

Targeting of >1.5 Mb of Human DNA into the Mouse X Chromosome Reveals Presence of *cis*-Acting Regulators of Epigenetic Silencing

Christine Yang,* Andrea J. McLeod,[†] Allison M. Cotton,* Charles N. de Leeuw,[†] Stéphanie Laprise,[†] Kathleen G. Banks,[†] Elizabeth M. Simpson,^{†,*} and Carolyn J. Brown*¹

*Department of Medical Genetics, Molecular Epigenetics Group, Life Sciences Institute, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada, and [†]Centre for Molecular Medicine and Therapeutics at the Child and Family Research Institute and

[‡]Department of Medical Genetics, Department of Psychiatry, University of British Columbia, Vancouver, BC, V5Z 4H4, Canada

ABSTRACT Regulatory sequences can influence the expression of flanking genes over long distances, and X chromosome inactivation is a classic example of *cis*-acting epigenetic gene regulation. Knock-ins directed to the *Mus musculus Hprt* locus offer a unique opportunity to analyze the spread of silencing into different human DNA sequences in the identical genomic environment. X chromosome inactivation of four knock-in constructs, including bacterial artificial chromosome (BAC) integrations of over 195 kb, was demonstrated by both the lack of expression from the inactive X chromosome in females with nonrandom X chromosome inactivation and promoter DNA methylation of the human transgene in females. We further utilized promoter DNA methylation to assess the inactivation status of 74 human reporter constructs comprising >1.5 Mb of DNA. 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 demonstration of expression from the inactive X chromosome. Integration of *PHB* resulted in lower DNA methylation of the flanking *HPRT* promoter in females, suggesting the action of a dominant *cis*-acting escape element. Female-specific DNA hypermethylation of CpG islands not associated with promoters implies a widespread imposition of DNA methylation during X chromosome inactivation; yet transgenes demonstrated differential capacities to accumulate DNA methylation when integrated into the identical location on the inactive X chromosome, suggesting additional *cis*-acting sequence effects. As only one of the human transgenes analyzed escaped X chromosome inactivation, we conclude that elements permitting ongoing expression from the inactive X are rare in the human genome.

X chromosome inactivation (XCI) transcriptionally silences an X chromosome early in mammalian development, thereby serving as a means of dosage compensation between the sexes (reviewed in Migeon 2011). XCI is a classic paradigm of epigenetic silencing, but while many of the molecular players in the inactivation process have been elucidated (Wutz 2011), less is known about the *cis*-acting DNA elements involved in the spread of silencing. A substantial number of X-linked genes continue to be expressed from the inactive X (Xi) in humans, and the clustered nature of

these escapees in humans suggests that escape from XCI may be regulated in domains (Carrel and Willard 2005). Intriguingly, the number of escapees in mouse is considerably more limited, with only ~4% of genes escaping XCI (~20 transcripts) in mice compared to ~15% (~94 transcripts) of analyzed human genes (Carrel and Willard 2005; Lopes *et al.* 2010; Yang *et al.* 2010; Splinter *et al.* 2011). The ongoing escape from inactivation of the escapee *Kdm5c* when integrated at four different locations on the mouse X chromosome, while its flanking genes maintain their inactive state, provides strong evidence for intrinsic regulatory sequences, which we refer to as escape elements, that allow genes to escape from XCI (Li and Carrel 2008).

Gartler and Riggs (1983) proposed the existence of waystations or booster elements that promote the spread of inactivation, based on the limited inactivation of autosomal regions of X; autosome translocations, and Lyon proposed

Copyright © 2012 by the Genetics Society of America

doi: 10.1534/genetics.112.143743

Manuscript received July 17, 2012; accepted for publication September 17, 2012

Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/content/suppl/2012/09/28/genetics.112.143743.DC1>.

¹Corresponding author: Department of Medical Genetics, 2350 Health Sciences Mall, Vancouver, BC, V6T 1Z3, Canada. E-mail: carolyn.brown@ubc.ca

that long interspersed elements (LINEs) that are enriched on the X chromosome could be such waystations (Waterston *et al.* 2002; Ross *et al.* 2005; Lyon 2006). Computational studies comparing the genomic neighborhoods of genes with different inactivation status have supported that inactivated and escape genes have different genomic environments with respect to the content of repetitive sequences (Bailey *et al.* 2000; Carrel *et al.* 2006; Wang *et al.* 2006; Nguyen *et al.* 2011). Several mouse escapees also escape from XCI in humans, and the difference in inactivation of flanking genes has been attributed to the presence or loss of the CTCF boundary element (Filippova *et al.* 2005; Goto and Kimura 2009). A direct comparison of the inactivation status of sequences integrated into the same genomic location would permit a direct analysis of the impact of *cis*-acting sequences on XCI and potentially identify sequences containing escape elements, waystations, and/or boundary elements.

