

DNA methylation demonstrates spread of X-chromosome inactivation
to human transgenes

by

Christine Yang

B.Sc., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2012

© Christine Yang, 2012

Abstract

X-chromosome inactivation is the process by which mammalian females achieve dosage compensation with males by silencing one of the two X chromosomes in female cells. Despite the chromosome-wide inactivation, a significant proportion of genes on the X chromosome in humans remain expressed on the inactive X chromosome. It has been long hypothesized that the genomic context plays an important role in influencing whether a gene is subject to or escapes from X-chromosome inactivation; however, *cis*-regulatory elements involved in X-chromosome inactivation have not yet been identified.

The objective of this thesis was to identify DNA elements that promote the escape of genes from X-chromosome inactivation in the human genome, through analyzing the X-chromosome inactivation statuses of human transgenes integrated at the *Hprt* locus on the mouse X chromosome and identifying the transgenes that escape from X-chromosome inactivation. DNA methylation was used to assess the inactivation status of 74 human reporter constructs comprising over 1.5 Mb of DNA. Transgenes that show low promoter DNA methylation in males and females would be potential escape genes. Of the 47 genes examined, only the *PHB* gene showed female DNA hypomethylation approaching the level seen in males, and escape from X-chromosome inactivation was verified by the demonstration of expression from the inactive X chromosome in females with non-random X-chromosome inactivation. Analysis on the repeat element content of five BAC-derived transgenes subject to X-chromosome inactivation suggested that local LINE-1 and Alu densities were insufficient to determine whether a gene would be subject to X-inactivation. Interestingly, CpG islands not associated with promoters also showed female-specific DNA hypermethylation, suggesting a dominant effect of X-chromosome inactivation on the regulation of DNA methylation. Different human transgenes show a differential capacity to accumulate DNA methylation when integrated into the identical location on the inactive X chromosome, indicating the presence of additional *cis*-acting epigenetic modulators. As only one of the human transgenes analyzed escaped from X-chromosome inactivation, we conclude that elements involved in ongoing expression from the inactive X are rare in the human genome and that mouse X-chromosome inactivation is very effective in silencing human transgenes.

Preface

The candidate (C. Yang) performed all the experiments and analysis on DNA methylation and the quantitation of expression from the region between genes *PHB* and *HPRT*.

Contribution of the Pleiades Promoter Project

- Designed and created the transgenes

Simpson laboratory

- Generation and breeding of transgenic mice
- Tissue dissections and ear notch sample collection
- *lacZ* and EGFP staining in tissues of transgenic mice to determine the expression status of transgenes

This review article is written by C. Yang and the Brown lab. It is cited in this thesis, but no text from this paper is used in the thesis.

- **Yang C**, Chapman AG, Kelsey AD, Minks J, Cotton AM, and Brown CJ. *X-chromosome inactivation: molecular mechanisms from the human perspective*. Human Genetics, 2011. **130**(2): p. 175-85.

Ethics approval was obtained from the Animal Care Committee at the University of British Columbia for doing research on mice. All the breeding and handling of mice were done by members of the Simpson lab. We received DNA and tissue samples from the Simpson lab and performed experiments on these samples. The ethics protocol numbers are listed below:

For working with mice that carry the Pleiades constructs

- Pleiades Promoter Project (2010): A09-0980
- Breeding: Pleiades Promoter Project (2010): A09-0981

For working with mice carrying an *Xist* deletion

- CanEuCre (Cre-Driver Strains): A10-0267
- Breeding: CanEuCre: A10-0268

Table of contents

Abstract	ii
Preface	iii
Table of contents	iv
List of tables	v
List of figures	vi
List of abbreviations	vii
List of gene names	viii
Acknowledgements	ix
1 Introduction	1
1.1 Thesis overview	2
1.2 Epigenetics and XCI	2
1.2.1 Regulation of genes in domains	3
1.3 DNA methylation	3
1.3.1 Distribution and deposition of DNA methylation	4
1.3.2 DNA methylation and correlation with expression	4
1.3.3 DNA methylation of X-linked promoters	5
1.3.4 Intragenic and intergenic DNA methylation on the X chromosome	6
1.4 Other epigenetic features of the Xa and Xi	7
1.5 Non-random XCI	8
1.5.1 <i>XIST/Xist</i> expression and role in XCI	8
1.5.2 Modifiers of randomness in XCI	8
1.6 Escape from XCI	9
1.6.1 Epigenetic signatures of escape genes	10
1.7 <i>Cis</i> -regulatory elements	11
1.7.1 Escape elements	11
1.7.2 Waystations	12
1.8 XCI status of transgenes	13
1.9 Thesis objective	14
2 Methods	18
2.1 Pleiades Promoter Project constructs	19
2.2 Generation of mice with <i>Xist</i> deletion and MaxiP	19
2.3 DNA and RNA tissue extraction and expression analysis	19
2.4 DNA methylation analysis	20
2.5 Repetitive element content analysis	21
2.6 Statistical analysis	21
3 Results	28
3.1 DNA hypermethylation reflects XCI of Pleiades constructs	29
3.2 DNA methylation of flanking genes reflects both skewing of XCI and differential capacity for DNA methylation on the Xi	29
3.3 <i>PITX2</i> is DNA hypermethylated at transcription start site as well as intragenic and intergenic CpG islands	31
3.4 A truncated gene on <i>NGFR</i> BAC construct partially escapes from XCI	31
3.5 Repeat content of candidate elements involved in XCI on the MaxiPs	32
3.6 All examined MiniP constructs appear to be subject to XCI	32
3.7 DNA methylation of <i>lacZ</i> reporter consistently reflects the pattern at CpG island promoters	33
4 Discussion	49
References	56

List of tables

Table 1.1	Mouse escape genes common across species in somatic cells.....	15
Table 2.1	qPCR primers used to determine relative transcription level of transgenes on the Xi normalized to <i>Pgk1</i>	22
Table 2.2	Pyrosequencing primers and cycling conditions for DNA methylation analysis	23
Table 3.1	Number of full-length LINE-1 on the MaxiPs.	48

List of figures

Figure 3.1	Experimental system in which human reporters (Pleiades constructs) were integrated at the <i>Hprt</i> locus on the mouse X.	35
Figure 3.2	DNA methylation agreed with the XCI status of the MaxiP constructs.	36
Figure 3.3	<i>NR2E1</i> promoter showed consistent DNA methylation pattern across tissues.	37
Figure 3.4	DNA methylation analysis of flanking <i>HPRT</i> and <i>Phf6</i> genes revealed differential susceptibility of MaxiP constructs to DNA methylation on the Xi.	38
Figure 3.5	DNA methylation of <i>HPRT</i> and <i>lacZ</i> was altered for <i>NGFR</i> (Ple133).	39
Figure 3.6	DNA methylation analysis of multiple regions in the <i>PITX2</i> BAC construct.	40
Figure 3.7	<i>PHB</i> on the <i>NGFR</i> BAC partially escaped from XCI.	41
Figure 3.8	Base pair composition of repetitive elements on the MaxiPs.	42
Figure 3.9	DNA methylation reflects XCI status independent of transgene expression.	43
Figure 3.10	Promoter DNA methylation of the MiniP constructs.	44
Figure 3.11	Promoter DNA methylation of genes flanking the MiniPs.	45
Figure 3.12	<i>lacZ</i> reporter exhibited female-specific DNA hypermethylation when present on the X chromosome.	46
Figure 3.13	<i>lacZ</i> DNA methylation can be used as a surrogate for promoter DNA methylation.	47

List of abbreviations

3'UTR = three prime untranslated region
BAC = bacterial artificial chromosome
bp = base pairs
cDNA = complementary deoxyribonucleic acid
CTCF = CCCTC binding factor protein
Cxxc1 = CXXC finger 1 (PHD domain), previously known as Cfp1
DNA = deoxyribonucleic acid
dNTP = deoxyribonucleotide
(E)GFP = (enhanced) green fluorescent protein
ENCODE = Encyclopedia of DNA Elements
ES = embryonic stem
FISH = fluorescent *in situ* hybridization
HC = high CpG density
IC = intermediate CpG density
ICF = immunodeficiency, centromere instability and facial abnormalities
kb = kilo base pairs
LINE = long interspersed nuclear element
MaxiP = MaxiPromoter
Mb = mega base pairs
MiniP = MiniPromoter
(q)PCR = (quantitative) polymerase chain reaction
RNA = ribonucleic acid
SINE = short interspersed nuclear element
UCSC = University of California – Santa Cruz
Xa = active X chromosome
XCI = X-chromosome inactivation
Xi = inactive X chromosome

List of gene names

Genes that are discussed in the body of the thesis are included in this list. Genes in capitalized letters refer to human genes, while genes in lower case refer to mouse genes or bacterial gene (*lacZ*)

AMOTL1 = angiomin like 1

AR = androgen receptor

CARTPT = CART prepropeptide

CCKBR = cholecystokinin B receptor

DCX = doublecortin

DDX3X/Ddx3x = DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked

Dnmt1 = DNA (cytosine-5-)-methyltransferase 1

Dnmt3a = DNA (cytosine-5-)-methyltransferase 3 alpha

DNMT3B/Dnmt3b = DNA (cytosine-5-)-methyltransferase 3 beta

GPX3 = glutathione peroxidase 3 (plasma)

HPRT/Hprt = hypoxanthine guanine phosphoribosyl transferase

ICMT = isoprenylcysteine carboxyl methyltransferase

IQSEC2/Iqsec2 = IQ motif and Sec7 domain 2

KDM5C/Kdm5c = lysine (K)-specific demethylase 5C (previously known as *JARID1C/Jarid1c*)

Kdm6a = lysine (K)-specific demethylase 6A (previously known as *UTX/Utx*)

lacZ = beta-D-galactosidase

LCT = lactase

MAOA = monoamine oxidase A

MCM6 = minichromosome maintenance complex component 6

NGFR = nerve growth factor receptor

NOV = nephroblastoma overexpressed

NR2E1 = nuclear receptor subfamily 2, group E, member 1

NR2F2 = nuclear receptor subfamily 2, group F, member 2

OXT = oxytocin, prepropeptide

PGK1/Pgk1 = phosphoglycerate kinase 1

PHB = prohibitin

Phf6 = PHD finger protein 6

PITX2 = paired-like homeodomain 2

POGZ = pogo transposable element with ZNF domain

Smc1a = structural maintenance of chromosomes 1A

SRY = sex determining region Y

STS = steroid sulfatase (microsomal), isozyme S

Tspy12 = TSPY-like 2

XIST/Xist = Xi-specific transcripts

Acknowledgements

I am thankful to the many people who helped me along in my thesis and made my life as a graduate student both enjoyable and valuable. First of all, I would like to thank my supervisor, my mentor, Dr. Carolyn Brown for her patience, guidance, and enthusiasm throughout my degree. I am also grateful to my committee members Dr. Elizabeth Simpson and Dr. Dixie Mager for their time and advice. Thanks to Dr. Catherine Van Raamsdonk for reading and examining my thesis.

I want to thank all present and past members of the Brown lab and the Molecular Epigenetics Group for their support and humour, with special thanks to Sarah Baldry, Dr. Allison Cotton, and Jakub Minks. I also want to thank members of the Simpson lab for the generation of transgenic mice and sample collection, and interesting insights on my project.

Finally, I want to thank my friends and family for their encouragement. To my aunt Marina and cousin Debbie Chang for always believing in me and supporting me for whatever I do, or plan to do. Importantly, I thank my parents for all the hard work they have been through for me and being there for me.

1 Introduction

1.1 Thesis overview

Gene regulation is important for the proper functioning of cells and therefore the organism as a whole. The mechanisms of regulating gene expression involve differential modifications and recruitment of proteins to genes of different activities. This thesis will be focused on a type of DNA (deoxyribonucleic acid) modification called DNA methylation, in the context of X-chromosome inactivation (XCI), which provides a unique opportunity to study differential regulation of active and inactive genes. The collaboration with the members of the Pleiades Promoter Project, who have integrated many human transgenes of various sizes on the X chromosome in mice, allowed me to analyze DNA methylation and the XCI statuses of these transgenes in this thesis. Overall, studying exogenous DNA on the X chromosome provides insight into the role of regulatory sequences in the mechanism of XCI, but also how the genomic context can influence gene silencing in general.

1.2 Epigenetics and XCI

Epigenetic regulation is a means by which genes exploit the resources in the cell, from modifications of the chromatin to the physical organization of the genome, to achieve differential expression of the genes. The fascination behind epigenetics is the ability of cells to attain different phenotypes with an identical genomic sequence, leading to cellular differentiation and adaptation to the environment. A classic example of epigenetic regulation is XCI, which is essentially a differential regulation of the X chromosomes in the cells of females. By studying how inactivation occurs on the X chromosome, we hope to gain more insight into how genes are regulated in our cells.

In placental mammals, males have one X chromosome and one Y chromosome while females have two X chromosomes. The sex chromosomes were originally a homologous pair of autosomes. Following the acquisition of the sex-determining gene, *SRY*, the Y chromosome has undergone extensive degeneration and become highly differentiated from the X chromosome (1; reviewed in 2). In order to compensate for the dosage difference between the sexes, females transcriptionally silence one of the two X chromosomes (X_i), leaving only one active X chromosome (X_a) per diploid cell (3). Therefore, expression of X-linked genes overall does not differ between males and females (4). In humans, the process of XCI is random so that either of the two X chromosomes could become the X_i , and the inactivation state is maintained through cell divisions (3). Females are therefore mosaic for two cell populations with either the paternal or the maternal X chromosome inactivated. A critical initiator of XCI is an untranslated RNA (ribonucleic acid) called X_i -specific transcripts (*XIST* in humans, *Xist* in mice), which is

exclusively expressed from the Xi and coats the chromosome in *cis* (5-9). *XIST/Xist* will be discussed further in section 1.5.1.

1.2.1 Regulation of genes in domains

In contrast to prokaryotes where co-regulated genes within the same pathway tend to be packaged into a single messenger RNA (also known as operons), non-random organization of genes contribute to the co-regulation of expression in eukaryotes. Chromosomal-wide silencing of the X is one of the most extreme examples of co-regulation of genes in domains, but domain-wide regulation is also evident in other parts of the human genome. For example, housekeeping genes with high breadth of tissue expression show strong clustering in the genome (10). The organization of linked pairs of housekeeping genes is conserved between humans and mice, suggesting that clustering of genes is advantageous (11). Experimentation with transgenes at 90 different integration sites in the genome has also shown that the transgene expression level correlates more strongly with the overall expression of the genes co-localized in the domain than with the expression of genes linked to the transgene at the integration site (12).

The chromatin environment and spatial organization of the genome inside the nucleus may provide a means of regulating multiple genes in a domain. In support of this view, domains with high transgene expression are more often found in the interior of the nucleus and appear to exhibit more open chromatin by FISH (fluorescent *in situ* hybridization) as compared to domains with weakly expressed transgenes (12). In addition, more than 1,300 genomic regions of 0.1-10 mega base pairs (Mb) in human fibroblasts were discovered to interact with the nuclear lamina, which typically is associated with a repressive influence on gene expression (13). Artificial tethering of transgenes to the nuclear lamina leads to repression of the transgene and some endogenous flanking genes (14). Genes within the lamina-associated domains are less active than genes outside of the domains, along with a decreased RNA polymerase II occupancy and reduced histone modifications associated with an active state (13). Interestingly, CCCTC-binding protein (CTCF), which is associated with insulating activity that protects target regions from position effects (reviewed in 15), demarcates the boundaries of the lamina-associated domains implicating a role for boundary elements and insulators in gene regulation in domains (13).

1.3 DNA methylation

DNA methylation is one of the epigenetic modifications important in gene regulation. The majority of DNA methylation occurs on the cytosines of CpG dinucleotides in the human genome, although methylated cytosines in non-CpG contexts are also found in embryonic stem (ES) cells (16). Because methylated cytosine is prone to spontaneous deamination resulting in a

mutation to thymine that is not efficiently repaired (reviewed in 17), DNA methylation is believed to be the cause of the CpG depletion in the highly methylated mammalian genomes (18). In fact, ~70% of the CpGs in the genome of human somatic cells are DNA methylated (16, 19), and a large proportion of methylated cytosines are found in repetitive elements (20). In both human and mouse, DNA methylation has been shown to serve as a defense mechanism to repress expression of repetitive elements such as retrotransposons (21-24), the activation of which can cause deleterious insertions or rearrangements in the genome (25; reviewed in 26).

1.3.1 Distribution and deposition of DNA methylation

DNA methylation is deposited by the DNA methyltransferases (*DNMT* in humans, *Dnmt* in mice). *Dnmt3a* and *Dnmt3b* are the major enzymes that catalyze *de novo* DNA methylation (27, 28), and mutation in *DNMT3B* causes ICF (immunodeficiency, centromere instability and facial abnormalities) syndrome in humans (29). In contrast, *Dnmt1* is responsible for methylating hemi-methylated DNA after DNA replication, so that the pattern of DNA methylation can be maintained in daughter cells (30, 31). Although the bulk of the genome is depleted in CpG dinucleotides, there are regions in the genome called CpG islands that retain a CpG density close to the expected level based on base compositions (32-34), which can be classified into high CpG density (HC) and intermediate CpG density (IC) (35). CpG islands are able to maintain CpG density compared to the rest of the genome because they are generally unmethylated and thus are not subject to the high mutability of methylated cytosines (20, 32), although IC islands are more likely to be DNA methylated than HC (35). The pattern of DNA methylation of CpG islands on the X chromosome is distinct from that of the autosomal CpG islands and will be discussed in section 1.3.3.

1.3.2 DNA methylation and correlation with expression

CpG islands are more frequently associated with promoters of housekeeping genes or widely expressed genes than tissue-specific genes (32, 35-37), suggesting a relationship between DNA methylation and gene expression. Although the majority of CpG islands, particularly the HC promoters, are free of DNA methylation, a small fraction of CpG island promoters are highly DNA methylated (3% of HC promoters, 21% of IC promoters; 35). Furthermore, there are CpG islands that exhibit a tissue-specific DNA methylation pattern (38, 39), although tissue-specific DNA methylation is more commonly found in gene bodies and the flanking region surrounding CpG islands, termed CpG island shores (40, 41). By examining DNA methylation and RNA polymerase II occupancy at 16,000 human promoters, Weber *et al.* (35) observed an inverse correlation between DNA methylation of CpG island promoters and gene activity. However, the IC promoters have a higher frequency of being DNA methylated in the absence of gene activity

than the HC promoters, as most HC promoters remain unmethylated even when the gene is not active. In general, non-CpG island promoters are highly DNA methylated regardless of gene activity (35), although non-CpG island promoters that show a correlation between DNA methylation and gene expression have been identified (42).

Beyond the promoter region, highly expressed genes actually show increased DNA methylation in the gene body compared to weakly expressed genes in different human cell types (43-45), but this paradoxical correlation between intragenic DNA methylation and expression is not observed in the gene body proximal to the promoter (44, 46). Even in the presence of an unmethylated promoter, DNA methylation within the 5' end of the gene body of a transgene is sufficient to repress transcription (47). There have been conflicting results as to whether DNA methylation of the gene body impedes RNA polymerase II elongation (48), but the discrepancy regarding increased intragenic DNA methylation with expression is perhaps due to differential regulation of DNA methylation on CpG-poor and CpG-rich gene bodies and/or due to different effects of DNA methylation depending on the proximity to the promoter.

1.3.3 DNA methylation of X-linked promoters

CpG island promoters on the X chromosome exhibit a unique pattern of DNA methylation compared to autosomal promoters. X-linked CpG island promoters present a female-specific DNA hypermethylation compared to males due to differential DNA methylation on the X_a and X_i. Early studies using DNA methylation-sensitive restriction digests to examine the level of DNA methylation on the X_a and X_i have shown that the 5' end of housekeeping genes such as hypoxanthine phosphoribosyltransferase (*HPRT* in humans, *Hprt* in mice) and phosphoglycerate kinase (*PGK1* in humans, *Pgk1* in mice) are unmethylated on the X_a and hypermethylated on the X_i (49-52). Furthermore, more recent large-scale studies on promoter DNA methylation on the X chromosome support that genes which are subject to XCI are partially DNA methylated at the CpG island promoters in females and unmethylated in males (53, 54). The majority of non-CpG island promoters are DNA methylated on both X_a and X_i, as they are on the autosomes (53).

DNA methylation is a relatively late modification on the X_i as the *Hprt* promoter does not become DNA methylated until several days post XCI; therefore, DNA methylation has been proposed to be the 'lock' in maintaining the inactive state of X-linked genes (55). Consistent with the role of DNA methylation in maintenance of XCI, treatment with the demethylating agent 5-azacytidine could lead to the reactivation of *Hprt* on the X_i (49, 50). DNA methylation is generally consistent between different tissues even for tissue-specific genes (56), suggesting that DNA methylation is not simply a marker for expression but for the XCI status of genes. For

example, the androgen receptor (*AR*) has a tissue-specific expression pattern (57) but still shows female-specific DNA hypermethylation when it is not expressed (58).

1.3.4 Intragenic and intergenic DNA methylation on the X chromosome

DNA methylation of promoters is well characterized on the X chromosome, but DNA methylation levels within and between transcription units on the X_a and X_i are less clear. Global DNA methylation on the X chromosome has been examined in several studies but the interpretations may differ depending on the method of detection. A majority of the studies have shown that the X_a is globally DNA hypermethylated compared to the X_i (59-61), and since promoters constitute only a small part of the X chromosome, global DNA methylation largely reflects intragenic and intergenic DNA methylation. Indeed, gene bodies on the X_a are found to be DNA hypermethylated compared to the X_i (62), but both X chromosomes exhibit some level of DNA methylation in the gene bodies and intergenic regions (53). In contrast, the majority of the CpG islands on the X chromosome, whether they be in promoters, gene bodies or intergenic regions, are DNA hypermethylated on the X_i compared to the X_a (63).

