S	ATCCCCTATAATATCTCCA
	DNAmPOU4F2_F	TTATTTGGGTGGGGTTGAGT
	DNAmPOU4F2_R	*CCCCTCAAACTTAAATCCTTTC
	DNAmPOU4F2_S	GGTGGGGTTGAGTGG
	DNAmSOX3_F	GGAAGGGTAGGTTTATTAAAATGTT
	DNAmSOX3_R	*CCCAACTAAAACCCAAACAAACTATAAAT
	DNAmSOX3_S	AATGTTTAGAGTTAAATTAGTAGGA
CRISPR-Cas9 guide RNAs with PAM sequence	gIqsec2_Intron2	CAGGCTAAGGGCTCCATTTG AGG
	gIqsec2_Intron6	GTGGCTGTAGGCCCAAAGCG TGG
	gHprt_Intron1	ACCTAAGCTAATACGACCTT TGG
	gHprt_Exon2	TAATCATTATGCCGAGGATT TGG
	gHprt_5'	CCACCGCCCGGCTGATTCTA TGG
Deletion characterization PCR	Iqsec2_2a6aF	TCTACCCACACATGCACATATAC
	Iqsec2_2a6aR	CGTCCAAATCCATAGAGGAGAAG
	HprtNull oEMS2240	GGCATCCAGTGCTCTTCACT
	Hprt_Exon2R	GCCAAATCAGTATGTATGCCCC
	Hprt_Exon2F	CTATTCAGCAGTAAGACGCAGC
	Hprt_Intron1F	TGGCAATCACATGAAAGACAATCC

Supplementary Table 4.3 Repeat analysis of genes on RP11-236P24

Repeat analysis of *KDM5C* and *KANTR* for LINE elements predicted to be involved in the spreading of XCI and Alu elements predicted to be involved in escape from XCI. Numbers are percentages of repetitive sequence base pairs out of total base pairs (column headings) analysed, calculated using RepeatMasker (http://www.repeatmasker.org, Institute for Systems Biology).

	DNA sequence features	**TSS +/-5kb**	**Length of transcription**	**TSS +15 kb upstream**	**Whole gene +/-50kb**
KDM5C *34102 bp*	Alu	2.92	11.38	4.8	14.1
	LINE1	6.18	8.02	4.18	11.4
	LINE2	6.76	3.02	6.16	5
KANTR *53852 bp*	Alu	24.82	28.85	5.93	20
	LINE1	8.95	39.28	8.12	20.3
	LINE2	10.17	5.24	1.97	2.73

Supplementary Table 4.4 Female ESCs with a smaller *Hprt* deletion have lower recombination frequencies than male cells with the original *Hprt*$^{b\text{-}m3}$ deletion

Numbers indicate how many clones were scored for each construct after HAT selection for BACs (>150 kb, blue) and plasmids (<20 kb, green), N/D= no data. Substrate indicates if the cells were grown on MEFs or gelatin prior to electroporation. Only experiments with data in all three ESC columns were performed at the same time.

ESC Strain	**C57BL/6J XY**	**129B6F1 XX**	**129B6F1 XX**
Hprt Allele	*b-m3* ~35 kb del	CRISPR ~5 kb del	CRISPR ~5 kb del
Substrate	MEFs	MEFs	Gelatin
KDM5C BAC (no stop)	2	N/D	0-2
RPS4X BAC	3-8	N/D	0
Kdm5c BAC	4-8	N/D	N/D
Empty homology plasmid	N/D	N/D	1
CanEuCre plasmid	618	2	0
EmGFP reporter plasmid	690	2	2

Appendix D Supplementary materials for Chapter 5

D.1 Supplementary Figures

Supplementary Figure 5.1 EmGFP reporter plasmid

Design based on previously used constructs (Portales-Casamar et al, 2010; Hickmott et al, 2016) includes an EmGFP reporter, candidate promoter elements, the *Hprt* homology arms and *HPRT* complementary sequence for selection of integration at the *Hprt* docking site in mouse.