Most transgenes on the X chromosome are subject to XCI with the only transgenes shown to consistently escape XCI being a chicken transferrin transgene and the mouse *Kdm5c* bacterial artificial chromosome (BAC) transgene (Goldman *et al.* 1987, 1998; Li and Carrel 2008), although a human collagen and an NF- κ B-dependent EGFP reporter gene may also at least partially escape from XCI (Wu *et al.* 1992; Magness *et al.* 2004). Lack of knowledge of the integration site for the chicken transferrin transgene confounds the identification of *cis*-acting regulatory elements. Targeted single-copy integrations into a single location would provide a consistent resource to examine the spread of XCI into sequences not normally X linked, and *Hprt* has been described as a permissive locus with relatively minimal effect on transgene expression (Bronson *et al.* 1996; Cvetkovic *et al.* 2000). To assess whether a gene is subject to XCI many studies have examined allelic expression in females with nonrandom inactivation (*e.g.*, Carrel and Willard 2005; Yang *et al.* 2010); however, an indirect means of assessing XCI is to study DNA methylation, which can be readily examined on banked DNA samples. CpG island promoters of X-linked genes that are subject to XCI on the Xi are typically DNA hypermethylated, while CpG island promoters are unmethylated on the active X (Xa), as on the autosomes (Wu *et al.* 1992; Chong *et al.* 2002; Matarazzo *et al.* 2002; Cotton *et al.* 2009; Yasukochi *et al.* 2010). Genes and transgenes that escape XCI demonstrate low levels of promoter DNA methylation on both the Xa and the Xi (Goodfellow *et al.* 1988; Goldman *et al.* 1998), resulting in overall low DNA methylation levels in both males and females, allowing DNA methylation to be used to assess XCI status of genes (Cotton *et al.* 2009; Yasukochi *et al.* 2010).

Integration of transgenes at a single locus on the X chromosome allows for detailed examination of *cis*-acting DNA elements and we herein report the analysis of 74 human autosomal and X-linked transgenes knocked into the mouse *Hprt* locus that were generated from the Pleiades Promoter Project (Yang *et al.* 2009; Portales-Casamar *et al.* 2010; J.-F. Schmouth and E. M. Simpson, unpublished data),

including eight human autosomal BACs of >195 kb integrated using the HuGX method [high-throughput human genes on the X chromosome (Schmouth *et al.* 2012)]. Overall there was a significant difference in DNA methylation between males and females, and most human promoters integrated into the X chromosome showed DNA hypermethylation in females, suggesting they were subject to XCI. Silencing from the Xi was verified by expression analysis of four BAC-derived transgenes present on the Xi in females that had nonrandom XCI. We identified one transgene that escaped from XCI and demonstrated that different constructs had differential capacity for DNA methylation when on the Xi, suggesting the presence of additional *cis*-acting epigenetic modulators.

Materials and Methods

Pleiades Promoter Project constructs

The Pleiades Promoter Project (Yang *et al.* 2009; Portales-Casamar *et al.* 2010) was an international 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 strains were made using homologous recombination at the *Hprt^{b-m3}* deletion locus on the mouse X chromosome; integration of the Pleiades constructs generated a chimeric *HPRT/Hprt* locus containing the human *HPRT* promoter that rescued the deletion (Bronson *et al.* 1996). MiniPromoters (MiniPs) were ≤ 4 kb in size and were composed of different combinations of small putative regulatory elements (<http://pleiades.org/>). In contrast, MaxiPromoters (MaxiPs) were human BAC-derived constructs that ranged from 100 to 195 kb of human genomic DNA, with a reporter inserted at the start codon of the gene of interest (J.-F. Schmouth and E. M. Simpson, unpublished data). For the MiniP constructs, the reporter *lacZ* or EGFP (or EGFP/cre) is ~ 200 bp or 50 bp downstream of the MiniP, respectively. Thirty-seven of 57 target genes contained a promoter CpG island, which was generally truncated in MiniPs compared to the endogenous islands (Supporting Information, Table S1). Approval for the generation and breeding of mice carrying the Pleiades constructs was obtained from the University of British Columbia Committee on Animal Care.