Studies of DNA methylation have disparately concluded either no difference in global DNA methylation between the X_a and the X_i (64) or global DNA hypermethylation of the X_i compared to the X_a (65). However, the apparently conflicting results do not necessarily contradict each other due to differing precision, sensitivity, and/or target regions of the techniques used to examine DNA methylation. The sequence target of the DNA methylation-sensitive enzyme *Hpa* II (5'-CCGG-3') used in the study by Prantera *et al.* (65) is enriched in CpG islands; therefore, the DNA hypomethylation observed on X_a may correspond to promoters where CpG islands are associated. Importantly, Bernardino *et al.* (59) observed relatively high variability in methyl-cytosine staining between autosomes within the same metaphase, which may have resulted in a statistically insignificant difference between X_a and X_i DNA methylation. The patterns of DNA methylation could also vary between metaphase plates depending on the length of digestion for nick translations (64). Therefore, visualization through metaphase chromosomes may not be as effective at detecting differences in DNA methylation between the X_a and the X_i, particularly when both chromosomes exhibit some level of DNA methylation (53). Interestingly, cells with abnormal karyotypes involving the X chromosome showed clearer and statistically significant difference in DNA methylation between the X_a and the X_i (59), suggesting that the type of cells examined is also important in the interpretation of the results. In agreement with the effect of cell types in DNA methylation analysis, different proliferation states of tissues are associated with different patterns of intragenic DNA methylation in the genome (45).

1.4 Other epigenetic features of the Xa and Xi

This thesis is focused on the DNA methylation of promoters on the X chromosome, but there are other epigenetic features that are different between the Xa and the Xi. In general, compared to the Xa, the Xi is enriched with modifications associated with the inactive state and depleted of those associated with the active state (reviewed in 66). The Xi shows a global depletion of the active marks of acetylation of histones H2A, H3 and H4 (67) and H3 lysine 4 methylation (68). Conversely, inactive marks such as H3 lysine 9 methylation, H3 lysine 27 trimethylation and H2A lysine 119 ubiquitination are enriched on the Xi (68); reviewed in 66). At the gene level, however, the individual histone marks are not exclusively found at the active or the inactive genes (69). In addition to modifications of histone tails, high level of the histone variant macroH2A is another feature of the Xi (70).

A classic way of distinguishing between the Xa and the Xi relies on the asynchronous replication of the Xa and Xi, the latter of which replicates late in S phase (66, 68, 71). Several origins of replication on the X chromosome have been mapped to the CpG island promoters of genes, and interestingly, the same origins of replication are used on the Xa and Xi (72). Although the efficiency of replication origin firing is equivalent on the Xa and Xi, the activity of the origins is delayed on the Xi, suggesting that the heterochromatic environment of the Xi influences the timing of replication (72). Although the inactivated genes on the Xi generally replicate later than the respective active alleles on the Xa (72), transcription *per se* is insufficient to alter replication timing as some genes can achieve earlier replication without reactivation following a treatment with 5-azacytidine (73).

Finally, not only are the Xa and Xi different in terms of chromatin structure and replication timing, their spatial organization in the nucleus also distinguishes them. Compared to the Xa, the territory occupied by the Xi is more compact and is known as the Barr body (9; reviewed in 74). The Xist/XIST-coated Xi forms a silent nuclear compartment that is depleted of RNA polymerase II and transcription factors (77, 78). In mice, upon silencing, X-linked genes are translocated from the periphery of or outside the Xist RNA domain to a more internal position within the silent domain (78). Furthermore, in humans, genes on the Xi tend to position at the periphery of the XIST domain regardless of whether they are subject to or escape from XCI (discussed further in section 1.6), while genes on the Xa are more frequently found inside the X territory than genes on the Xi (77). The scaffold attachment factor A, which is implicated in higher order organization of the genome, is found to be concentrated on the Xi territory, and its stable association with the Xi is largely dependent on an interaction with RNA, for which the XIST RNA is a strong candidate (75, 76). Many mysteries of XCI remain to be solved, and each

difference between the Xi and the Xa represents an opportunity for us to understand the mechanism and consequence of XCI.

1.5 Non-random XCI

Although mice and humans generally have random XCI where either X chromosome in a female cell can undergo XCI, mice have imprinted XCI in the extra-embryonic tissues where only the paternal X chromosome is inactivated (79). The randomness of XCI can also be disrupted due to genetic determinants, such as mutation in the regulator of XCI, *XIST/Xist*.

1.5.1 *XIST/Xist* expression and role in XCI

In mice, *Xist* expression is first detected in 4-cell pre-implantation embryos, prior to overt differentiation of the extra-embryonic tissue and XCI (80). By the blastocyst stage, *Xist* is upregulated and continues to be expressed in adult female mice (81). It has also been shown that undifferentiated female mouse ES cells carry two Xa's, and XCI and *Xist* upregulation occurs upon differentiation (81). *XIST* expression has been detected in pre-implantation embryos in humans as well (82, 83), but there have been conflicting results on the timing of XCI in humans. Through RNA FISH, van den Berg *et al.* (84) observed female-specific *XIST* RNA accumulation in human blastocysts that coincided with transcriptional silencing, while a more recent study showed that *XIST* is expressed in both male and female blastocysts with no apparent transcriptional silencing (85). Therefore, more work is required to clarify the timing of XCI in humans.

In addition to the concordance between *Xist* expression and timing of XCI, knock-out experiments have shown that *Xist* is required for both random and imprinted XCI in mice (86, 87). Paternally-inherited *Xist* deletions result in female embryonic lethality, as the paternal X chromosome cannot inactivate and thus dosage compensation cannot be achieved in the extra-embryonic tissues (87). Similarly for differentiated ES cells and embryos, the X chromosome carrying the null *Xist* allele is always the Xa, demonstrating non-random XCI *in vitro* and *in vivo* (86). However, *XIST/Xist* is not required to keep genes silenced after XCI has been established (88).

1.5.2 Modifiers of randomness in XCI

There are also less extreme cases where one X of the two X chromosomes in females are preferentially, but not exclusively, inactivated. Having the same X chromosome inactivated in $\geq 75\%$ of cells in a female is called skewing of XCI (reviewed in 89). In mice, there is an X-controlling element (*Xce*) locus that modifies the randomness of XCI and different strains appear to have different *Xce* alleles (90). The relative strength of the three defined alleles is

$Xce^a < Xce^b < Xce^c$, where the weaker allele is more likely to be inactivated in a heterozygous female. The degree of skewing can reach ~25:75 for a Xce^a/Xce^c heterozygous female, where the X chromosome carrying the Xce^c allele is the Xa in ~75% of the cells (90). Through correlating the XCI pattern with strain genotypes on the X chromosome in female mice with recombinant X chromosomes, Chadwick *et al.* (91) narrowed the candidate *Xce* region to 1.85 Mb, which encompasses the *Xist* locus.

It remains unclear whether humans also have an *Xce* that modifies randomness of XCI. Females with mutations at the *XIST* promoter have been shown to exhibit skewing of XCI (92-94), but skewing of XCI does not appear to be heritable from mother to neonates (95). In a study examining the pattern of XCI in more than 1000 phenotypically normal females, 25% of females showed skewing to the degree of >70:30 (inactivating the same X chromosome in at least 70% of the blood cells), but only a small proportion of females showed skewing degree of >80:20 (96). In addition, increased skewing of XCI has been associated with ageing (95-98), suggesting that skewing of XCI may be largely an acquired trait in the human population.

1.6 Escape from XCI

Despite the chromosome-wide silencing of XCI, not all genes on the X chromosome are inactivated. Based on allele-specific expression in females with non-random XCI, and expression level of X-linked genes in mouse-human somatic cell hybrids containing a human Xa or Xi, Carrel and Willard (99) have shown that ~15% of X-linked genes escape XCI in humans, with an additional 10% of X-linked genes showing variable XCI status between individuals or somatic cell hybrids. In mice, XCI is also incomplete, with ~8.9% (35/393) of transcripts found to escape from XCI through different methods of detection (see Table 1.1 for a list of mouse escape genes and their XCI status in different species). The XCI statuses of genes are only partially conserved, as only about seven of the 35 mouse escape genes including *XIST/Xist* have been shown to escape from XCI in humans as well. In addition, XCI statuses of genes do not always agree between studies (Table 1.1), which may reflect variable XCI pattern of the genes between individuals or cell types or different definitions of escape from XCI based on the method of detection used. Although genes that are expressed from the Xi with >10% expression of the Xa are generally considered to escape from XCI, escape genes do show a wide range of expression level on the Xi, and the expression level from the Xi can differ between individuals as well (99). In fact, most escape genes are not fully expressed from the Xi compared to the Xa in both humans and mice (99-101). The variable XCI status of genes and degree of escape may be allele-specific whereby a certain genotype allows for escape, since two genes *Ddx3x* and *Smc1a* have been shown to escape from XCI in voles when present on the Xi with one genetic

background but not another (102). Furthermore, different human populations also appear to have differing expression levels of some escape genes (eg. *DDX3X* and *STS*) from the Xi (100). Therefore, not only can XCI status of genes differ between females, the degree of escape in terms of expression levels is also variable between individuals.

The majority of genes escaping from XCI in humans are clustered on the short arm of the X chromosome (99, 100), while the escape genes in mice are randomly distributed on the X chromosome and therefore generally have neighbouring genes subject to XCI (101). In mice, there have been four non-coding RNAs found to escape from XCI in close proximity to a known escape gene (103), and whether these non-coding RNAs have a functional role in the escape of the neighbouring gene requires further investigation. In general, the distribution of escape genes on the chromosome suggests that escape is controlled at the domain level in humans as opposed to the gene-by-gene basis proposed in mice, and perhaps genes escape from XCI via different mechanisms in humans and mice although a few escape genes are conserved across species.

1.6.1 Epigenetic signatures of escape genes

Genes that escape from XCI show epigenetic features more reminiscent of the active state, and thus have patterns of DNA methylation and chromatin marks more similar to the genes on the Xa. Large-scale studies of DNA methylation on the X chromosome have shown a strong correlation between promoter DNA hypomethylation on the Xi and escape from XCI (54, 56, 63). Studies that compare DNA methylation levels between sexes and/or between mouse-human somatic cells carrying an Xa or Xi have also demonstrated low promoter DNA methylation at escape genes on both Xa and Xi (104, 105). Also, similar to actively expressed genes, escape genes show DNA hypermethylation at the gene body (63). Escape genes on the Xi also tend to possess histone modifications that are associated with active genes on the Xa. Most of the escape genes analyzed have shown hyperacetylation of histones and increased H3 lysine 4 methylation, and reduced inactive marks such as H3 lysine 9 trimethylation and H3 lysine 27 trimethylation at the promoter region (68, 101, 106). MacroH2A, a histone variant enriched on the Xi, is also depleted at several escape genes in mice (107).

Interestingly, mouse Xist RNA has been found to coat the promoter and transcribed region of genes subject to XCI but not genes that escape from XCI such as *Kdm6a* and *Kdm5c* (108). This agrees with the relocation of silenced genes into the Xist RNA compartment during XCI, while the escape gene *Kdm5c* (previously known as *Jarid1c*) remains at the periphery of the Xist RNA domain (78). However, in humans, regardless of XCI status, all analyzed genes are located at the periphery of the XIST RNA compartment in the nucleus (77).

1.7 Cis-regulatory elements

The mechanism(s) by which genes escape from the chromosome-wide silencing on the X chromosome is an interesting question to address in the field of XCI. It has been hypothesized that the genomic environment surrounding the genes can influence whether they would be subject to or escape from XCI. In 1983, Gartler and Riggs (109) initially proposed the existence of DNA sequences that promote the silencing of genes, termed waystations or booster elements, and in recent years, the idea of another type of DNA element that can promote escape from XCI, called escape elements, has also been introduced. The studies that aim to identify candidates for waystations to date are all done bioinformatically, examining content of various DNA sequences surrounding genes of different XCI status (discussed in section 1.7.2), while the existence of escape elements is supported by experimental results.

1.7.1 Escape elements

The evidence for the existence of escape elements is presented in a study by Li and Carrel (110), where they examined the inactivation status of mouse BAC (bacterial artificial chromosome) derived transgenes at different loci on the X chromosome in female ES cells. The BACs contain an escape gene known to escape from XCI in both humans and mice, *KDM5C/Kdm5c* (111), and its neighbouring genes, *Tspyl2* and *Iqsec2*. Despite substantial sequence conservation in the 235 kb (kilo base pairs) region around *KDM5C/Kdm5c* between humans and mice, humans possess a larger domain of escape genes (including *IQSEC2*) while *Kdm5c* is the only escape gene in the region for mice (111). Single-copy transgenes were examined at four different locations on the mouse X chromosome, and both *Tspyl2* and *Iqsec2* on the BACs remained properly inactivated as seen for the endogenous alleles. In contrast, the transgenic *Kdm5c* was shown to escape from XCI at all four integration sites by RNA/DNA FISH. Using allele-specific expression analysis, Li and Carrel also demonstrated for two transgenic lines that the transgenic *Kdm5c* is expressed from the Xi at ~40% expression of the Xa. Previously reported expression level from the Xi of endogenous *Kdm5c* generally ranges from ~20-70% of the Xa depending on the cell types (112, 113). The developmental stage of the cells is also an important factor in the degree of escape, as both the level of *Kdm5c* expression from the Xi and the proportion of cells in which *Kdm5c* escapes increase with time since following XCI in early development (114).

Although the local environment of the transgenic *Kdm5c* analyzed by Li and Carrel (110) is the same at all integration sites, the escape of *Kdm5c* indicates that it is protected from the long-distance spread of silencing on the X chromosome, and that the 112 kb region of overlap between the BACs contains *cis*-regulatory element(s) promoting the escape of *Kdm5c*. The

insulator protein CTCF is found to bind the 5' end of *Kdm5c* *in vitro* and *in vivo* (115), and CTCF sites are also present on the BACs analyzed by Li and Carrel (110). In contrast, CTCF binding is absent at the 5' end of the human *KDM5C*, suggesting the loss of the insulator activity contributes to the larger domain of escape in humans (115). Although CTCF has been implicated in insulating *Kdm5c* from the inactivated neighbourhood, boundary elements containing CTCF binding sites have been shown to be insufficient for an X-linked transgene to escape from XCI (116). Thus, further investigation is required to decipher the role of boundary elements and insulators in the mechanism of escape from XCI.

1.7.2 Waystations

X;autosome translocations in both humans (117-120) and mice (121) have demonstrated incomplete or variable inactivation of the autosomal segments linked to the X chromosome. Most of the X;autosome studies examined inactivation of the translocated autosomes through cytogenetic observations associated with the Xi, such as histone deacetylation, late replication timing, and XIST/Xist RNA painting. However, a few studies have assayed the expression of selected genes on the translocated autosomes and showed the spreading of silencing into the translocated autosomes can be continuous or discontinuous from the breakpoint (119) and approximately 30% of autosomal genes escaped from XCI (120; reviewed in 122). The attenuated silencing on the autosomes, as shown for the inducible *Xist* transgene system as well, led to the hypothesis that there are genomic features called waystations that are repetitive along the X chromosome and propagate the inactivating signals of XCI (109). The waystations should be more abundant but not exclusive to the X chromosome, since a proportion of autosomal genes can be silenced. Lyon proposed that LINE-1 (long interspersed nuclear element-1) is a candidate for waystations (123), as supported by the two-fold enrichment on the mammalian X chromosomes compared to the autosomes (~29% on the X chromosome in mice and humans; 124), especially in the long arm of the X where inactivated genes are enriched (1). In particular, the over-representation of LINE-1 on the X chromosome is due to the younger LINE-1, which are more likely to be full-length and thus retain retrotransposition capability than the older LINE-1 (1, 125). Intriguingly, two families of full-length, young LINE-1 in mice have been shown to be expressed from the Xi during XCI in an *Xist*-dependent manner (126), further supporting the role of LINE-1 in XCI. Other candidate waystations include other repetitive elements, oligmers, and sequence motifs (125, 127-130).

In order to identify candidates for waystations important in the spread of inactivation, several bioinformatic studies have compared the content of various repetitive elements surrounding genes of different XCI status (127, 128), with the expectation that waystations would be

enriched around genes that are subject to XCI but depleted around genes which escape from XCI. The caveat of identifying waystations through this approach is that the apparent enrichment or depletion of certain repeats may be a reflection of the evolutionary history in different segments (strata) of the X chromosome, since genes subject to and escaping XCI are primarily located on the older (long arm) and the younger strata (short arm) of the X chromosome, respectively (1, 99). Therefore, Wang *et al.* (127) and Carrel *et al.* (128) have confirmed the enrichment of candidate waystations around genes subject to XCI within the younger strata where there are a mix of genes subject to and escaping XCI. Among the many genomic features proposed to be candidate waystations, both studies support the role of LINE-1 in XCI (127, 128). In contrast, Alu elements are enriched around genes escaping from XCI (127); similarly, [GATA]_n is over-represented in the region of the X chromosome that escapes from XCI compared to rest of the X chromosomes and the autosomes (129). Although most bioinformatics studies have focused on genomic features that can distinguish between genes of different XCI status at different regions of the X chromosome, based on the differing distribution of escape genes across the chromosome, it is plausible that the genes on the short arm and the long arm utilize different mechanism to escape from XCI.

1.8 XCI status of transgenes

Numerous transgenes have been inserted on the X chromosome (eg. 131, 132), some of which were targeted to the mouse *Hprt* locus (eg. 133, 134) because the *Hprt* locus has been identified as a permissive docking site where transgenes exhibit appropriate and consistent expression patterns when on the Xa (135, 136). X-linked transgenes vary in copy number, integration site, and also construct size, but are predominantly subject to XCI. Transgenes subject to XCI include DNA of mammalian (135-138), viral (139), and avian origins (140). Even highly expressed promoters such as chicken and human *beta*-actin (135, 138) are subject to the regulation of XCI when present on the X chromosome. The XCI status of transgenes is generally assessed by comparing the overall expression and/or DNA methylation level of the transgene between homozygous females, heterozygous females, and hemizygous males, with the expectation that a gene subject to XCI would show lower expression in heterozygous females and DNA hypermethylation in females compared to males. In addition, a gene escaping from XCI would exhibit increased expression in homozygous females compared to males. Xi-specific expression or DNA methylation can also be tested in transgenic mice with an X;autosome translocation resulting in non-random XCI, or in cells selected to carry the transgene on the Xi (131, 132).

Two transgenes of non-X origin have been shown to escape from XCI in somatic cells: chicken transferrin and human collagen. The nuclear factor- κ B GFP (green fluorescent protein) is another transgene potentially escaping from XCI as there are similar numbers of GFP-positive cells between males and females (141), but further experiments are required to confirm the XCI status. The chicken transferrin transgene is randomly integrated at band D on the mouse X chromosome, with 11 tandem copies amounting to ~187 kb (142). The chicken transferrin transgene is expressed from the Xi in females with non-random XCI at a similar level as the Xa in males (142) and shows equivalent DNA methylation level on the Xi and Xa (131), demonstrating that the chicken transgene fully escapes from XCI. In contrast, the 39.6 kb human collagen transgene located close to the C and D bands only partially escapes from XCI. By comparing expression levels between homozygous females, heterozygous females and hemizygous males, Wu *et al.* (143) have shown that human collagen is expressed on the Xi at ~20-25% of the Xa, with the assumption that Xa is equivalently expressed in males and females. Further analysis of Xi-specific transgene expression in females with non-random XCI supports the partial escape from XCI. Surprisingly, only ~3% of cells show expression of human collagen from the Xi by RNA *in situ* hybridization, thus the transgene in the majority of the cells is subject to XCI. In agreement, DNA methylation of the transgene is higher in females with non-random XCI (when the transgene is on the Xi) compared to heterozygous females with random XCI, who show higher DNA methylation than males. Therefore, this study highlights the importance of different methods of detection, which address the same question through different perspectives and can give different results regarding the XCI status of genes.

1.9 Thesis objective

The objective of my thesis is was identify escape elements in the human genome by examining the XCI status of various human transgenes targeted to the *Hprt* locus on the mouse X chromosome as part of the Pleiades Promoter Project. Promoter DNA methylation was used to assess the XCI status of the human transgenes. DNA methylation of gene body and intergenic CpG islands were also examined.

Table 1.1 Mouse escape genes common across species in somatic cells

Gene	Mouse	Vole	Human	Cow	Elephant
1810030O07Rik/ CXorf38	E (101) ²		E (99) ³		
2010308F09Rik	E (103) ¹				
2610029G23Rik/ CXorf26	E (103) ¹ (144) ¹ (101) ² (107) ²		S (99) ³		
5530601H04Rik	E (103) ¹ (107) ²				
6720401G13Rik	E (101) ²				
Akap14	E (145) ²		S (99) ³		
Atrx	E : (146) ⁵ S : (101) ²	S (102) ²	S (146) ⁵ (99) ³		S (146) ⁵
B230206F22Rik/Ftx	E (147) ^{2,5}				
BC022960	E (101) ²				
Bgn	E (101) ²		S : (99) ³ E or S : (56) ⁴		
Car5b	E (101) ²		E (99) ³ (100) ¹		
Cdk16/Pctk1	E : (145) ² S : (101) ²		E (99) ³ (100) ¹		
Chm	E : (145) ² S : (101) ²	S (102) ²	E or S (99) ³ (56) ⁴		
D330035K16Rik	E (103) ¹				
D930009K15Rik	E (103) ¹				
Ddx3x	E (103) ¹ (144) ¹ (101) ²	E or S (102) ²	E (99) ³ (100) ¹		

Gene	Mouse	Vole	Human	Cow	Elephant
<i>Eif2s3x</i>	E (103) ¹ (144) ¹ (101) ² (148) ² (145) ²		E (148) ¹ (99) ³ (100) ¹		
<i>Enox/Jpx</i>	E (149) ^{2,4} (150) ^{2,5}				
<i>Fmr1</i>	E : (145) ² S : (101) ² (151) ⁴		S (151) ⁴	S (151) ⁴	E (146) ⁵
<i>Hcfc1</i>	E (145) ²		E or S (99) ³		
<i>Huwe1</i>	E : (146) ⁵ S : (101) ²		S (146) ⁵ (99) ³		S (146) ⁵
<i>Kdm5c</i>	E : (103) ¹ (144) ¹ (101) ² (145) ² (110) ^{2,5} S : (146) ⁵ E or S : (151) ⁴		E (146) ⁵ (151) ⁴ (99) ³ (100) ¹	S : (152) ¹ E or S : (151) ⁴ E : (153) ¹	E (146) ⁵
<i>Kdm6a/Utx</i>	E (144) ¹ (101) ² (154) ² (151) ⁴	E (102) ^{2,4}	E (154) ³ (151) ⁴ (99) ³ (100) ¹	E (151) ⁴	
<i>Mid1</i>	E : (101) ² S : (110) ²	S (102) ²	S (99) ³		
<i>Nkap</i>	E : (145) ² S : (101) ²		S (99) ³		
<i>Pnck</i>	E (144) ¹		S (56) ⁴		
<i>Pola1</i>	E : (146) ⁵ S : (101) ²		NE: (146) ⁵ S : (99) ³		
<i>Rbmx</i>	E : (146) ⁵ S : (101) ²		S (146) ⁵ (99) ³		NE (146) ⁵
<i>Shroom4</i>	E (101) ²		S (99) ³		
<i>Suv39h1</i>	E : (145) ² S : (101) ²		S (99) ³		
<i>Sts</i>	E : (155) ¹ S : (144) ¹ (156) ⁶	S (157) ⁶	E (99) ³ (146) ⁵		
<i>Taf1 and/or Ogt</i>	E (145) ⁵		Both S (99) ³		

Gene	Mouse	Vole	Human	Cow	Elephant
<i>Usp9x</i>	E: (145) ² S: (101) ² (146) ⁵		E (146) ⁵ (99) ³ (100) ¹		E (146) ⁵
<i>Vbp1</i>	E: (145) ² S: (101) ²		S (99) ³		
<i>Xist</i>	E (7) ² (103) ¹ (144) ¹ (101) ²	E (102) ^{2,4-5}	E (5) ^{1,3} (6) ⁵	E (152) ¹	

Method of detection: Superscript 1=expression differences between cells with different number of X chromosomes (eg. male/female/aneuploids); 2=allele-specific expression; 3=mouse-human somatic cell hybrids; 4=DNA methylation; 5=RNA FISH; 6= enzyme activity difference between animals with different number of X chromosomes.

E, escape from XCI (green). S, subject to XCI (red). E or S, variable XCI status (blue). NE, not expressed. Genes escaping in both humans and mice are bolded.

List of genes escaping in more than one species but not in mice: *UBA1* (100, 146, 151, 158); *ZFX* (100, 146, 151, 152); *SMC1A/sb1.8* (100, 102); *MED14/CRSP2* (151); *RPS4X* (153)

2 Methods

2.1 Pleiades Promoter Project constructs

The Pleiades Promoter Project (159, 160) is a collaborative effort to develop various human promoters driving specific expression patterns in the mouse brain, eye, and spinal cord. Most of the promoters originated from human autosomal genes, with only two X-linked promoters, *DCX* and *MAOA* being assessed. All Pleiades constructs were integrated by homologous recombination at the *Hprt* locus on the mouse X chromosome. MiniPromoters (MiniP) were 4 kb and less and were composed of different combinations of small putative regulatory elements. In contrast, MaxiPromoters (MaxiP) were human BAC-derived constructs that ranged from 100 to 200 kb, with the reporter (*lacZ*, enhanced GFP [EGFP], or EGFP/cre) inserted at the start codon of the gene of interest. Since MaxiPs were intact BACs rather than a combination of selected elements joined together, there was only one MaxiP for each target human gene used to design the promoter. The MaxiPs were named in the format Gene-X, with the “X” indicating that it was a MaxiP. However, for MiniPs, there were several combinations of the regulatory elements used to generate a construct, thus resulting in multiple constructs for one gene. These constructs were named Gene-A, B, etc., with the letter at the end denoting the different variation of the promoter. Every MiniP and MaxiP construct can be identified with a unique Pleiades number (Ple#).

2.2 Generation of mice with *Xist* deletion and MaxiP

The Simpson laboratory crossed female mice carrying an *Xist* gene flanked by lox sites to 129-ACTBCre males to generate the *Xist* deletion. Female mice with the *Xist* deletion (129-*XXist*^{1lox}/X) were then crossed with males carrying the Pleiades construct integrated at the *Hprt* locus (B6-*X*^{MaxiP}/Y) to generate 129-*XXist*^{1lox}/B6-*X*^{MaxiP} females. This knock-out of *Xist* has been shown to render the X chromosome unable to inactivate (161), thus resulting in the knock-in X chromosome with intact *Xist* becoming the Xi.

2.3 DNA and RNA tissue extraction and expression analysis

DNA and RNA extractions from 50-100 mg of mouse livers and/or brains were done using TRIzol Reagent (Invitrogen), according to the manufacturer’s protocol. DNA and RNA were extracted from mice carrying the NR2E1-F (Ple140), NR2E1-X (Ple142), NGFR-X (Ple133) with and without the *Xist* deletion, and MKI67-D (Ple131) constructs.

Staining of *lacZ* with X-gal and EGFP with anti-GFP in the brains and ear notches was performed by the Simpson laboratory to determine the expression status of the Pleiades constructs. For analyzing transcription level, approximately 2 µg of RNA extracted from livers was converted to complementary deoxyribonucleic acid (cDNA) with standard reverse transcription conditions using M-MLV (Invitrogen) at 42° for 2 h followed by a 5 min incubation at

95°. Quantitative PCR (qPCR) was used to determine relative transcription levels of *PHB*, *HPRT*, and the intergenic region between *PHB* and *HPRT* compared to *Pgk1* in mice carrying the Ple133 construct (*NGFR* BAC), using a StepOnePlus™ Real-Time PCR System (Applied Biosystems, Darmstadt, Germany), using Maxima® Hot Start Taq (Fermentas) and EvaGreen® dye (Biotium). Conditions for qPCR were as follows: 95° for 5 min, followed by 40 cycles of [95° for 15 s, 60° for 30 s, 72° for 1 min], and a melt curve stage of [95° for 15 s, 60° for 1 min, increase of 0.3° until 95°]. Serial dilutions of genomic DNA from the *NGFR* female tEMS 9703 (without the *Xist* deletion) were used as the standards to which each sample cDNA was compared, to generate a relative quantity of *PHB*, *HPRT*, *Pgk1*, and intergenic transcription between *PHB* and *HPRT*. Expression levels were normalized to *Pgk1* expression level, and quantifications were done in triplicates, of which any outlier identified by the StepOne Software v2.2.2 is excluded in the analysis. Primer sequences are found in Table 2.1.

2.4 DNA methylation analysis

DNA methylation was analyzed predominantly in DNA obtained from ear notch samples, but *NR2E1* promoter DNA methylation was also examined in liver and brains samples. Lysed ear notches of about 2 mm in diameter from transgenic mice of approximately four weeks old were obtained from the Simpson laboratory. Using the EZ DNA Methylation-Gold Kit or the EZ-96 DNA Methylation-Gold Kit (Zymo Research Corporation), 500 ng of DNA obtained from the lysed ear notches or liver and brain samples were used for bisulfite conversion, following the manufacturer's instructions. Internal bisulfite conversion controls are included in the pyrosequencing assays to monitor complete conversion of DNA. Each 25 µL pyrosequencing PCR was performed with 1x PCR Buffer (Qiagen), 0.2 mM dNTPs (deoxynucleotides), 0.625 U Hot Start *Taq* DNA polymerase (Qiagen), 0.25 µM forward primer, 0.25 µM reverse primer, and 12-35 ng bisulfite-converted DNA. Assays for *CCKBR*, *ICMT*, *NOV*, and *NR2E1* were performed with 0.5 µM forward and reverse primers. Conditions for PCR were: 95° for 15 min, 50 cycles of 94° for 30 s, annealing temperature for 30 s (See Table 2.2), 72° for 1 min, and finally 72° for 10 min. One of the forward or reverse primer was biotinylated, depending on which strand contains the target region to be sequenced, in order to subsequently isolate the strand of interest for pyrosequencing. Template preparation for pyrosequencing was done according to the manufacturer's protocol, using 10 to 15 µL of PCR products. CDT tips were used to dispense the nucleotides for pyrosequencing using the PyroMark MD machine (Qiagen).

Variability in pyrosequencing results within a sample was observed for some DNAs, which we attributed to degradation of ear notch DNAs stored for up to 3 years. All promoter assays were replicated in ear notch samples (but not all tissue samples) at least twice and the average is

presented. If the standard deviation of a sample for a particular assay was large enough to be considered an outlier using the modified Z-score method (see below), the data point was not included in the analyses. *HPRT*, *Phf6* and *lacZ* assays were replicated on sufficient samples that we were confident of their reliability (average standard deviation of 5%, 3%, 5%, respectively), and therefore for these three assays not all samples were replicated. Each human promoter assay was tested in at least one mouse sample without the target transgene to ensure the specificity of the human primers.

We used the UCSC (University of California – Santa Cruz) definition of a CpG island (GC content of at least 50%, length >200 bp, Observed_{CpG}/Expected_{CpG} ratio >0.6), which was first proposed by Gardiner-Garden and Frommer (32). Primers were designed for 40 CpG islands using the PSQ Assay Design software (Qiagen). At least three CpG's were analyzed for each CpG island examined, and the distance of the analyzed CpG's from the transcription start sites can be found in Table S1.

Promoter identification of *PITX2* (in Ple158) was based on the ENCODE (Encyclopedia of DNA Elements) chromatin states track downloaded from UCSC genome browser (162).

2.5 Repetitive element content analysis

The repeat content for LINE-1 and Alu on the MaxiPs was expressed as the percent base pair composition of the repetitive element of interest on the BACs. The coordinates of LINE-1 and Alu in the genome was obtained from Repeat Masker (<http://www.repeatmasker.org/>, Institute for Systems Biology) and calculating the overlap between the repetitive elements and the BACs was done using Galaxy (163-165). Detection of full-length LINE-1 on the MaxiPs was performed using L1Xplorer (166).

2.6 Statistical analysis

Statistical analyses were done using GraphPad Prism 5.02. An alpha value of 0.05 was used for testing significance. Mann-Whitney t-test and one-way ANOVA Kruskal-Wallis test were used to test for significant differences in DNA methylation level between male and female mice. Mouse strains with modified Z-scores higher than 3.5 in absolute value were considered outliers (167). Spearman correlation was performed to examine the relationship between DNA methylation levels of neighbouring genes.

Table 2.1 qPCR primers used to determine relative transcription level of transgenes on the Xi normalized to *Pgk1*

Assay	Construct	Sequence (5'-3')	Location of assay
qhHPRT	All	F2: CCTGCTTCTCCTCAGCTTCAG R2: CGGGAAAGCCGAGAGGTT	Exon 1 7 bp 3' of TSS
qHPRT-5'A	All	F1: CAAATCTCCTGCCATCACATACC R1: AGTGCCAGCACATAGTTGGT	Intergenic 1121 bp 5' of <i>HPRT</i> TSS
qHPRT-5'B	All	F1: GCCACAGGTAGTGCAAGGTCTT R1:CCAGTCATCGCGTGAATCCT	Intergenic 258 bp 5' of <i>HPRT</i> TSS
qPgk1-e1	Endogenous	F1: CGTCTGCCGCGCTGTT R1:AACACCGTGAGGTCGAAAGG	Exon 1 64 bp 3' of TSS
qPHB-3'UTR	NGFR-X: Ple 133 (RP11-158L10)	F1: CTGTCACTGATGGAAGGTTTGC R1: AGGCCTGCCTTCTCAGTTCA	3'UTR

3'UTR, three prime untranslated region

Table 2.2 Pyrosequencing primers and cycling conditions for DNA methylation analysis

Assay	Construct	Distance of closest CpG from TSS (bp)	Sequence (5'-3')	Annealing temperature (°C)	Position of CpGs analyzed
AMOTL1	Ple 5 (RP11-936P10)	72 (5')	F1: GGGATAAAGGAAGGGATGTTG R1: *TCACTAAAACCTACACTCCACC S2: GGAGGGTGTGTTGTAGA	55	8-13
ATP6V1C2	Ple 7	544 (5')	F1: AGGTGGGAGTTTTTTGGGTAAT R1: *CAAAAAAATCACCTACTCCCAAATATCT S1: GGGAGTTTTTTGGGTAA	53.9	1-5
CARTPT	Ple 20-21	296 (5')	F1: GTAAATGTGGTTGTTTGGAGGTAATA R1: *TCCCAACACCTAACAATAATAACAAC S1: TGGTTGTTTGGAGGTAAT	55	1-3
CCKBR	Ple 24-25	424 (5')	F1: GAGGAGTTGTAGGGAATTA R1: *AATACTTTAATCTAAACCTAAAACC S1: GAGGAGTTGTAGGGAAT	55	1-5
CLDN5	Ple 34	820 (3')	F1: AGTTGTTAGAGGTTTTGTGATTG R1: *AAAAATACCCTCTTTAAAAATTC S1: GTTGTTAGAGGTTTTGTGA	53.3	1-5
DRD1	Ple 61-62	38 (3')	F1: TATTGTTATAGGTTTTTGAGAGGT R1: *CCTTCAACCCTACAAAACAAA S1: ATTGTTATAGGTTTTTGAGA	53.3	1-5
FEV	Ple 66-67	1 (3')	F1: *GGAGGGGGAGGAGAGTGTA R1: CCCTCCCTAAAACCCTTCTTC S1: AAAACCCTTCTTCCAA	53.9	1-4
GPX3	Ple 97-98	39 (3')	F1: TGGGGAGTTGAGGGTAAGT R1: *CCCAACCACCTTTCAAAC S2: GGGAGTTGAGGGTAAGT	55	1-5
GRP	Ple 99-100	216 (5')	F1: AGAGGGAGGAGTTTATTAAATTGTGTT R1: *CATTACCCCTCTTTTTTCT S1: AAATTGTGTTGGATGGA	55	1-5
HAP1	Ple 103-104, 106	262 (5')	F1: GGAGGGGTTGTTTTAGTTAGGG R1: *ATTTTTCTACCCTCTCCATCTCC S1: GTTGTTTTTAGTTAGGGATT	53.9	1-4

Assay	Construct	Distance of closest CpG from TSS (bp)	Sequence (5'-3')	Annealing temperature (°C)	Position of CpGs analyzed
HBEGFc	Ple 107-109	209 (5')	F1: GTTTGGGGAAAGGTAGGAAT R1: *TCACAATTTTTAAAACCAAACC S1: GTTTGGGGAAAGGTA	55	1-5
HPRTb	All constructs	94 (5')	F1: GGAATTAGGGAGTTTTTTGAATAGG R1: *CCTACCAATTTACAACTCACTAAATA S1: GGGAGGGAAAGGGGT	55	1-3
HTR1A	Ple 119	266 (5')	F1: *TTTGGGATTGGAGATTGTTTGT R1: ACTCCAATAAAAACTAAAATTAACCT S1: CTAAAAAACTAAAATTAACC	55	5-8
ICMT	Ple 123-124	206 (5')	F1: GGAATTTTTTGAGTTTGGGATTAA R1: *CATCCCAACTCTAAACCAAACCTCTATA S1: TGGGATTAAGTTTGGATA	58.3	1-4
LacZ-meth	Reporter	~724 (3' of <i>lacZ</i> seq)	F1: TGTATTGGAGGTTGAAGTTTAGATGT R1: *TTTCACCCTACCATAAAAAAACTATTAC S1: TGGAGGTTGAAGTTTAGAT	55	1-4
LacZ-3'-meth	Reporter	~2577 (3' of <i>lacZ</i> seq)	F2: TGGTAGTATTAGGGGAAAATTTTATTTA R2: *CCAATAACAATTCAAACCAATC S2: GATTGATGGTAGTGGTAAA	55	1-5
MCM6	Ple 126 (RP11-406M16)	190 (5')	F1: *GTGGAATGATTTAAAGAATATTTGAAAA R1: CCTTCTAAAAAAAACCCATCTACCTT S1: CTTCTAAAAAAAACCCATC	55	3-7
mPhf6	Endogenous	305 (5')	F1: GTAAGGGTTAAGGTTTGTGTATTTGT R1: *CCAAAAACCTAAAACCAAATCCT S1: GTTAAGGTTTGTGTATTTGTT	55	1-3
NGFR	Ple 133 (RP11-158L10)	279 (5')	F1: AGGAAGATGGGTAAGAGAGTGAATT *TCCCTACCTTATCCCTTAAACCT GGTAAGAGAGTGAATTTTGT	55	1-4
NOV	Ple 134 (RP11-840I14)	26 (5')	F33: GTTTTTTATTTTTTGGGAAAAGTT R33: *ACAATTAATAAATACTACTCTCCTTAAA S1: TTTTTTGGGAAAAGTTAG	48	1-6

Assay	Construct	Distance of closest CpG from TSS (bp)	Sequence (5'-3')	Annealing temperature (°C)	Position of CpGs analyzed
NR2E1	Ple 140, 142 (RP11-144P8)	22 (3')	F1: *TTAGGAGTTGGGGGAAAAGTTAA R1: AACTAAATCCCCTATAATATCTCCAAAA S1: ATCCCCTATAATATCTCCA	55	2-4
NTSR1	Ple 144, 146-147	368 (5')	F1: GTTGGGGGAGGTGTATAGTT R1: *TACCACCCTCTTCCCTATT S1: TTGGGGGAGGTGTAT	58.3	1-3
OLIG1	Ple 148, 150-151	36 (5')	F1: GAGGGAGGTTGTTTTGAGTAGA R1: *CCCTACCCCTTTAAACCC S1: GGTATAAGTAGTTAATGAATA	55	1-11
OXT	Ple 152-153	107 (5')	F1: GTTTTGTTAATGAAGAGGAAAAGTT R1: *ACCTAACCTTTTTATACCTAAACAT S1: TTGTTAATGAAGAGGAAAAGT	55	1-3
PHB-IC	Ple 133 (RP11-158L10)	1 (5')	F1: GAATTAGGGTGAGGTTTTAAGTTATTTT R1: *ACATAAATTCCCCAACCAACACA S1: GGGTGAGGTTTTAAGTTAT	58.3	1-5
PITX2-CpG18	Ple 158 (RP11-268I1)	Gene body 1385 (5') of Alt. TSS	F1: GGGATTGGGGTTAATTAGTTTTTGG R1: *AACTCCCTCCCCTTTCAAATTTT S1: AGGGATTGGGGTTAA	58.3	1-4
PITX2-CpG22	Ple 158 (RP11-268I1)	Gene body	F1: *AAATTTGTAGTTTATTTGAAAGGTGTTT R1: ACAACTAATACAATTTCCCCTAAAAATA S1: AACTAATACAATTTCCCCTA	58.3	1-3
PITX2-CpG29	Ple 158 (RP11-268I1)	Intergenic	F1: *GTTTTGATTTGGAGGAGGTATTAGT R1: AACCTAACCCACCAATACTCC S1: AACCTAACCCACCA	58.3	1-5
PITX2-CpG46	Ple 158 (RP11-268I1)	Gene body	F92: GTATTTTTTTAGGTTTGTGTTGTGGTAGAG R92: *CCCCAACCAACCAAATCTTTTT S1: TGGTAGAGAAGGGGGA	48	1-5
PITX2-CpG59	Ple 158 (RP11-268I1)	Intergenic	F1: TGATTAGGATTTTTTGGATTTATGAATT R1: *CCATATCATTAACCAAAAACATAACATT S1: GGATTTTTTGGATTTATGA	55	1-7

Assay	Construct	Distance of closest CpG from TSS (bp)	Sequence (5'-3')	Annealing temperature (°C)	Position of CpGs analyzed
PITX2-CpG100b	Ple 158 (RP11-268I1)	Gene body	F1: TGGAGTGGAAAAGTGGTTTAATA R1: *AAACCTAAATAACTAAATAAACCCCTAAT S1: GTGGAAAAGTGGTTTAATA	56.3	1-6
PITX2-CpG196b	Ple 158 (RP11-268I1)	18 (3') of Alt. TSS	F1: TGGTTTTAAGATGTTAGGTTAATAGGG R1: *ACTCAACTCCAAACACCCAAA S1: GATGTTAGGTTAATAGGGAA	55	1-6
PITX2-exon1	Ple 158 (RP11-268I1)	146 (3')	F1: *AAAGGTTAGAGGGATTAATATATAGGT R1: ACTTCCCTTCTACAACAATTTTCT S1: ACTTCCCTTCTACAACAAT	58.5	1-3
PITX2-intron2	Ple 158 (RP11-268I1)	Gene body 1580 (3') of Alt. TSS	F1: AGATATTAATAATTTATAGGGTGTTGAA R1: *AAACTTTATACCCAACCCTTTATCT S1: TAATTTATAGGGTGTTGAAG	53.3	1-3
PITX3	Ple 160-162	78 (5')	F1: GAGTTTTAGTAGGGTAGTTGGAAAGG R20: *CCATTCACCTTTATAACAAACCAAAA S1: GTAGGGTAGTTGGAAAGG	55	1-10
POGZ	Ple 167-168, 170	84 (5')	F1: GTAGGGGTTTGGATGAGTTTATGA R1: *CTTTTTACCCACCTCCCAATTA S1: GGGTTTGGATGAGTTTA	55	1-4
RLBP1L2b	Ple 179-181	396 (5')	F1: TGGGGAGGTTGGAAAGTATG R1: *CCCCACTCCTCAACAACTACT S1: GGGGAGGTTGGAAAG	58.3	1-5
SLC6A4	Ple 198	134 (5')	F1: *TGTTAGGTTTTAGGAAGAAAGAGAGA R68: CATCCTAACTTTCTACTCTTTAACTTTA S1: AACTACACAAAAAAACAAAT	58.3	6-10
TAC1	Ple 214, 217	2 (3')	F1: GAATTTAATTGGGTTTAGATGTTATGGG R1: *TTTAATTAACCCCTCCTCTCCTTT S1: GGGTTTAGATGTTATGGGTA	55	1-6
THY1	Ple 229	24 (5')	F1: GGAGGTGGGTTTTAGTTGAAA R1: *AAAAACATTATCCTCCTCCCTAAA S1: TGAAAAGGAAATGTGGA	58.3	1-3

Assay	Construct	Distance of closest CpG from TSS (bp)	Sequence (5'-3')	Annealing temperature (°C)	Position of CpGs analyzed
UGT8	Ple 241-242	88 (5')	F1: GTGGGTGGTGGTAGAAAG R1: *CCCACTCTTCCCTCTTTA S1: TGGGTGGTGGTAGAA	58.3	1-4

Each pyrosequencing assay consists of a forward primer, reverse primer and a sequencing primer. One of the forward and reverse primers was biotinylated at the 5' end (indicated by an asterisk *). Position of analyzed CpG is relative to the sequencing primer. Alt. TSS, alternative transcription start site.

3 Results

3.1 DNA hypermethylation reflects XCI of Pleiades constructs

The Pleiades Promoter Project has generated reporter constructs with human promoters and targeted the transgenes to the *Hprt* locus on the mouse X chromosome by homologous recombination (159, 160). Integration of the Pleiades promoter constructs created a chimaeric *HPRT/Hprt* locus that consisted of human *HPRT* promoter and exon 1 and mouse *Hprt* exons 2-9 (Figure 3.1). Upstream of the *HPRT/Hprt* locus is a MiniP or MaxiP construct. MiniP constructs contained a human promoter of 4 kb or less in size that drives a reporter (*lacZ*, EGFP, or EGFP/cre), while MaxiPs were derived from BACs of up to 195 kb with the reporter (*lacZ* or EGFP) inserted at the start codon of the gene of interest on the human BAC. To determine whether the constructs were subject to the *cis*-regulation of XCI now that they were X-linked, the Simpson laboratory generated female mice carrying an *Xist* deletion (*Xist*^{1lox}; 161) on the X chromosome lacking the knock-in, thereby causing the Pleiades knock-in to always be on the Xi. The MaxiPs *AMOTL1* (Ple5), *MAOA* (Ple127), *NOV* (Ple134), *NR2E1* (Ple142), and *NR2F2* (Ple143) were not expressed in the brains of *Xist*^{1lox} females based on *lacZ* reporter staining performed by the Simpson laboratory, but were expressed in various parts of the brain in females without the *Xist* deletion (examples of *lacZ* staining shown in Figure 3.2A), indicating that these MaxiPs are only expressed when present on the Xa and are thus subject to XCI. DNA methylation analysis on DNA from hemizygous male and heterozygous female mice transgenic for *AMOTL1*, *NOV*, and *NR2E1* showed that CpG island promoters on the BACs were significantly DNA hypermethylated in ear notch samples of females compared to males of the same strain ($p < 0.05$; Figure 3.2B), in agreement with the XCI statuses assessed by *lacZ* expression in the females with non-random XCI. *NR2E1* promoter DNA methylation was also examined in a limited number of liver and brain samples, and consistently showed a female-specific DNA hypermethylation as in the ear notch samples (Figure 3.3). We conclude that DNA methylation is a reliable indicator of XCI status for transgenes at *Hprt*. We therefore used DNA methylation to determine the XCI of the remaining MaxiPs that lacked detectable expression even when present on the Xa. *PITX2* showed less female DNA methylation than the other MaxiPs, which could either reflect the presence or absence of *cis*-acting regulators of XCI, or a tendency to be preferentially located on the Xa. To examine the latter possibility we examined the DNA methylation of the genes flanking the MaxiP, *Phf6* and *HPRT*.

3.2 DNA methylation of flanking genes reflects both skewing of XCI and differential capacity for DNA methylation on the Xi

The MaxiP construct and *HPRT* are closely linked, and thus if substantial skewing of XCI is present, then their DNA methylation levels will be correlated among different samples. In contrast, *Phf6* DNA methylation should not be affected by skewing since it is present on both X

chromosomes. Both promoters demonstrated significant DNA hypermethylation in females compared to males (*HPRT*: female average 38%, male average 5%, $p < 0.0001$; *Phf6*: female average 34%, male average 5%, $p < 0.0001$), indicating that both neighbouring genes were generally subject to XCI. Compared to *Phf6* DNA methylation, *HPRT* showed higher variability in promoter DNA methylation levels between female mice (standard deviations: 10% for *HPRT*, 4% for *Phf6*), suggesting variability in skewing of XCI in the samples analyzed. A correlation between the DNA methylation levels at the MaxiP promoter and *HPRT*, but not with *Phf6*, was observed (Figure 3.4), suggesting that there is skewing of XCI in the analyzed ear notch samples. Intriguingly, different MaxiPs showed different slopes in the correlation of their DNA methylation level with *HPRT* (Figure 3.4B), suggesting that different Pleiades promoters have different capacities for DNA methylation when located at the same site on the Xi. To confirm that different constructs had different levels of DNA methylation on the Xi, we analyzed the promoter and *HPRT* DNA methylation in females homozygous for the knock-in and in *Xist*^{1lox} females who carry the knock-in solely on the Xi. The *AMOTL1*, *NOV* and *NR2E1* MaxiPs showed similar levels of *HPRT* DNA methylation on the Xi (~70%) but slightly different levels of promoter DNA methylation on the Xi (Figure 3.4B). DNA methylation levels at *PITX2* and *NGFR* were strikingly different from the other MaxiPs.

PITX2 (CpG island 46) showed a much lower range of DNA methylation compared to DNA methylation of *AMOTL1*, *NOV*, and *NR2E1*, and DNA hypermethylation of *HPRT* indicates that the low female DNA methylation in *PITX2*-CpG46 (Figure 3.4B) is not attributable to skewing of XCI but to its intrinsic resistance to accumulate DNA methylation. The promoter assayed in the *NGFR* BAC, in contrast, showed a lower *HPRT* DNA methylation range (13-33%) compared to the other MaxiPs, suggesting that the capacity of *HPRT* to accumulate DNA methylation is altered in this construct. We designed a DNA methylation assay ~720 bp downstream of the start codon in the *lacZ* reporter, which showed similar DNA methylation levels on the Xi for all constructs except for the *NGFR* BAC (Figure 3.5). DNA methylation levels of *lacZ* and *HPRT* were also analyzed in female mice homozygous for the *MAOA* transgene. The *NGFR* BAC showed lower levels of *HPRT* and *lacZ* DNA methylation on the Xi than expected (*HPRT* average 41%, outlier; *lacZ* average 56%), suggesting the region is subject to substantial influence from the genomic context. Therefore, *PITX2* showed the most decrease in capacity to accumulate promoter DNA methylation and the *NGFR* BAC showed an impact on *HPRT* DNA methylation. To understand the *cis*-modulatory effect of integrated DNA, we explored the *PITX2* and *NGFR* BAC transgenes in more detail.

3.3 *PITX2* is DNA hypermethylated at transcription start site as well as intragenic and intergenic CpG islands

The DNA hypomethylated CpG island 46 is not annotated as the start of a transcript, so we analyzed the DNA methylation levels at eight additional locations on the *PITX2* BAC including exon 1 and intron 2 (non-CpG island), three internal CpG islands, the promoter CpG island of the alternative isoform, and two intergenic CpG islands (Figure 3.6). Although the first exon does not contain a CpG island, it still showed significantly higher DNA methylation in females than in males ($p=0.0084$; Figure 3.6). In fact, all the locations tested in *PITX2* generally showed DNA hypermethylation in females compared to males, including the CpG island at the alternative promoter and the intergenic CpG islands. In a recent study, Ernst *et al.* (162) mapped multiple chromatin modifications across nine human cell types and classified the different patterns of chromatin modifications into 15 chromatin states corresponding to various genomic regions such as promoters, enhancers, and insulators. The chromatin modifications associated with active promoters were found to overlap the assays in intron 2 and CpG islands 46 and 196 on the *PITX2* BAC (Figure 3.6; 162), suggesting *PITX2* has additional internal promoters. Intergenic CpG islands 59 and 29 show no or very weak chromatin modifications associated with promoters or enhancers (respectively) yet both CpG islands showed female-specific DNA hypermethylation (Figure 3.6). Interestingly, *lacZ* showed a clear difference in DNA methylation levels between males and females (Figure 3.6), in agreement with the DNA methylation status of multiple sites in *PITX2*. Thus, while CpG islands 18 and 46 showed lower female DNA methylation (average 14%), because other locations in the gene consistently showed DNA hypermethylation in females at levels consistent with XCI, we conclude that *PITX2* is likely subject to XCI based on DNA methylation.

3.4 A truncated gene on *NGFR* BAC construct partially escapes from XCI

A distinguishing characteristic of the *NGFR* construct from the other MaxiPs is the presence of a truncated gene at the end of the BAC that is adjacent to the *HPRT/Hprt* locus (Figure 3.7A). The *PHB* gene is truncated within the 3'UTR approximately 200 bp from the end of the gene, and we hypothesized that if *PHB* escapes from XCI, the run-on transcription from *PHB* through the *HPRT/Hprt* locus positioned ~1.7 kb downstream could be the cause of the reduced *HPRT* DNA methylation on the Xi (Figure 3.4B). We therefore examined the transcription levels of *PHB* and the intergenic region between *PHB* and *HPRT/Hprt* in males and in females with and without the *Xist* deletion (Figure 3.7B). By qPCR, we showed that *PHB* was not a highly expressed gene (using *Pgk1* as control), but could be expressed from the Xi in *Xist*^{1lox} females up to 30% of the level of expression in males (Figure 3.7B), while females with random XCI showed a level of *PHB* expression close to 60% of that in males. Furthermore, one of the *Xist*^{1lox} females showed

no expression of *PHB* from the Xi, indicating that *PHB* was subject to XCI in this individual. The differences in *PHB* expression level from the Xi between females may reflect the variability in XCI status between individuals. However, while the expression level at ~580 bp downstream of the truncated *PHB* gene in the intergenic region was essentially the same as at the 3'UTR of *PHB* (Figure 3.7B), this transcription had ceased by ~1.4 kb downstream of *PHB* (~250 bp upstream of *HPRT/Hprt*), indicating that there is no substantial run-on transcription through the *HPRT/Hprt* locus. Analysis of *HPRT* expression showed that *HPRT/Hprt* remained inactivated on the Xi despite its proximity to an escapee. In agreement with the *PHB* expression analysis, the promoter of *PHB* has an IC island that showed relatively low DNA methylation in females who had higher levels of *PHB* expression from the Xi, but still distinct from the level of DNA methylation on the Xa (Figure 3.7C). Overall, it appears that *PHB* partially escapes from XCI; however, run-on transcription through *HPRT/Hprt* is not the cause of altered *HPRT* DNA methylation capacity on the Xi.

3.5 Repeat content of candidate elements involved in XCI on the MaxiPs

Since LINE-1 and Alu have been proposed to be involved in the spread of silencing in XCI, the repeat content of LINE-1 and Alu on the MaxiPs was analyzed. Based on bioinformatic studies (127, 128) and the Lyon LINE-1 hypothesis (123), we would expect genes that are subject to XCI to be enriched in LINE-1 and depleted in Alu, and *vice versa* for the genes that escape from XCI. However, our analysis of the MaxiPs suggests that the repeat content of LINE-1 and Alu does not correlate strongly with the XCI of the MaxiPs (Figure 3.8 A and B). The LINE-1 and Alu densities were also plotted across the *NGFR* BAC and the *PITX2* BAC in overlapping 10 kb windows (Figure 3.8 C and D). Both *NGFR* and *PITX2* are located in regions of low LINE-1 content, while the gene *PHB* that escapes from XCI possesses higher level of LINE-1 at its promoter region, arguing against the role of LINE-1 as a dominant waystation in this region. Furthermore, in contrast to the *PITX2* BAC where LINE-1 and Alu densities show differing distribution, LINE-1 and Alu densities on the *NGFR* BAC both peaked at the downstream region of *NGFR* and the upstream region of *PHB*.

Using the L1Base (166), we also examined the content of putative full-length LINE-1 on the MaxiPs and the respective chromosomes from which the MaxiP BACs originated (Table 3.1). The X chromosome possesses 12 of the 145 intact LINE-1 in the genome, with a substantial over-representation of intact LINE-1 relative to the average of all chromosomes.

3.6 All examined MiniP constructs appear to be subject to XCI

Since our MaxiP results agreed with previous reports that DNA methylation is an accurate marker for XCI status (35, 131), we next analyzed promoter DNA methylation of the MiniPs to

assess their XCI statuses. Overall, heterozygous females showed significantly higher DNA methylation levels at promoter CpG islands compared to males (female average 45%, male average 12%; $p < 0.0001$). This male-female difference in DNA methylation in the ear notch samples was independent of whether the transgenes were expressed in the brain (Figure 3.9A). The difference in DNA methylation between sexes was not significant when the transgenes were not expressed in the ear notch tissue (Figure 3.9B), likely due to reduced number of MiniPs analyzed for expression in the ear notch. DNA methylation levels were analyzed individually at 46 island-containing MiniP constructs, which originated from 23 human genes, to determine whether there were MiniPs that escaped XCI. For MiniPs that were generated from the same gene and thus shared the same core promoter sequence, the same CpGs were examined for DNA methylation levels. Almost all MiniPs showed promoter DNA hypermethylation in females compared to males, with a female and male average of 44% and 4% respectively, with the outliers removed in the analysis (Figure 3.10). In order for a transgene to be qualified as a potential escapee, we required consistent low promoter DNA methylation in multiple heterozygous female mice. Since our MiniP constructs generally showed elevated DNA methylation in female averages compared to males, we concluded that none of the MiniP constructs appeared to escape XCI. MiniPs based on *CARTPT*, *GPX3*, *ICMT*, *OXT*, and *POGZ*, were detected as outliers for DNA methylation at the promoter in males. *ICMT*, *OXT*, and *POGZ* MiniPs showed some expression in the brain according to the *lacZ* or EGFP staining performed by the Simpson laboratory, indicating that DNA hypermethylation at the promoter on the Xa was not correlated with silencing. Similar to mice carrying the MaxiPs, females carrying the MiniPs showed DNA hypermethylation at *HPRT* and *Phf6* compared to males (Figure 3.11). In general, we analyzed fewer mice per construct for the MiniPs; however, overall MiniP constructs showed higher levels of DNA methylation compared to the MaxiP constructs (average 45% and 33% respectively, $p < 0.0001$), perhaps reflecting the closer association of the MiniPs with the X-linked *cis*-acting elements.

3.7 DNA methylation of *lacZ* reporter consistently reflects the pattern at CpG island promoters

Since *lacZ* DNA methylation resembled the DNA methylation pattern of the promoter region in *PITX2*, we assessed the DNA methylation at *lacZ* in other Pleiades constructs. Similar to CpG island promoters, female mice overall showed significantly higher *lacZ* DNA methylation than males ($p < 0.0001$), with male and female average DNA methylation levels of 26% and 49%, respectively. The constructs that were analyzed for promoter DNA methylation generally showed female-specific DNA hypermethylation (Figure 3.12A). The mice with EGFP/cre as the reporter for the Pleiades constructs had an autosomal *lacZ* reporter that could be activated by

cre [*Gt(ROSA)26Sor^{tm1Sor}*; (168)]. At the autosomal location, *lacZ* showed no significant difference in DNA methylation levels between males and females, which both averaged ~45% with considerable variability (Figure 3.12B), indicating that the difference in the DNA methylation levels of the X-linked *lacZ* between the sexes is a consequence of the epigenetic regulation on the X chromosome. Although *lacZ* showed overall higher DNA methylation than the CpG island promoters (male $p < 0.0001$; female $p = 0.0029$), *lacZ* DNA methylation showed a significant correlation with DNA methylation of the promoter island in females (Figure 3.13A). Since constructs with and without CpG islands in the promoter both showed a significant difference between female and male *lacZ* DNA methylation levels ($p < 0.0001$ and $p = 0.0008$, respectively), we used *lacZ* DNA methylation as a surrogate for promoter DNA methylation and assessed the XCI status of additional Pleiades constructs for which there was no assay for promoter DNA methylation, due to difficulty in assay design, assay failure, and/or the absence of a CpG island in the promoter (Figure 3.13B). Consistent with the lack of *lacZ* expression from the Xi in *XXist^{1lox}/X^{MAOA}* females (Figure 3.2A), *MAOA* showed female-specific *lacZ* DNA hypermethylation (Figure 3.13B), further supporting the usage of this locus as a surrogate to determine XCI status. Furthermore, a DNA methylation assay designed at the 3' end of the *lacZ* reporter showed a similar pattern as the *lacZ* assay located at ~720 bp downstream of the TSS (Figure 3.13C). However, compared to promoter DNA methylation, males more often showed DNA hypermethylation of the *lacZ* reporter (Figure 3.12A and Figure 3.13). Using the criteria of non-overlapping standard deviations of DNA methylation between the sexes and a male average DNA methylation level below two standard deviations of the female average of all strains, we called an additional 11 constructs subject to XCI based on *lacZ* DNA methylation.

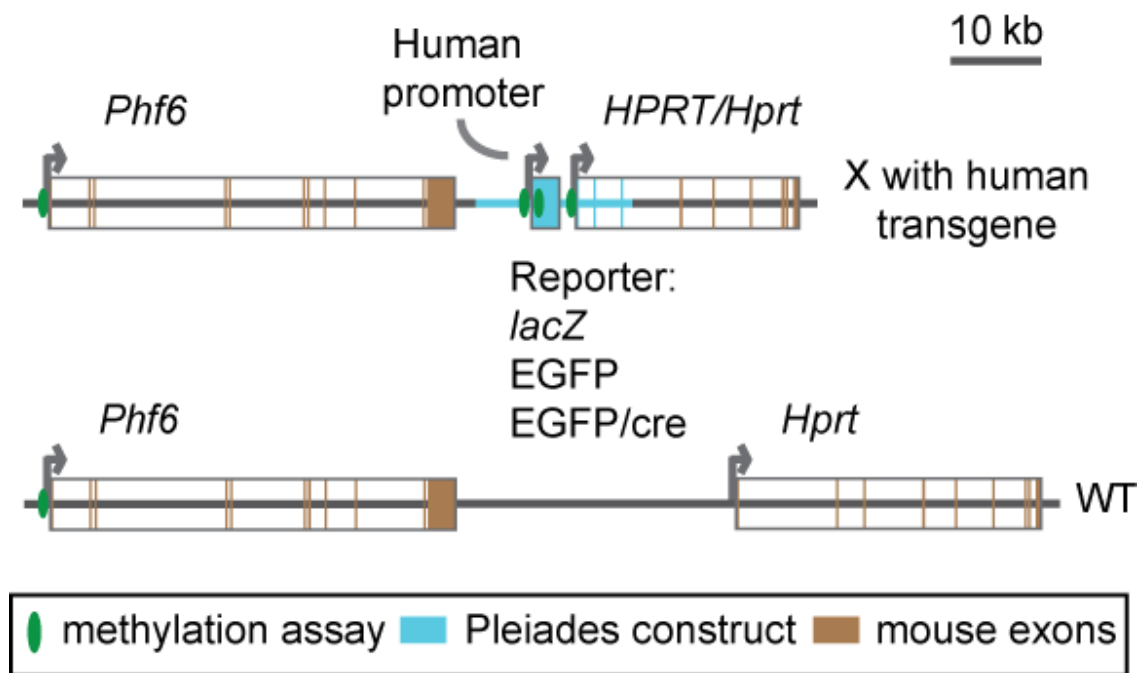


Figure 3.1 Experimental system in which human reporters (Pleiades constructs) were integrated at the *Hprt* locus on the mouse X.

A chimaeric *HPRT/Hprt* locus was generated, consisting of the human *HPRT* promoter and first exon and the mouse counterpart for the rest of the gene. The majority of the female mice examined were heterozygous for the human knock-in. The wild-type mouse locus is shown below the knock-in chromosome. The size of the Pleiades construct and the internal exons are variable and not shown to scale.

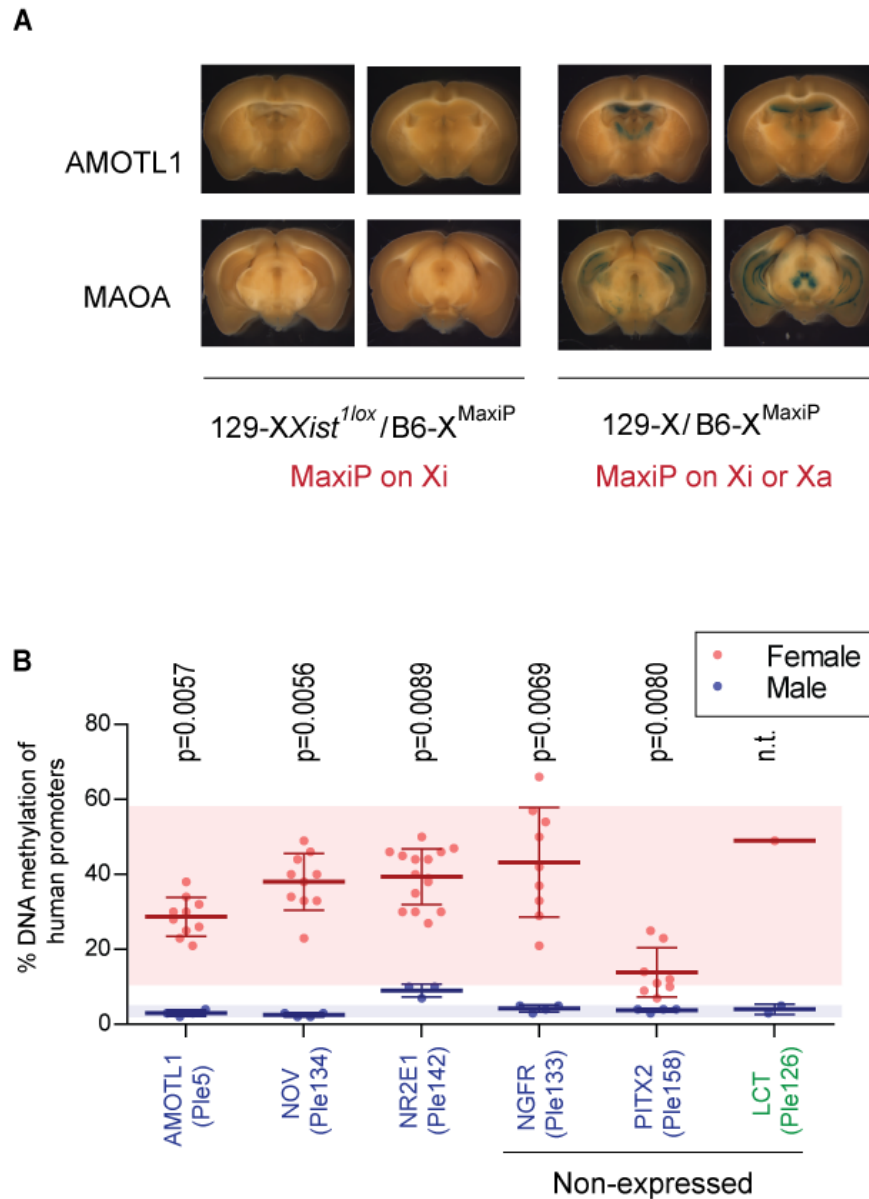


Figure 3.2 DNA methylation agreed with the XCI status of the MaxiP constructs.

(A) MaxiPs *AMOTL1*, *NOV*, *NR2E1*, *MAOA*, and *NR2F2* were not expressed when present on the Xi based on staining of *lacZ* reporter by the Simpson laboratory. The above figure shows two examples of the *lacZ* staining of *AMOTL1* and *MAOA* in the brains of females with (129-XXist^{1lox}/B6-X^{MaxiP}) and without (129-X/B6-X^{MaxiP}) the *Xist* deletion. In females with the *Xist* deletion, the MaxiP was solely present on the Xi, while in females with random XCI, the MaxiP was present on either the Xa or the Xi. Generation of females with the *Xist* deletion and *lacZ* staining of brains were performed by the Simpson laboratory. (B) MaxiPs were DNA hypermethylated in females compared to males. Constructs with *lacZ* and EGFP as the reporter are labelled in blue and green, respectively. The DNA methylation shown for the *LCT* BAC (Ple126) is the promoter DNA methylation at the *MCM6* gene present on the same MaxiP, not at the promoter of the *LCT* gene itself. Significance was tested using Mann-Whitney t-test. n.t., not tested due to the limited sample numbers. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain; shaded regions, two standard deviations from the average DNA methylation level.

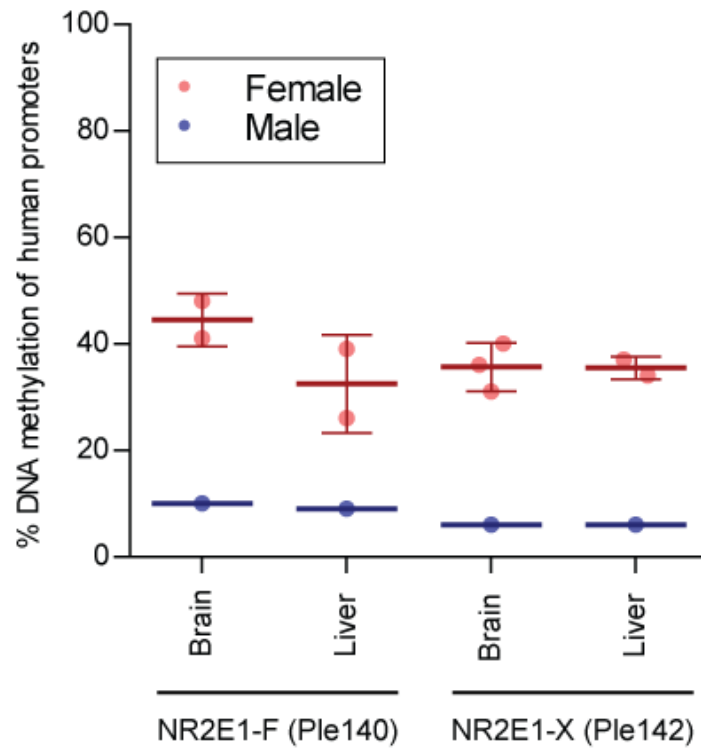


Figure 3.3 *NR2E1* promoter showed consistent DNA methylation pattern across tissues.

NR2E1 was called subject to XCI in the ear notch (Figure 3.2), brain and liver samples based on promoter DNA hypermethylation in females compared to males. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain.

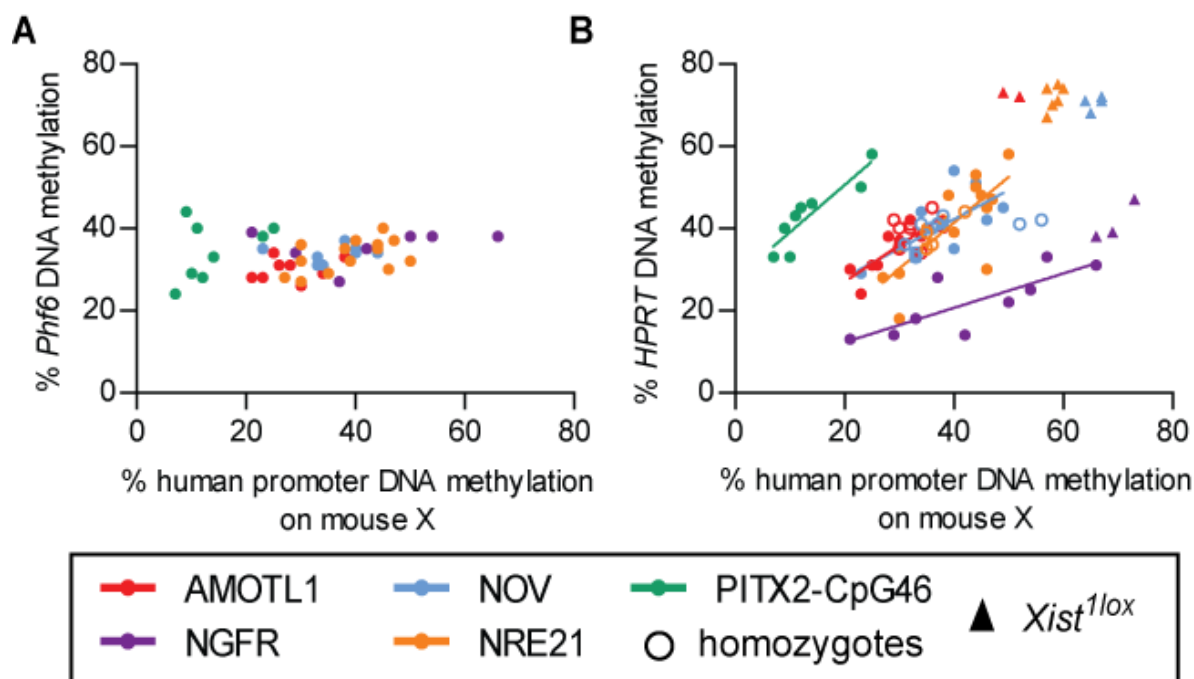


Figure 3.4 DNA methylation analysis of flanking *HPRT* and *Phf6* genes revealed differential susceptibility of MaxiP constructs to DNA methylation on the Xi.

(A) Spearman correlation analysis between *Phf6* and the MaxiP DNA methylation levels showed no significant correlations. *AMOTL1* $r=0.2532$, $p=0.5206$; *NGFR* $r=0.1482$, $p=0.7825$; *NOV*-X $r=0.2270$, $p=0.5821$; *NR2E1* $r=0.4633$, $p=0.0953$; *PITX2* $r=0.2635$, $p=0.5364$. (B) Spearman correlation analysis between *HPRT* and the MaxiP DNA methylation levels showed significant correlations, but differential capacity for DNA methylation was also observed. *AMOTL1* $r=0.8869$, $p=0.0011$; *NGFR* $r=0.8285$, $p=0.0083$; *NOV* $r=0.7439$, $p=0.0174$; *NR2E1* $r=0.6864$, $p=0.0047$; *PITX2* $r=0.9581$, $p=0.0011$. *Xist*^{1lox} females (triangles) carrying different MaxiP constructs showed different levels of DNA methylation on the Xi. Each circle or triangle shows DNA methylation levels from the ear notch sample of an individual mouse.

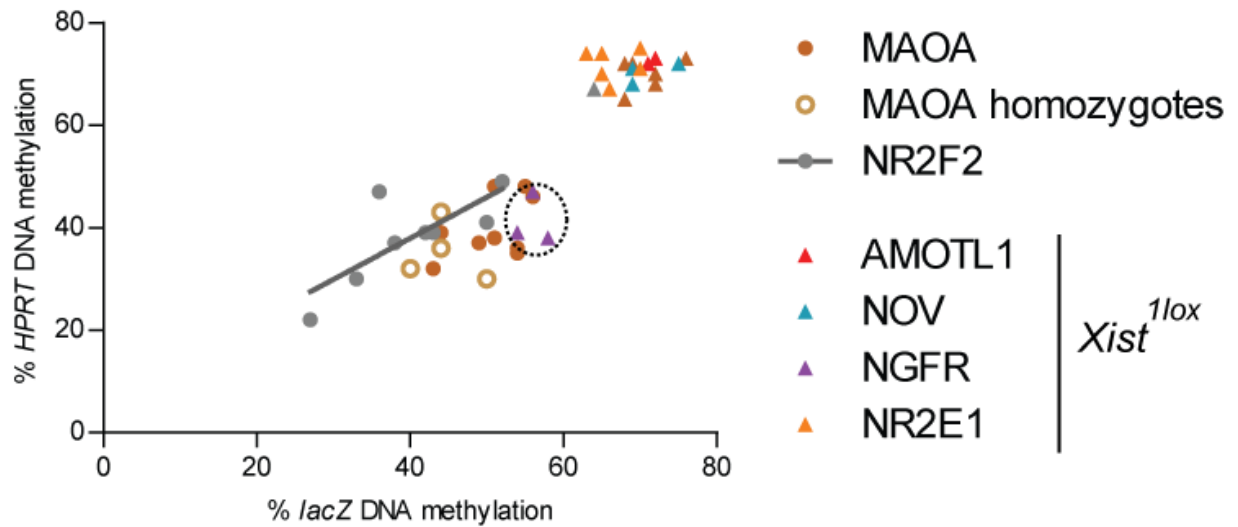


Figure 3.5 DNA methylation of *HPRT* and *lacZ* was altered for *NGFR* (Ple133).

Spearman correlation between *HPRT* and *lacZ* DNA methylation levels. *NR2F2* (Ple143): $r=0.7545$; $p=0.0368$; *MAOA* (Ple127): $r=0.4093$; $p=0.2696$. DNA methylation levels of *NGFR* in *Xist*^{1lox} females (encircled on the graph) were substantially lower than *AMOTL1*, *NOV*, and *NR2E1* in *Xist*^{1lox} females. Filled dots, DNA methylation in individual female mice which were heterozygous for the knock-in and carried wild-type *Xist* on both X chromosomes. Triangle, DNA methylation of the knock-in on the inactive X chromosome in females heterozygous for an *Xist* deletion (*Xist*^{1lox}). DNA methylation levels in four mice homozygous for the *MAOA* MaxiP were also examined (open circle).

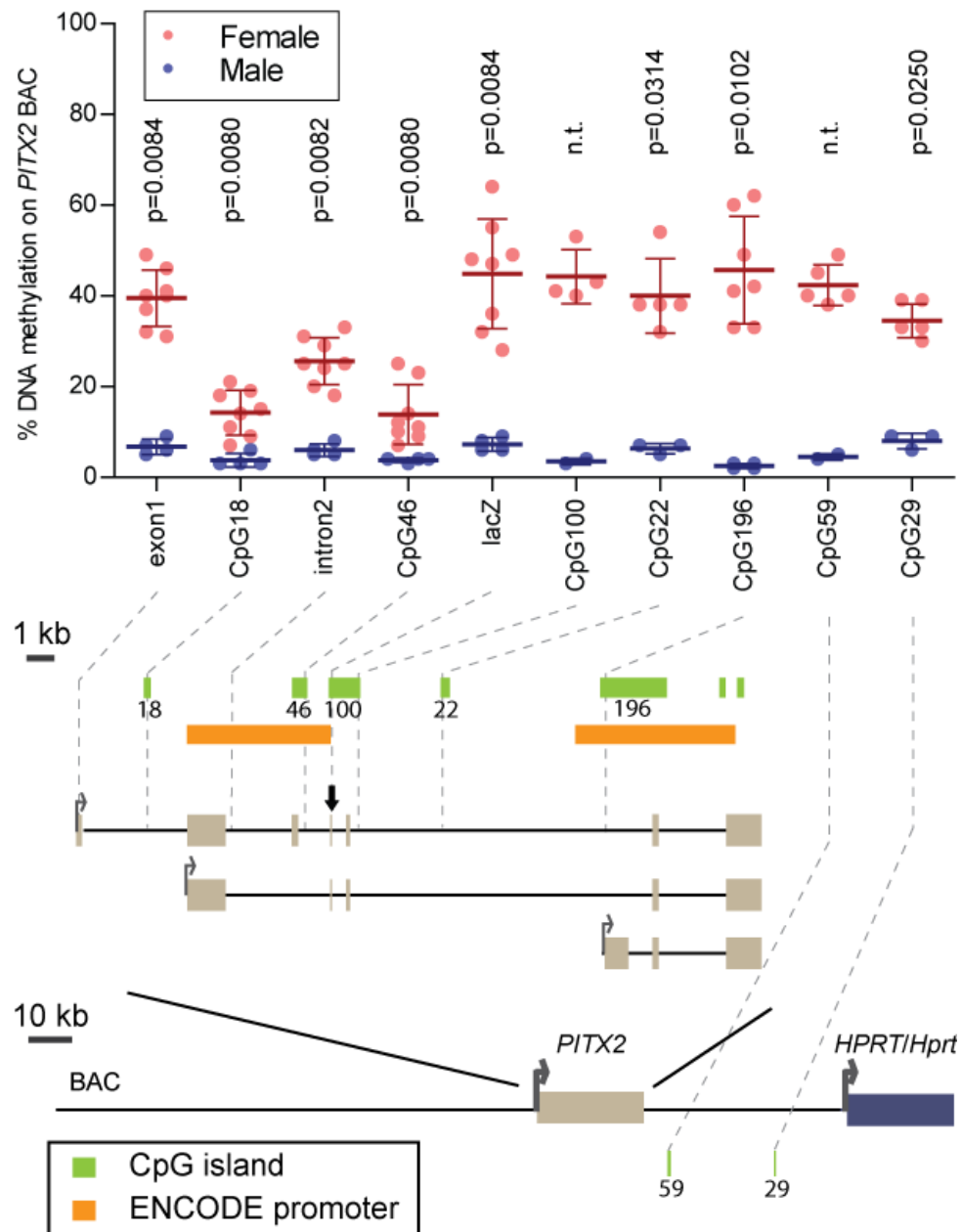


Figure 3.6 DNA methylation analysis of multiple regions in the *PITX2* BAC construct.

DNA methylation assays were designed in the gene body, within or outside of CpG islands, as well as in the intergenic CpG islands. The corresponding locations of the DNA methylation assays are shown with dotted lines to the sites on the construct. The gene is depicted with three transcript isoforms (169) and the BAC (RP11-268I1) is shown below. Location of the *lacZ* insertion is indicated with a downward arrow. Promoter track is based on the ENCODE study on chromatin states by Ernst *et al.* (162). The promoter track presented in this figure represents the the enrichment of chromatin marks associated with active promoters. Circles, DNA methylation of the individual ear notch sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain. Significance is tested using Mann-Whitney t-test. n.t., not tested.

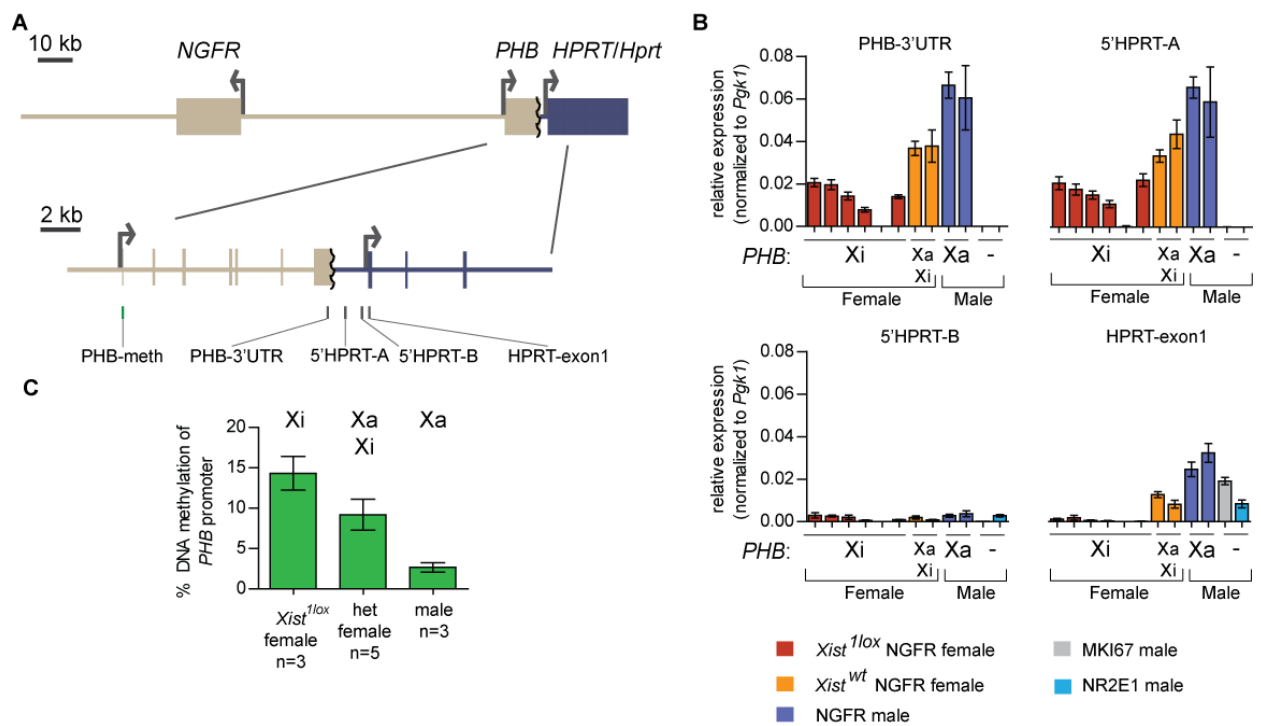


Figure 3.7 *PHB* on the *NGFR* BAC partially escaped from XCI.

(A) Structure of the *NGFR* MaxiP construct. *PHB* is truncated in the 3'UTR on the BAC, located ~1.7 kb from the *HPRT/Hprt* locus. Grey and green bars below the MaxiP indicate the locations of the expression and DNA methylation assays, respectively. Internal exons not shown to scale. (B) Expression of *PHB*, intergenic region between *PHB* and *HPRT/Hprt*, and *HPRT/Hprt* (exon1), normalized to *Pgk1*. DNA from *MKI67* (Ple131) and *NR2E1* (Ple142) served as negative controls (-) since they lack *PHB* gene. Error bars indicate standard deviation between two qPCR runs. The x-axis indicates whether the Pleiades construct was present only on the Xi or the Xa in a given mouse, or on either X chromosomes (Xa Xi), as in the case for females with random XCI. (C) DNA methylation of *PHB* promoter in mice carrying *NGFR* MaxiP. Error bars indicate the standard deviation between mice.

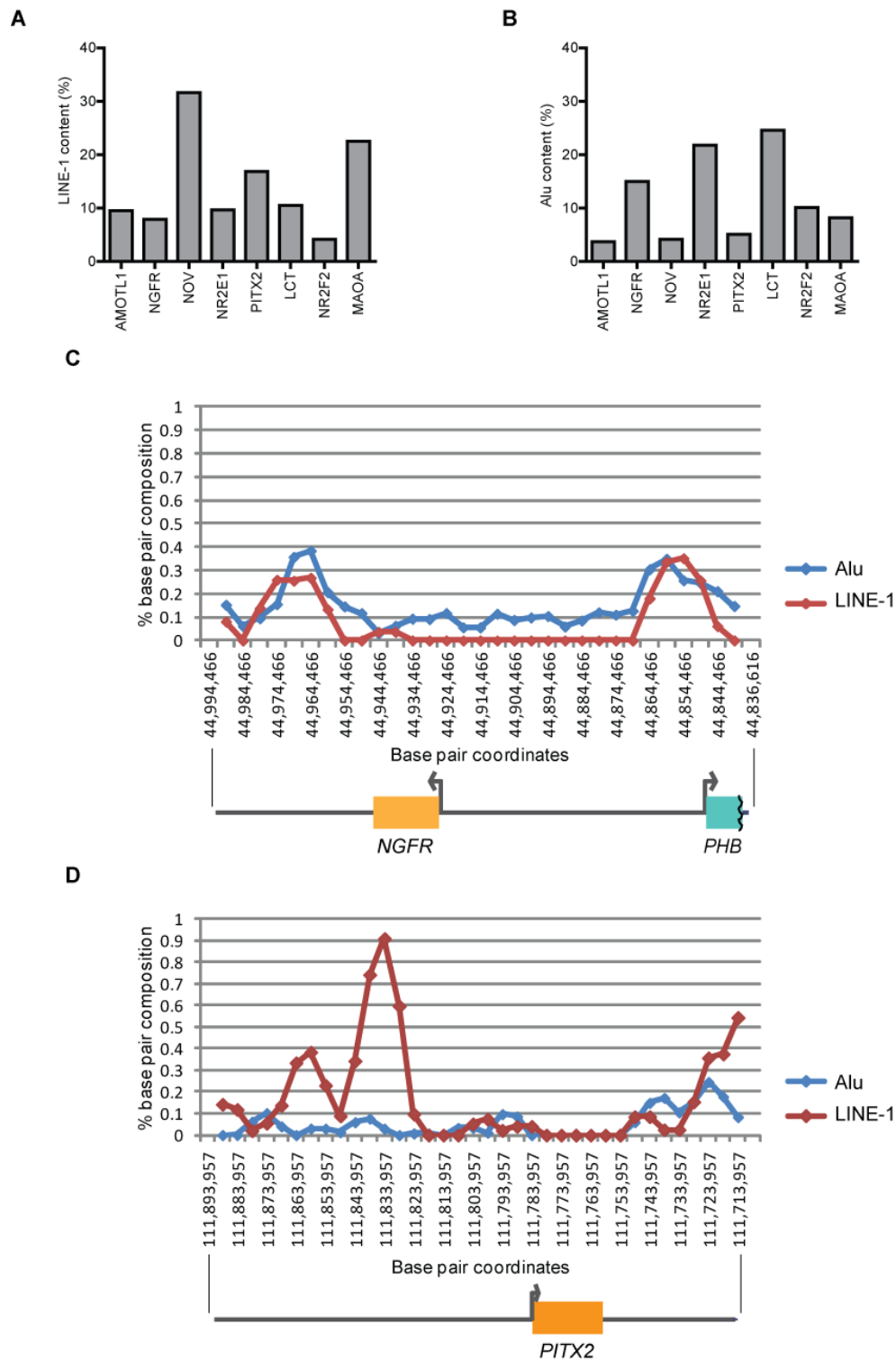


Figure 3.8 Base pair composition of repetitive elements on the MaxiPs.

Repeat content for LINE-1 (A) and Alu (B) on the BACs did not correlate strongly with the XCI statuses of the MaxiPs. Genes that are subject to XCI are predicted to be enriched in LINE-1 and depleted in Alu, while genes that escape from XCI are predicted to be depleted in LINE-1 and enriched in Alu. Distribution plots of LINE-1 and Alu across overlapping 10 kb windows across the *NGFR* BAC and the *PITX2* BAC are shown in (C) and (D), respectively. The coordinates show the midpoint of each window.

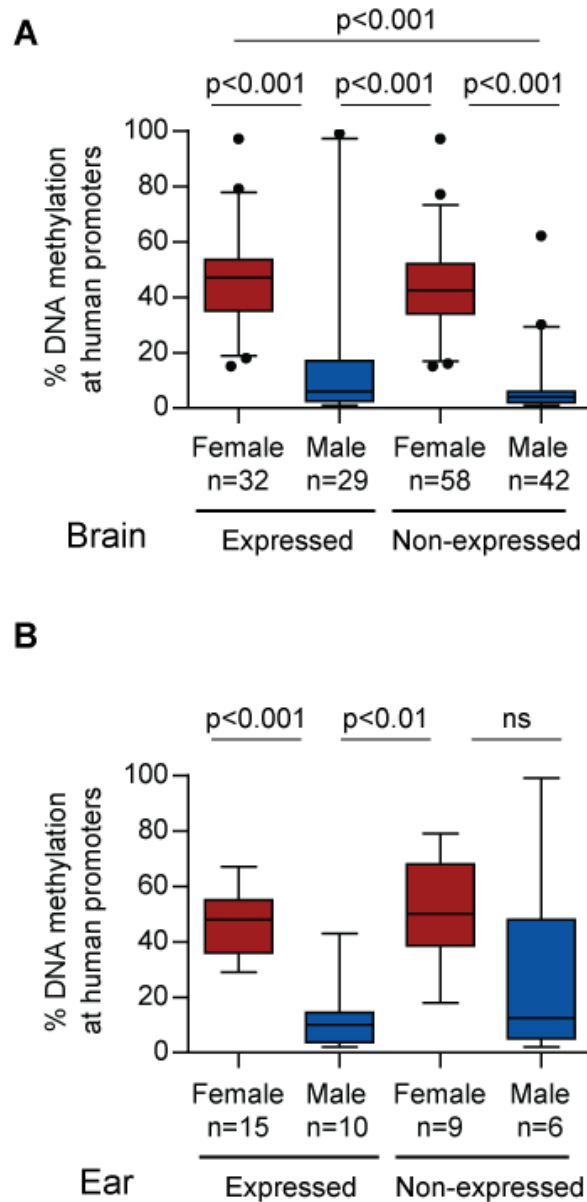


Figure 3.9 DNA methylation reflects XCI status independent of transgene expression.

(A) Females showed significant DNA hypermethylation of MiniPs in ear notch samples compared to males regardless of whether the transgene was expressed in the brain. (B) Females showed partial DNA methylation of MiniPs independent of transgene expression in the ear notch but females did not show significant DNA hypermethylation compared to males when the transgene is not expressed in the ear notch. Expression status (expressed and non-expressed) was based on *lacZ* or EGFP staining in the brains or ear notches of mice. Significance was determined using one-way ANOVA Kruskal-Wallis test. ns, not significant.

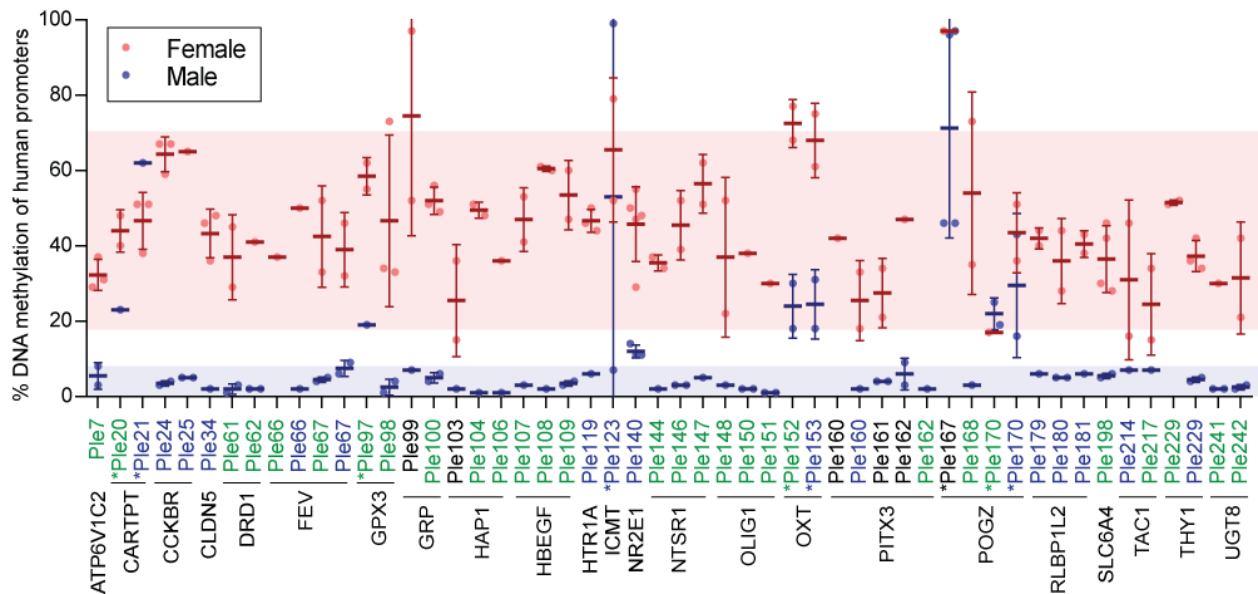


Figure 3.10 Promoter DNA methylation of the MiniP constructs.

DNA methylation levels of individual MiniP constructs in ear notch samples. Constructs with *lacZ*, EGFP, and EGFP/cre as the reporter are colored in blue, green and black, respectively. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain; shaded regions, two standard deviations from the average DNA methylation level with outlier strains removed. Outliers are marked with asterisks (*). Female outlier: *POGZ* (Ple167). Male outliers: *CARTPT* (Ple20, Ple21), *GPX3* (Ple97), *ICMT* (Ple123), *OXT* (Ple152, Ple 153), *POGZ* (Ple167, Ple170). Modified Z-score greater than 3.5 in absolute values were marked as outliers (167).

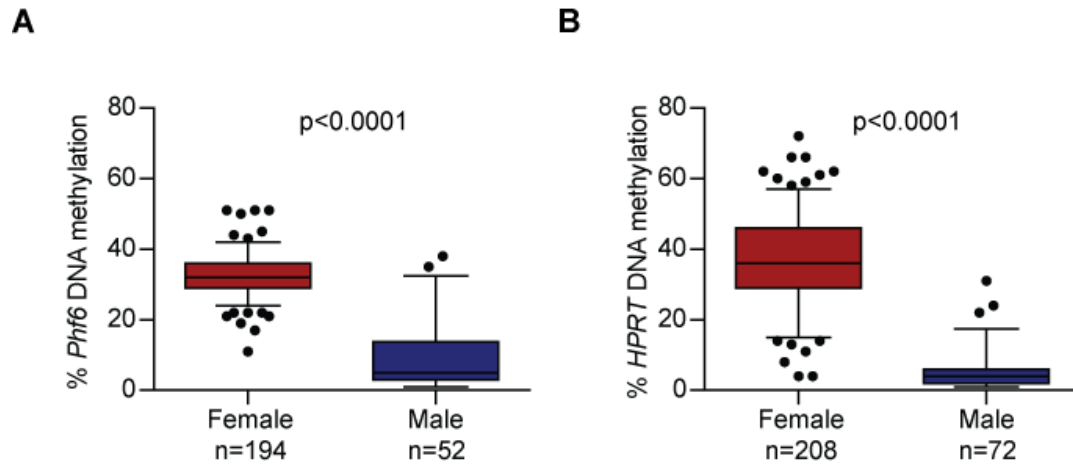


Figure 3.11 Promoter DNA methylation of genes flanking the MiniPs.

(A) *Phf6* promoter and (B) *HPRT* promoter both showed significant difference in DNA methylation levels between male and female ear notch samples. Mann-Whitney t-test was used to test for significance. Boxplot whiskers are 5-95% percentile. Circles, DNA methylation levels of individual mice.

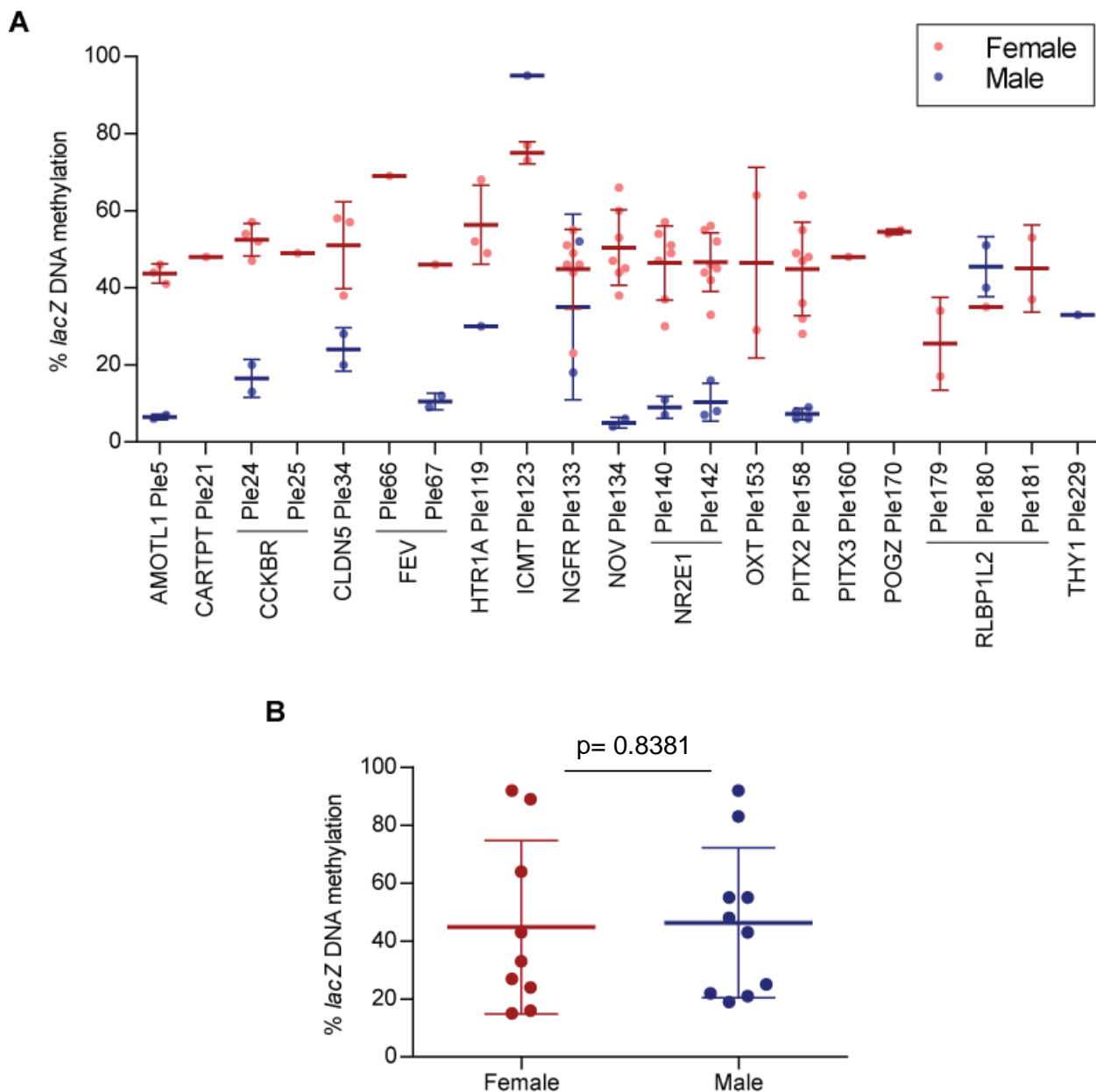


Figure 3.12 *lacZ* reporter exhibited female-specific DNA hypermethylation when present on the X chromosome.

(A) DNA methylation of the reporter *lacZ* in constructs where the promoter DNA methylation has also been analyzed. *lacZ* generally shows DNA hypermethylation in females compared to males but is more frequently DNA hypermethylated in males compared to promoter DNA methylation. Circles, DNA methylation levels from an individual mouse; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain (B) DNA methylation of autosomal *lacZ* reporter. Each circle presents the level of DNA methylation in an individual mouse that carried an autosomal *lacZ* that could be activated by cre. The promoters driving cre may differ between individuals. Bar, average; error bars, standard deviation between mice. Significance was tested using Mann-Whitney t-test.

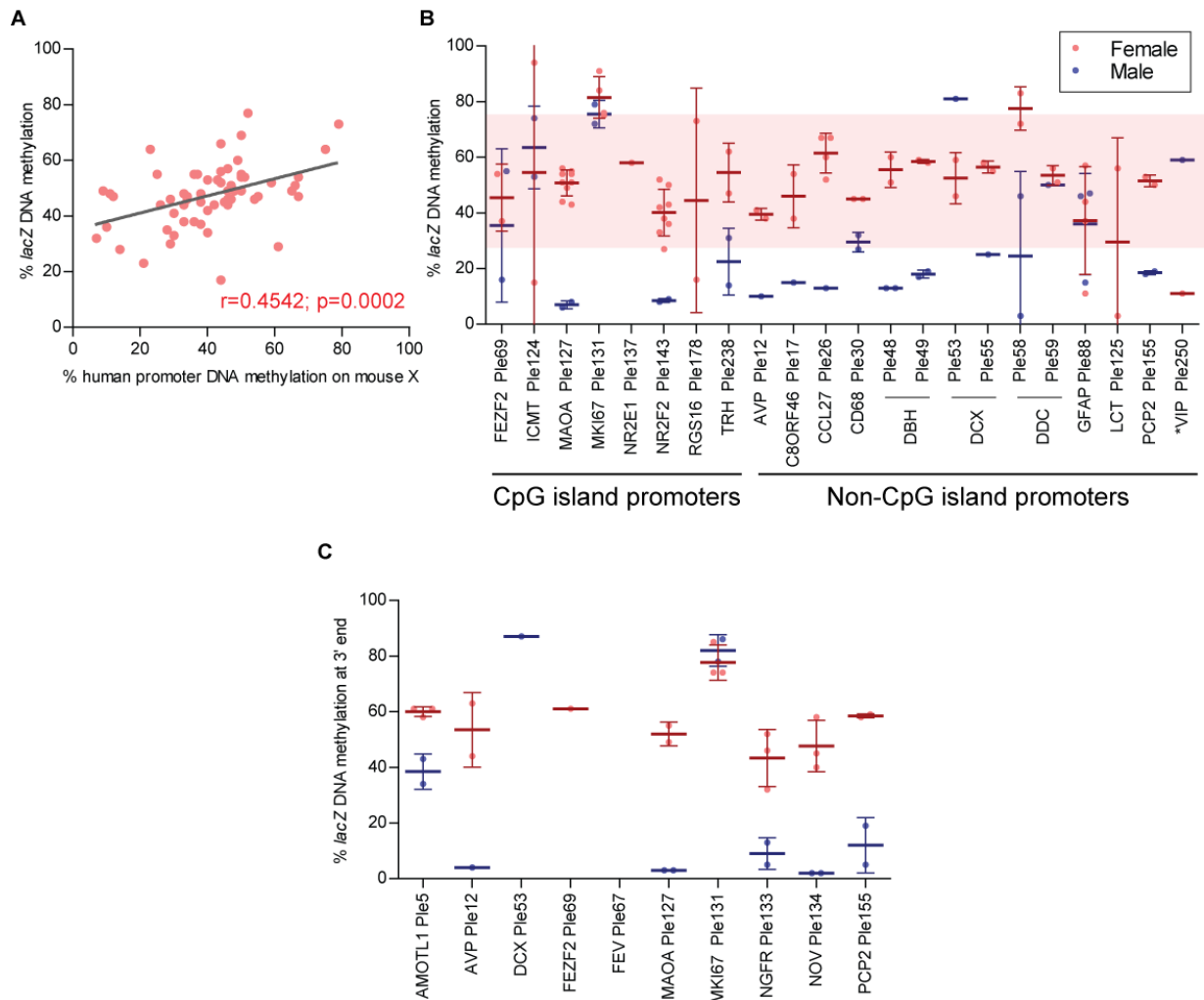


Figure 3.13 *lacZ* DNA methylation can be used as a surrogate for promoter DNA methylation.

(A) Spearman correlation between *lacZ* and promoter DNA methylation levels in females. Circles, DNA methylation levels from an individual mouse. (B) *lacZ* DNA methylation levels of the Pleiades constructs that do not have a DNA methylation assay in the promoter region due to difficulty in assay design, assay failure, or absence of a CpG island. All constructs shown here have *lacZ* as the reporter on the X chromosome. Circles, DNA methylation levels from an individual mouse; bar in the center of the error bars, average DNA methylation for the strain; error bars, standard deviations between mice for the strain; shaded regions are the two standard deviations from the female average DNA methylation level with the outlier strains removed. Female outlier (*): *VIP* (Ple250). (C) DNA methylation levels at ~480 bp from the end of the *lacZ* reporter. DNA methylation levels at the 3' end of *lacZ* are generally similar to the levels at the 5' end.

Table 3.1 Number of full-length LINE-1 on the MaxiPs.

According to the L1Base, there are 145 putative full-length LINE-1 in the human genome and the number of full-length LINE-1 on the chromosome of interest was obtained from the L1Base (166). Detection of full-length LINE-1 on the MaxiPs was done using the L1Xplorer. The expected number of full-length LINE-1 for the chromosomes was calculated based on proportion of the chromosome size to the genome.

MaxiP	Chr	Number of full-length LINE-1 on the chr	Expected number of full-length LINE-1 on the chr	Number of full-length LINE-1 on MaxiP
<i>AMOTL1</i> (RP11-936P10)	11	8	6.5	0
<i>LCT</i> (RP11-406M16)	2	10	11.7	0
<i>MAOA</i> (RP11-475M12)	X	12	7.5	1
<i>NGFR</i> (RP11-158L10)	17	1	3.9	0
<i>NOV</i> (RP11-840I14)	8	7	7.1	4
<i>NR2E1</i> (RP11-144P8)	6	10	8.3	0
<i>NR2F2</i> (RP11-134D15)	15	5	4.9	0
<i>PITX2</i> (RP11-268I1)	4	10	9.3	0

Chr, chromosome. The MaxiPs that carry full-length LINE-1 are bolded.

4 Discussion

Arguably, the most dramatic example of *cis*-regulation is the silencing of one X chromosome in females. The *cis*-acting elements involved in spreading silencing along the ~155 Mb X chromosome remain unknown. Having 74 transgenes integrated into the X chromosome presented us with an opportunity to assess *cis*-regulation of ~1.5 Mb of DNA. DNA methylation has been shown to be a reliable predictor of XCI status of X-linked genes, such that DNA hypermethylation at the promoter in females suggests inactivation of the gene (54, 56). Therefore, in this study, we examined DNA methylation of human transgenes integrated at the *Hprt* locus on the X chromosome to determine their XCI statuses in mice. Crossing the knock-in mice with an *Xist* deletion to cause non-random XCI of the transgenic chromosome validated that *MAOA*, *AMOTL1*, *NOV*, *NR2F2* and *NR2E1* are subject to XCI, as predicted by DNA methylation. In addition, *PHB* exhibits low promoter DNA methylation in males and females, consistent with escape of the gene based on expression analysis. We found that 92% of constructs analyzed showed DNA hypermethylation of the human promoters in female mice compared to males (Figure 3.2 and Figure 3.10), suggesting that these constructs are subject to XCI. While assessment of XCI status can be confounded by either skewing of XCI or variability between females in their XCI status for the same X-linked genes (99), the inclusion of *HPRT* DNA methylation and a criteria for consistently low promoter DNA methylation in multiple females make the use of DNA methylation a robust method for the rapid determination of XCI status.

In general, promoter CpG islands are unmethylated on the autosomes; however, ~4% are reported to show DNA methylation, often with variability between tissues (39). Thirty-one of 35 autosomal CpG islands (promoter or non-promoter associated) analyzed showed low DNA methylation levels in both male and female cell lines and/or blood samples as expected [data not shown; (170)], indicating that the female-specific DNA hypermethylation in the knock-in mice was a consequence of movement to the X chromosome and subsequent XCI. DNA methylation >20% was observed at four CpG islands (*CARTPT*, *OXT*, *THY1*, and CpG island 29 on the *PITX2* BAC) and exon 1 of *PITX2* at the endogenous loci. All of these loci lost DNA methylation on the Xa in males except for *CARTPT* and *OXT*, which retained some DNA methylation at the *Hprt* integration site on the Xa in males. Three loci, *GPX3*, *ICMT* and *POGZ* appeared to variably gain DNA methylation when integrated into *Hprt*. In general we observed a dominant regulation of XCI on promoter DNA methylation, with the majority of the promoters exhibiting an altered pattern of DNA methylation when positioned on the X chromosome. This dominant control of XCI is consistent with our results that showed female-specific promoter DNA hypermethylation regardless of whether the transgenes were expressed (Figure 3.9), and is

consistent with genes in humans, such as the human X-linked *AR* gene which shows DNA hypermethylation on the Xi relative to the Xa independently of expression (58).

While promoter CpG islands are recognized to be DNA hypermethylated on the Xi when the genes are subject to XCI, it is less clear what the DNA methylation pattern is on the rest of the X chromosome, with reports of Xa-specific DNA methylation in gene bodies and non-promoter CpG islands (62, 63). Our results with the *PITX2*-containing BAC demonstrated that the influence of XCI on DNA methylation of transgenes not only applies to promoters, but also gene body and intergenic CpG islands, since all analyzed locations on the BAC demonstrated female-specific DNA hypermethylation that was not observed on the endogenous human chromosome (Figure 3.6). While some of the analyzed regions may be unannotated promoters, CpG island 29 does not appear to be associated with promoters (162) or overlap with conserved transcription factor binding sites (TRANSFAC Biobase, <http://www.gene-regulation.com/pub/databases.html>), yet showed DNA hypermethylation in the females compared to males. Therefore, it is possible that the default state of CpG islands on the X chromosome is to acquire DNA methylation on the Xi, consistent with the majority of the CpG islands being DNA hypermethylated on the Xi (63).

The *lacZ* reporter, ~3000 bp with a GC content of 56% and an observed/expected CpG ratio of 1.196, is essentially a single large CpG island. The CpG-rich *lacZ* may be recognized and regulated more like a promoter than a gene body, and therefore shows DNA hypermethylation in females compared to males on the X chromosome (Figure 3.12 and Figure 3.13). Promoter-less artificial CpG islands inserted into the 3'UTR of an autosomal and an X-linked gene have been shown to recruit unmethylated CpG-binding protein Cxxc1 (previously known as Cfp1) and the promoter histone mark H3 lysine 4 methylation even in the absence of RNA polymerase II binding (171), although the X-linked locus has some DNA methylation presumably due to XCI. The ability of CpG-rich sequences to acquire characteristics of promoters further supports using *lacZ* DNA methylation as a surrogate for promoter DNA methylation to predict whether transgenes are subject to XCI. However, *lacZ* DNA methylation is not as robust a predictor of inactivation as the promoter DNA methylation, due to a higher frequency of males with DNA hypermethylation.

Through analyzing MaxiP DNA methylation of female mice with complete non-random XCI from an *Xist* deletion, we showed that different MaxiPs could accumulate DNA methylation at the promoters to different extents on the Xi (Figure 3.4). However, the shared DNA sequences on the MaxiP constructs, the *HPRT* promoter and *lacZ* reporter, generally exhibited similar levels of DNA methylation (Figure 3.5), suggesting that the capacity to accumulate DNA methylation is a

characteristic of the DNA sequence. Intriguingly, there may be differences between hemizygous and heterozygous, or between homozygous and heterozygous situations, as the observed promoter DNA methylation levels in females with the *Xist* deletion tend to be lower than the expected level of DNA methylation on the Xi based on the assumption that the DNA methylation on the Xa is equivalent between males and females.

The limited spread of XCI into autosomes has led to the proposal of waystations or booster elements that spread the silencing signal along the chromosome. Because MiniPs are small transgenes, it is not surprising that they are subject to XCI, since they are in close proximity to X-linked DNA and putative waystations. Indeed, the majority of previous studies on X-linked transgenes have reported silencing of the transgenes on the Xi. Eleven copies of the chicken transferrin gene that amounted to ~187 kb of foreign DNA escaped from XCI (131, 142). This is of the size of the MaxiPs analyzed and thus we anticipated that the MaxiPs might have a higher probability of carrying *cis*-regulatory sequences that may influence their XCI state causing escape from XCI. Alternatively, the MaxiPs could potentially escape from XCI by lack of waystations on the autosomal BACs. Our results examining DNA methylation of MaxiP constructs (Figure 3.2 and Figure 3.13) demonstrated that mouse XCI is consistently capable of inactivating foreign transgenes up to 195 kb; therefore, the escape of the chicken transferrin cannot simply be explained by the inability of XCI to spread over the large size of the transgene. It is possible that the chicken transferrin escaped from XCI due to the presence of an escape element, the lack of waystations within a large domain size, and/or the genomic nature of the integration site, which may contain an escape element in proximity or is a waystation-poor region relative to *Hprt*. Interestingly, none of the conserved mouse escape genes (bolded in Table 1.1) are located at band A5, where *Hprt* is located, so the genomic environment at the *Hprt* locus may have a strong influence in promoting XCI and thus renders most of our MaxiPs subject to XCI.

If young LINE-1s can serve as waystations important in spreading the silencing of XCI, then the escape of chicken transferrin could be due to the lack of waystations since the chicken genome is depleted in young LINE-1. However, genes likely require the depletion of waystations in conjunction with a large domain size, since the 21.4 kb chicken lysozyme transgene at the *Hprt* locus is still subject to XCI (140), suggesting that waystations can act over large distances. This then raises the question why most of our MaxiPs, which are of a similar size to the chicken transferrin transgene, do not escape from XCI. Based on the LINE-1 and Alu repeat content on the MaxiPs, we showed that some MaxiPs remain subject to XCI despite appearing to possess an environment that resembles the genomic context of escapees (Figure 3.8 and Table 3.1).

The *NGFR* BAC which harbours both the X-inactivated *NGFR* and the partially escaping *PHB*, contains relatively low LINE-1 and high Alu content compared to the other MaxiPs. The repeat content of the *NGFR* BAC suggests that its genomic environment may be more favourable for genes to escape from XCI, in agreement with the XCI status of *PHB*, but at the same time shows that LINE-1 and Alu elements likely do not play a dominant role in determining the XCI status of genes in this case, since *NGFR* remains subject to XCI. Similar to the *NGFR* BAC, the repeat content of LINE-1 and Alu would have predicted *NR2E1* and the *MCM6* gene (on the *LCT* BAC) to escape from XCI, and thus disagrees with the XCI statuses of *NR2E1* and *MCM6* based on promoter DNA methylation and/or *lacZ* expression analyses. In contrast, *MAOA*, *NOV*, and *PITX2* BACs contain relatively high LINE-1 and low Alu content, which is consistent with their XCI statuses. The repeat content for LINE-1 and Alu are both relatively low on the *AMOTL1* and the *NR2F2* BACs, therefore it is unclear whether the genomic environment promotes the spread of silencing or the escape from XCI. Furthermore, our analysis of intact LINE-1 content showed that chromosome 17 (from which the *NGFR* BAC originated) contains the least number of full-length LINE-1 of the chromosomes analyzed, lower than the expected number of 3.9. It is therefore not surprising that the *NGFR* BAC does not contain an intact LINE-1. The most striking observation is that the *NOV* BAC harbours four of the seven full-length LINE-1 present on chromosome 8. The presence of intact LINE-1 on *NOV* and *MAOA* BACs is consistent with their XCI statuses, but full-length LINE-1 is not required on the BACs for the spread of silencing since the other X-inactivated MaxiPs do not contain intact LINE-1. Similarly, in mice, genes in proximity to full-length LINE-1 appear to be silenced more efficiently than the LINE-1-deprived genes, but the presence of intact LINE-1 alone is insufficient for efficient silencing (126).

One possible explanation for the discrepancy between the XCI status and the repeat content is that the LINE-1 content, albeit low in base coverage on the MaxiPs, was still sufficient to propagate the inactivating signals on the X chromosome, while the chicken transferrin transgene was completely depleted in waystations. Overall, it seems that high content of waystations in proximity is not necessary for the spread of silencing, since waystations likely act over long distances greater than 100 kb. Furthermore, since LINE-1 and Alu content does not strongly correlate with the XCI status of MaxiPs, there may be additional repeats or a combination of other factors that, together with LINE-1 and Alu, contribute to the silencing or the escape of genes. Perhaps, waystations are composed of a variety of different types of sequences, and thus are redundant in function, so examining a limited number of repetitive elements is inconclusive in determining the XCI statuses of genes. Finally, much of the discussion here on elements involved in XCI has been based on the assumption that the

analyzed repeats LINE-1 and Alu are regulatory sequences involved in XCI, but it remains possible that waystations are other DNA elements that have not yet been identified.

In contrast to waystations, escape elements are proposed to aid in the escape of genes from XCI. Escape elements are presumably outside the promoter as 46 MiniPs are subject to XCI. We determined that the *PHB* gene escapes from inactivation. As waystations are reduced in abundance on autosomes and *NGFR* is subject to XCI while being farther from the mouse X-linked DNA, we believe that *PHB* likely carries an escape element in order to escape from XCI. As this gene is poorly expressed in human fibroblasts (data not shown), escape from XCI cannot simply be due to a strong promoter. We also observed a domain protected from DNA hypermethylation in *PITX2* between CpG islands 18 and 46, which span approximately 6 kb, while DNA hypomethylation on the *NGFR* MaxiP was evident at both *PHB* and *HPRT* promoters located over 10 kb apart. Interestingly, it appears that levels of DNA methylation can be modulated without impacting the XCI status of the transgenes. Despite lower DNA methylation and proximity to an escapee, *HPRT/Hprt* remains subject to XCI (Figure 3.7). This suggests that 41% DNA methylation on the Xi was sufficient to keep *HPRT/Hprt* in the inactivated state, while *PHB* showed expression on the Xi with ~15% DNA methylation, albeit not at the full expression level of the Xa. In addition, since *HPRT* DNA methylation was only examined in female mice that showed evidence of *PHB* escaping XCI, determining the DNA methylation of *HPRT* in female mice that showed no *PHB* expression on the Xi (Figure 3.7) will give more insights on whether the lowered *HPRT* DNA methylation with the *NGFR* BAC transgene is a result of being in close proximity to an escapee or a *cis*-regulatory element that independently promotes a more open chromatin state. At the endogenous location, there are at least three predicted insulators between *PHB* and the neighbouring gene approximately 50 kb downstream (162). It would be interesting to see whether *HPRT* DNA methylation is restored to the expected level on Xi if the insulators were included in the construct.

Overall, MaxiP constructs containing human BACs of up to 195 kb offered us an opportunity to examine whether XCI can spread across over 100 kb of human DNA on a mouse X chromosome. Our analysis of over 1.5 Mb of DNA identified only one of the 47 genes examined that escaped from XCI, which we propose reflects the existence of a *cis*-acting escape element in order to escape from XCI in an inactivated region on the X chromosome. Our study suggests that escape elements are not commonly found at the promoters on autosomes, consistent with the hypothesis that more genes escape from XCI in X;autosome translocations (~30% of autosomal genes, reviewed in 122) due to a depletion of waystations on the autosomes. Silencing of the MaxiPs on the Xi suggests that the autosomal BACs may contain waystations

that can propagate the inactivating signals during XCI or that the waystations native to the X chromosome could act over long distance across the MaxiPs. In addition, the most promising candidate waystation to date, LINE-1, whether full-length or truncated, likely does not serve as a dominant waystation, as the presence of LINE-1 alone was neither required nor sufficient to cause the inactivation of genes. It may be that waystations are yet to be identified, or perhaps, 'dominant waystations' simply do not exist and that waystations are in fact multiple factors that work synergistically to spread the inactivating signals of XCI. In the future, targeting smaller fraction of the *NGFR* BAC and certainly the X-linked human genes known to escape from XCI to the mouse *Hprt* locus will not only narrow down the identification of the different *cis*-regulatory elements important in XCI, but also decipher the nature of how waystations and escape elements function in humans and mice.

References

1. Ross, M.T., *et al.*, *The DNA sequence of the human X chromosome*. Nature, 2005. **434**(7031): p. 325-37.
2. Graves, J.A., *Sex chromosome specialization and degeneration in mammals*. Cell, 2006. **124**(5): p. 901-14.
3. Lyon, M.F., *Gene action in the X-chromosome of the mouse (Mus musculus L.)*. Nature, 1961. **190**: p. 372-3.
4. Deng, X., *et al.*, *Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster*. Nat Genet, 2011. **43**(12): p. 1179-85.
5. Brown, C.J., *et al.*, *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome*. Nature, 1991. **349**(6304): p. 38-44.
6. Brown, C.J., *et al.*, *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus*. Cell, 1992. **71**(3): p. 527-42.
7. Brockdorff, N., *et al.*, *Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome*. Nature, 1991. **351**(6324): p. 329-31.
8. Brockdorff, N., *et al.*, *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus*. Cell, 1992. **71**(3): p. 515-26.
9. Clemson, C.M., J.A. McNeil, H.F. Willard, and J.B. Lawrence, *XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure*. J Cell Biol, 1996. **132**(3): p. 259-75.
10. Lercher, M.J., A.O. Urrutia, and L.D. Hurst, *Clustering of housekeeping genes provides a unified model of gene order in the human genome*. Nat Genet, 2002. **31**(2): p. 180-3.
11. Singer, G.A., A.T. Lloyd, L.B. Huminiecki, and K.H. Wolfe, *Clusters of co-expressed genes in mammalian genomes are conserved by natural selection*. Mol Biol Evol, 2005. **22**(3): p. 767-75.
12. Gierman, H.J., *et al.*, *Domain-wide regulation of gene expression in the human genome*. Genome Res, 2007. **17**(9): p. 1286-95.
13. Guelen, L., *et al.*, *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions*. Nature, 2008. **453**(7197): p. 948-51.
14. Reddy, K.L., J.M. Zullo, E. Bertolino, and H. Singh, *Transcriptional repression mediated by repositioning of genes to the nuclear lamina*. Nature, 2008. **452**(7184): p. 243-7.
15. Bell, A.C., A.G. West, and G. Felsenfeld, *Insulators and boundaries: versatile regulatory elements in the eukaryotic genome*. Science, 2001. **291**(5503): p. 447-50.
16. Lister, R., *et al.*, *Human DNA methylomes at base resolution show widespread epigenomic differences*. Nature, 2009. **462**(7271): p. 315-22.
17. Holliday, R. and G.W. Grigg, *DNA methylation and mutation*. Mutat Res, 1993. **285**(1): p. 61-7.

18. Bird, A.P., *DNA methylation and the frequency of CpG in animal DNA*. Nucleic Acids Res, 1980. **8**(7): p. 1499-504.
19. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*. Nucleic Acids Res, 1982. **10**(8): p. 2709-21.
20. Rollins, R.A., et al., *Large-scale structure of genomic methylation patterns*. Genome Res, 2006. **16**(2): p. 157-63.
21. Kochanek, S., D. Renz, and W. Doerfler, *Transcriptional silencing of human Alu sequences and inhibition of protein binding in the box B regulatory elements by 5'-CG-3' methylation*. FEBS Lett, 1995. **360**(2): p. 115-20.
22. Woodcock, D.M., et al., *Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon*. J Biol Chem, 1997. **272**(12): p. 7810-6.
23. Hata, K. and Y. Sakaki, *Identification of critical CpG sites for repression of L1 transcription by DNA methylation*. Gene, 1997. **189**(2): p. 227-34.
24. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation*. Nat Genet, 1998. **20**(2): p. 116-7.
25. Kazazian, H.H., Jr., et al., *Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man*. Nature, 1988. **332**(6160): p. 164-6.
26. Han, J.S. and J.D. Boeke, *LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?* Bioessays, 2005. **27**(8): p. 775-84.
27. Xie, S., et al., *Cloning, expression and chromosome locations of the human DNMT3 gene family*. Gene, 1999. **236**(1): p. 87-95.
28. Okano, M., D.W. Bell, D.A. Haber, and E. Li, *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*. Cell, 1999. **99**(3): p. 247-57.
29. Hansen, R.S., et al., *The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome*. Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14412-7.
30. Lei, H., et al., *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells*. Development, 1996. **122**(10): p. 3195-205.
31. Yoder, J.A., N.S. Soman, G.L. Verdine, and T.H. Bestor, *DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe*. J Mol Biol, 1997. **270**(3): p. 385-95.
32. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes*. J Mol Biol, 1987. **196**(2): p. 261-82.
33. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
34. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
35. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome*. Nat Genet, 2007. **39**(4): p. 457-66.
36. Larsen, F., G. Gundersen, R. Lopez, and H. Prydz, *CpG islands as gene markers in the human genome*. Genomics, 1992. **13**(4): p. 1095-107.

37. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters*. Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1412-7.
38. Song, F., et al., *Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression*. Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3336-41.
39. Shen, L., et al., *Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters*. PLoS Genet, 2007. **3**(10): p. 2023-36.
40. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores*. Nat Genet, 2009. **41**(2): p. 178-86.
41. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters*. Nature, 2010. **466**(7303): p. 253-7.
42. Han, H., et al., *DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter*. Hum Mol Genet, 2011. **20**(22): p. 4299-310.
43. Jones, P.A., *The DNA methylation paradox*. Trends Genet, 1999. **15**(1): p. 34-7.
44. Ball, M.P., et al., *Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells*. Nat Biotechnol, 2009. **27**(4): p. 361-8.
45. Aran, D., G. Toperoff, M. Rosenberg, and A. Hellman, *Replication timing-related and gene body-specific methylation of active human genes*. Hum Mol Genet, 2010. **20**(4): p. 670-80.
46. Brenet, F., et al., *DNA methylation of the first exon is tightly linked to transcriptional silencing*. PLoS One, 2011. **6**(1): p. e14524.
47. Appanah, R., et al., *An unmethylated 3' promoter-proximal region is required for efficient transcription initiation*. PLoS Genet, 2007. **3**(2): p. e27.
48. Lorincz, M.C., D.R. Dickerson, M. Schmitt, and M. Groudine, *Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells*. Nat Struct Mol Biol, 2004. **11**(11): p. 1068-75.
49. Wolf, S.F., et al., *Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation*. Proc Natl Acad Sci U S A, 1984. **81**(9): p. 2806-10.
50. Lock, L.F., D.W. Melton, C.T. Caskey, and G.R. Martin, *Methylation of the mouse Hprt gene differs on the active and inactive X chromosomes*. Mol Cell Biol, 1986. **6**(3): p. 914-24.
51. Singer-Sam, J., et al., *Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation*. Mol Cell Biol, 1990. **10**(9): p. 4987-9.
52. Keith, D.H., J. Singer-Sam, and A.D. Riggs, *Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase*. Mol Cell Biol, 1986. **6**(11): p. 4122-5.
53. Cotton, A.M., et al., *Inactive X chromosome-specific reduction in placental DNA methylation*. Hum Mol Genet, 2009. **18**(19): p. 3544-52.
54. Yasukochi, Y., et al., *X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils*. Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3704-9.

55. Lock, L.F., N. Takagi, and G.R. Martin, *Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation*. Cell, 1987. **48**(1): p. 39-46.
56. Cotton, A.M., et al., *Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation*. Hum Genet, 2011. **130**(2): p. 187-201.
57. Su, A.I., et al., *Large-scale analysis of the human and mouse transcriptomes*. Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4465-70.
58. Allen, R.C., et al., *Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation*. Am J Hum Genet, 1992. **51**(6): p. 1229-39.
59. Bernardino, J., et al., *DNA methylation of the X chromosomes of the human female: an in situ semi-quantitative analysis*. Chromosoma, 1996. **104**(7): p. 528-35.
60. Bernardino, J., M. Lombard, A. Niveleau, and B. Dutrillaux, *Common methylation characteristics of sex chromosomes in somatic and germ cells from mouse, lemur and human*. Chromosome Res, 2000. **8**(6): p. 513-25.
61. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells*. Nat Genet, 2005. **37**(8): p. 853-62.
62. Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome*. Science, 2007. **315**(5815): p. 1141-3.
63. Sharp, A.J., et al., *DNA methylation profiles of human active and inactive X chromosomes*. Genome Res, 2011. **21**(10): p. 1592-600.
64. de la Torre, J., A.T. Sumner, J. Gosalvez, and L. Stuppia, *The distribution of genes on human chromosomes as studied by in situ nick translation*. Genome, 1992. **35**(5): p. 890-4.
65. Pranter, G. and M. Ferraro, *Analysis of methylation and distribution of CpG sequences on human active and inactive X chromosomes by in situ nick translation*. Chromosoma, 1990. **99**(1): p. 18-23.
66. Ng, K., D. Pullirsch, M. Leeb, and A. Wutz, *Xist and the order of silencing*. EMBO Rep, 2007. **8**(1): p. 34-9.
67. Belyaev, N., A.M. Keohane, and B.M. Turner, *Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells*. Hum Genet, 1996. **97**(5): p. 573-8.
68. Boggs, B.A., et al., *Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes*. Nat Genet, 2002. **30**(1): p. 73-6.
69. Brinkman, A.B., et al., *Histone modification patterns associated with the human X chromosome*. EMBO Rep, 2006. **7**(6): p. 628-34.
70. Costanzi, C. and J.R. Pehrson, *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals*. Nature, 1998. **393**(6685): p. 599-601.
71. Willard, H.F. and S.A. Latt, *Analysis of deoxyribonucleic acid replication in human X chromosomes by fluorescence microscopy*. Am J Hum Genet, 1976. **28**(3): p. 213-27.
72. Gomez, M. and N. Brockdorff, *Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage*. Proc Natl Acad Sci U S A, 2004. **101**(18): p. 6923-8.

73. Hansen, R.S., T.K. Canfield, A.D. Fjeld, and S.M. Gartler, *Role of late replication timing in the silencing of X-linked genes*. Hum Mol Genet, 1996. **5**(9): p. 1345-53.
74. Morey, C. and P. Avner, *The demoiselle of X-inactivation: 50 years old and as trendy and mesmerising as ever*. PLoS Genet, 2011. **7**(7): p. e1002212.
75. Helbig, R. and F.O. Fackelmayer, *Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain*. Chromosoma, 2003. **112**(4): p. 173-82.
76. Fackelmayer, F.O., *A stable proteinaceous structure in the territory of inactive X chromosomes*. J Biol Chem, 2005. **280**(3): p. 1720-3.
77. Clemson, C.M., et al., *The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences*. Proc Natl Acad Sci U S A, 2006. **103**(20): p. 7688-93.
78. Chaumeil, J., P. Le Baccon, A. Wutz, and E. Heard, *A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced*. Genes Dev, 2006. **20**(16): p. 2223-37.
79. Takagi, N. and M. Sasaki, *Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse*. Nature, 1975. **256**(5519): p. 640-2.
80. Kay, G.F., S.C. Barton, M.A. Surani, and S. Rastan, *Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development*. Cell, 1994. **77**(5): p. 639-50.
81. Kay, G.F., et al., *Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation*. Cell, 1993. **72**(2): p. 171-82.
82. Daniels, R., et al., *XIST expression in human oocytes and preimplantation embryos*. Am J Hum Genet, 1997. **61**(1): p. 33-9.
83. Ray, P.F., R.M. Winston, and A.H. Handyside, *XIST expression from the maternal X chromosome in human male preimplantation embryos at the blastocyst stage*. Hum Mol Genet, 1997. **6**(8): p. 1323-7.
84. van den Berg, I.M., et al., *X chromosome inactivation is initiated in human preimplantation embryos*. Am J Hum Genet, 2009. **84**(6): p. 771-9.
85. Okamoto, I., et al., *Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development*. Nature, 2011. **472**(7343): p. 370-4.
86. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation*. Nature, 1996. **379**(6561): p. 131-7.
87. Marahrens, Y., et al., *Xist-deficient mice are defective in dosage compensation but not spermatogenesis*. Genes Dev, 1997. **11**(2): p. 156-66.
88. Brown, C.J. and H.F. Willard, *The human X-inactivation centre is not required for maintenance of X-chromosome inactivation*. Nature, 1994. **368**(6467): p. 154-6.
89. Minks, J., W.P. Robinson, and C.J. Brown, *A skewed view of X chromosome inactivation*. J Clin Invest, 2008. **118**(1): p. 20-3.
90. Plenge, R.M., I. Percec, J.H. Nadeau, and H.F. Willard, *Expression-based assay of an X-linked gene to examine effects of the X-controlling element (Xce) locus*. Mamm Genome, 2000. **11**(5): p. 405-8.

91. Chadwick, L.H., et al., *Genetic control of X chromosome inactivation in mice: definition of the Xce candidate interval*. Genetics, 2006. **173**(4): p. 2103-10.
92. Plenge, R.M., et al., *A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation*. Nat Genet, 1997. **17**(3): p. 353-6.
93. Tomkins, D.J., H.L. McDonald, S.A. Farrell, and C.J. Brown, *Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay*. Eur J Hum Genet, 2002. **10**(1): p. 44-51.
94. Pugacheva, E.M., et al., *Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation*. Hum Mol Genet, 2005. **14**(7): p. 953-65.
95. Bolduc, V., et al., *No evidence that skewing of X chromosome inactivation patterns is transmitted to offspring in humans*. J Clin Invest, 2008. **118**(1): p. 333-41.
96. Amos-Landgraf, J.M., et al., *X chromosome-inactivation patterns of 1,005 phenotypically unaffected females*. Am J Hum Genet, 2006. **79**(3): p. 493-9.
97. Sharp, A., D. Robinson, and P. Jacobs, *Age- and tissue-specific variation of X chromosome inactivation ratios in normal women*. Hum Genet, 2000. **107**(4): p. 343-9.
98. Hatakeyama, C., et al., *The dynamics of X-inactivation skewing as women age*. Clin Genet, 2004. **66**(4): p. 327-32.
99. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females*. Nature, 2005. **434**(7031): p. 400-4.
100. Johnston, C.M., et al., *Large-scale population study of human cell lines indicates that dosage compensation is virtually complete*. PLoS Genet, 2008. **4**(1): p. e9.
101. Yang, F., T. Babak, J. Shendure, and C.M. Disteche, *Global survey of escape from X inactivation by RNA-sequencing in mouse*. Genome Res, 2010. **20**(5): p. 614-22.
102. Dementyeva, E.V., et al., *Difference between random and imprinted X inactivation in common voles*. Chromosoma, 2010. **119**(5): p. 541-52.
103. Reinius, B., et al., *Female-biased expression of long non-coding RNAs in domains that escape X-inactivation in mouse*. BMC Genomics, 2010. **11**: p. 614.
104. Goodfellow, P.J., et al., *Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X, and the Y chromosome*. Proc Natl Acad Sci U S A, 1988. **85**(15): p. 5605-9.
105. Goto, Y. and H. Kimura, *Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene*. Nucleic Acids Res, 2009. **37**(22): p. 7416-28.
106. Valley, C.M., L.M. Pertz, B.S. Balakumaran, and H.F. Willard, *Chromosome-wide, allele-specific analysis of the histone code on the human X chromosome*. Hum Mol Genet, 2006. **15**(15): p. 2335-47.
107. Changolkar, L.N., et al., *Genome-wide distribution of macroH2A1 histone variants in mouse liver chromatin*. Mol Cell Biol, 2010. **30**(23): p. 5473-83.
108. Murakami, K., et al., *Identification of the chromatin regions coated by non-coding Xist RNA*. Cytogenet Genome Res, 2009. **125**(1): p. 19-25.
109. Gartler, S.M. and A.D. Riggs, *Mammalian X-chromosome inactivation*. Annu Rev Genet, 1983. **17**: p. 155-90.

110. Li, N. and L. Carrel, *Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus*. Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17055-60.
111. Tsuchiya, K.D., et al., *Comparative sequence and X-inactivation analyses of a domain of escape in human Xp11.2 and the conserved segment in mouse*. Genome Res, 2004. **14**(7): p. 1275-84.
112. Carrel, L., P.A. Hunt, and H.F. Willard, *Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene*. Hum Mol Genet, 1996. **5**(9): p. 1361-6.
113. Sheardown, S., D. Norris, A. Fisher, and N. Brockdorff, *The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation*. Hum Mol Genet, 1996. **5**(9): p. 1355-60.
114. Lingenfelter, P.A., et al., *Escape from X inactivation of Smcx is preceded by silencing during mouse development*. Nat Genet, 1998. **18**(3): p. 212-3.
115. Filippova, G.N., et al., *Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development*. Dev Cell, 2005. **8**(1): p. 31-42.
116. Ciavatta, D., S. Kalantry, T. Magnuson, and O. Smithies, *A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation*. Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9958-63.
117. Hall, L.L., et al., *Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin*. Hum Mol Genet, 2002. **11**(25): p. 3157-65.
118. Keohane, A.M., et al., *H4 acetylation, XIST RNA and replication timing are coincident and define x;autosome boundaries in two abnormal X chromosomes*. Hum Mol Genet, 1999. **8**(2): p. 377-83.
119. Sharp, A.J., et al., *Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations*. Hum Mol Genet, 2002. **11**(25): p. 3145-56.
120. White, W.M., H.F. Willard, D.L. Van Dyke, and D.J. Wolff, *The spreading of X inactivation into autosomal material of an X;autosome translocation: evidence for a difference between autosomal and X-chromosomal DNA*. Am J Hum Genet, 1998. **63**(1): p. 20-8.
121. Popova, B.C., et al., *Attenuated spread of X-inactivation in an X;autosome translocation*. Proc Natl Acad Sci U S A, 2006. **103**(20): p. 7706-11.
122. Yang, C., et al., *X-chromosome inactivation: molecular mechanisms from the human perspective*. Hum Genet, 2011. **130**(2): p. 175-85.
123. Lyon, M.F., *LINE-1 elements and X chromosome inactivation: a function for "junk" DNA?* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6248-9.
124. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome*. Nature, 2002. **420**(6915): p. 520-62.
125. Bailey, J.A., L. Carrel, A. Chakravarti, and E.E. Eichler, *Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis*. Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6634-9.
126. Chow, J.C., et al., *LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation*. Cell, 2010. **141**(6): p. 956-69.

127. Wang, Z., H.F. Willard, S. Mukherjee, and T.S. Furey, *Evidence of influence of genomic DNA sequence on human X chromosome inactivation*. PLoS Comput Biol, 2006. **2**(9): p. e113.
128. Carrel, L., *et al.*, *Genomic environment predicts expression patterns on the human inactive X chromosome*. PLoS Genet, 2006. **2**(9): p. e151.
129. McNeil, J.A., K.P. Smith, L.L. Hall, and J.B. Lawrence, *Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]_n in the X escape region*. Genome Res, 2006. **16**(4): p. 477-84.
130. Nguyen, D.K., *et al.*, *Clcn4-2 genomic structure differs between the X locus in Mus spretus and the autosomal locus in Mus musculus: AT motif enrichment on the X*. Genome Res, 2011. **21**(3): p. 402-9.
131. Goldman, M.A., *et al.*, *Comparative methylation analysis of murine transgenes that undergo or escape X-chromosome inactivation*. Chromosome Res, 1998. **6**(5): p. 397-404.
132. Krumlauf, R., *et al.*, *Differential expression of alpha-fetoprotein genes on the inactive X chromosome in extraembryonic and somatic tissues of a transgenic mouse line*. Nature, 1986. **319**(6050): p. 224-6.
133. Heaney, J.D., A.N. Rettew, and S.K. Bronson, *Tissue-specific expression of a BAC transgene targeted to the Hprt locus in mouse embryonic stem cells*. Genomics, 2004. **83**(6): p. 1072-82.
134. Guillot, P.V., *et al.*, *Targeting of human eNOS promoter to the Hprt locus of mice leads to tissue-restricted transgene expression*. Physiol Genomics, 2000. **2**(2): p. 77-83.
135. Bronson, S.K., *et al.*, *Single-copy transgenic mice with chosen-site integration*. Proc Natl Acad Sci U S A, 1996. **93**(17): p. 9067-72.
136. Cvetkovic, B., B. Yang, R.A. Williamson, and C.D. Sigmund, *Appropriate tissue- and cell-specific expression of a single copy human angiotensinogen transgene specifically targeted upstream of the HPRT locus by homologous recombination*. J Biol Chem, 2000. **275**(2): p. 1073-8.
137. Evans, V., *et al.*, *Targeting the Hprt locus in mice reveals differential regulation of Tie2 gene expression in the endothelium*. Physiol Genomics, 2000. **2**(2): p. 67-75.
138. Farivar, S., S. Yamaguchi, M. Sugimoto, and N. Takagi, *X-chromosome inactivation in differentiating mouse embryonic stem cells carrying X-linked GFP and lacZ transgenes*. Int J Dev Biol, 2004. **48**(7): p. 629-35.
139. Collick, A., W. Reik, S.C. Barton, and A.H. Surani, *CpG methylation of an X-linked transgene is determined by somatic events postfertilization and not germline imprinting*. Development, 1988. **104**(2): p. 235-44.
140. Chong, S., J. Kontaraki, C. Bonifer, and A.D. Riggs, *A Functional chromatin domain does not resist X chromosome inactivation: silencing of cLys correlates with methylation of a dual promoter-replication origin*. Mol Cell Biol, 2002. **22**(13): p. 4667-76.
141. Magness, S.T., *et al.*, *In vivo pattern of lipopolysaccharide and anti-CD3-induced NF-kappa B activation using a novel gene-targeted enhanced GFP reporter gene mouse*. J Immunol, 2004. **173**(3): p. 1561-70.
142. Goldman, M.A., *et al.*, *A chicken transferrin gene in transgenic mice escapes X-chromosome inactivation*. Science, 1987. **236**(4801): p. 593-5.

143. Wu, H., *et al.*, *An X-linked human collagen transgene escapes X inactivation in a subset of cells*. *Development*, 1992. **116**(3): p. 687-95.
144. Lopes, A.M., *et al.*, *Transcriptional changes in response to X chromosome dosage in the mouse: implications for X inactivation and the molecular basis of Turner Syndrome*. *BMC Genomics*, 2010. **11**: p. 82.
145. Splinter, E., *et al.*, *The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA*. *Genes Dev*, 2011. **25**(13): p. 1371-83.
146. Al Nadaf, S., *et al.*, *A cross-species comparison of escape from X inactivation in Eutheria: implications for evolution of X chromosome inactivation*. *Chromosoma*, 2012. **121**(1): p. 71-8.
147. Chureau, C., *et al.*, *Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region*. *Hum Mol Genet*, 2011. **20**(4): p. 705-18.
148. Ehrmann, I.E., *et al.*, *Characterization of genes encoding translation initiation factor eIF-2gamma in mouse and human: sex chromosome localization, escape from X-inactivation and evolution*. *Hum Mol Genet*, 1998. **7**(11): p. 1725-37.
149. Johnston, C.M., A.E. Newall, N. Brockdorff, and T.B. Nesterova, *Enox, a novel gene that maps 10 kb upstream of Xist and partially escapes X inactivation*. *Genomics*, 2002. **80**(2): p. 236-44.
150. Tian, D., S. Sun, and J.T. Lee, *The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation*. *Cell*, 2010. **143**(3): p. 390-403.
151. Yen, Z.C., I.M. Meyer, S. Karalic, and C.J. Brown, *A cross-species comparison of X-chromosome inactivation in Eutheria*. *Genomics*, 2007. **90**(4): p. 453-63.
152. Basrur, P.K., *et al.*, *Expression pattern of X-linked genes in sex chromosome aneuploid bovine cells*. *Chromosome Res*, 2004. **12**(3): p. 263-73.
153. Nino-Soto, M.I., *et al.*, *Differences in the pattern of X-linked gene expression between fetal bovine muscle and fibroblast cultures derived from the same muscle biopsies*. *Cytogenet Genome Res*, 2005. **111**(1): p. 57-64.
154. Greenfield, A., *et al.*, *The UTX gene escapes X inactivation in mice and humans*. *Hum Mol Genet*, 1998. **7**(4): p. 737-42.
155. Keitges, E., M. Rivest, M. Siniscalco, and S.M. Gartler, *X-linkage of steroid sulphatase in the mouse is evidence for a functional Y-linked allele*. *Nature*, 1985. **315**(6016): p. 226-7.
156. Gartler, S.M. and M. Rivest, *Evidence for X-linkage of steroid sulfatase in the mouse: steroid sulfatase levels in oocytes of XX and XO mice*. *Genetics*, 1983. **103**(1): p. 137-41.
157. Wiberg, U.H. and K. Fredga, *Steroid sulphatase levels are higher in males than in females of the root vole (Microtus oeconomus). Yet another rodent with an active Y-linked allele?* *Hum Genet*, 1987. **77**(1): p. 6-11.
158. Carrel, L., *et al.*, *X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse*. *Hum Mol Genet*, 1996. **5**(3): p. 391-401.
159. Yang, G.S., *et al.*, *Next generation tools for high-throughput promoter and expression analysis employing single-copy knock-ins at the Hprt1 locus*. *Genomics*, 2009. **93**(3): p. 196-204.
160. Portales-Casamar, E., *et al.*, *A regulatory toolbox of MiniPromoters to drive selective expression in the brain*. *Proc Natl Acad Sci U S A*, 2010. **107**(38): p. 16589-94.

161. Csankovszki, G., *et al.*, *Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation*. Nat Genet, 1999. **22**(4): p. 323-4.
162. Ernst, J., *et al.*, *Mapping and analysis of chromatin state dynamics in nine human cell types*. Nature, 2011. **473**(7345): p. 43-9.
163. Goecks, J., A. Nekrutenko, and J. Taylor, *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences*. Genome Biol, 2010. **11**(8): p. R86.
164. Blankenberg, D., *et al.*, *Galaxy: a web-based genome analysis tool for experimentalists*. Curr Protoc Mol Biol, 2010. **Chapter 19**: p. Unit 19 10 1-21.
165. Giardine, B., *et al.*, *Galaxy: a platform for interactive large-scale genome analysis*. Genome Res, 2005. **15**(10): p. 1451-5.
166. Penzkofer, T., T. Dandekar, and T. Zemojtel, *L1Base: from functional annotation to prediction of active LINE-1 elements*. Nucleic Acids Res, 2005. **33**(Database issue): p. D498-500.
167. Iglewicz, B. and D. Hoaglin, *How to Detect and Handle Outliers*. ASQC basic references in quality control. Vol. 16. 1993, Milwaukee: ASQC Quality Press.
168. Soriano, P., *Generalized lacZ expression with the ROSA26 Cre reporter strain*. Nat Genet, 1999. **21**(1): p. 70-1.
169. Dreszer, T.R., *et al.*, *The UCSC Genome Browser database: extensions and updates 2011*. Nucleic Acids Res, 2012. **40**(1): p. D918-23.
170. Straussman, R., *et al.*, *Developmental programming of CpG island methylation profiles in the human genome*. Nat Struct Mol Biol, 2009. **16**(5): p. 564-71.
171. Thomson, J.P., *et al.*, *CpG islands influence chromatin structure via the CpG-binding protein Cfp1*. Nature, 2010. **464**(7291): p. 1082-6.