Generation of mouse strains

The floxed *Xist* strain 129-*Xist^{tm2Jae}* [stock no. 029172-UNC (Csankovszki *et al.* 1999)] was obtained heterozygous from the Mutant Mouse Regional Resource Center, and the cre-deleter strain FVB/N-Tg(ACTB-cre)2Mrt [stock no. 003376 (Lewandoski *et al.* 1997)] was obtained heterozygous from The Jackson Laboratory (JAX). Both strains were maintained by backcrossing to strain 129S1/SvImJ [stock no. 002448 (Simpson *et al.* 1997)] obtained from JAX. The floxed *Xist* strain was crossed to the cre-deleter strain at backcross generations JAX-plus N2 and N7, respectively, to

generate females carrying the *Xist* deletion (129-XX*ist*^{1lox}/X). Females with the *Xist* deletion were then crossed to males with the Pleiades construct integrated at the *Hprt* locus (B6-X^{MaxiP}/Y) to generate 129B6F1-XX*ist*^{1lox}/X^{MaxiP} females. This *Xist* knockout has been shown to render the X chromosome carrying it unable to inactivate (Gribnau *et al.* 2005), thus resulting in the MaxiP knock-in X chromosome with an intact *Xist* becoming the Xi. Complete nonrandom XCI was verified by examining the relative expression levels of the B6 and 129 alleles at a single-nucleotide polymorphism of the *Fln* locus (data not shown).

The B6.129S4-*Gt(ROSA)26Sor*^{tm1Sor}/J strain [stock no. 003474 (Soriano 1999)] was obtained homozygous from JAX at N5, maintained by backcrossing to strain C57BL/6J (stock no. 000664) obtained from JAX, and then made homozygous again at N10. The allele was introduced into the Pleiades strains at N5–N10 (Portales-Casamar *et al.* 2010).

DNA and RNA extraction from tissues

An ear notch of ~1 mm in diameter was taken from each mouse postweaning (~4 weeks old) and digested with 200 μ l of mouse homogenization buffer [50 mM KCL, 10 mM Tris-HCL, pH 8.3, 2 mM MgCl₂, 0.1 mg/ml gelatin, 0.45% IGEPAL CA-630 (Sigma-Aldrich), 0.45% Tween 20 (Sigma-Aldrich), and 24 μ g of Proteinase K (Sigma-Aldrich)] overnight at 55°. The digested samples were then heat inactivated at 95° for 10 min. DNA and RNA extractions from 50–100 mg of mouse livers and brains were done using TRIzol Reagent (Invitrogen, Carlsbad, CA), according to the manufacturer's protocol.

Expression analysis

Expression of constructs was assessed by staining of *lacZ* with X-gal as previously described (Portales-Casamar *et al.* 2010). *lacZ* staining was performed on a limited number of ear notch samples (Table S1 and Table S2) and on the brains of female mice carrying the *Xist* deletion. The images of *lacZ* staining of the mouse brain sections (Figure 1) were adjusted for brightness and contrast using Photoshop, but all images for the same construct (test and control samples) received the same adjustments. For analysis of transcription, ~2 μ g of RNA extracted from tissues was converted to cDNA with standard reverse transcription conditions, using M-MLV (Invitrogen) at 42° for 2 hr 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 sec, 60° for 30 sec, and 72° for 1 min; and a melt curve stage of 95° for 15 sec, 60° for 1 min, and an increase of 0.3° until 95°. Serial dilutions of genomic DNA from an *NGFR* female (without the *Xist* deletion) were used as the

standards to which each sample cDNA was compared, to generate a relative quantity of the *PHB*, *HPRT*, *Pgk1*, and intergenic transcription between *PHB* and *HPRT*. Expression levels were normalized to *Pgk1* expression level, and quantifications were done in triplicate, with any outlier excluded from the analysis. Primer sequences and distances from the assay to the transcription start site are found in Table S3.

DNA methylation analysis

Using the EZ DNA Methylation-Gold Kit or the EZ-96 DNA Methylation-Gold Kit (Zymo Research), 500 ng of DNA obtained from the lysed ear notches and a limited number of liver and brain samples were bisulfite converted, following the manufacturer's instructions. Internal bisulfite conversion controls were included in the pyrosequencing assays to monitor complete conversion of DNA. Each 25- μ l pyrosequencing PCR was performed with 1 \times PCR buffer (QIAGEN, Valencia, CA), 0.2 mM dNTPs, 0.625 unit Hot Start Taq DNA polymerase (QIAGEN), 0.25 μ M forward primer, 0.25 μ M reverse primer, and 12–35 ng bisulfite-converted DNA. Assays for *CCKBR*, *ICMT*, *NOV*, and *NR2E1* were performed with 0.5 μ M forward and reverse primers. Conditions for PCR were 95° for 15 min, 50 cycles of 94° for 30 sec, annealing temperature for 30 sec (see Table S3), 72° for 1 min, and finally 72° for 10 min. One forward or reverse primer was biotinylated, depending on which strand contained the target region to be sequenced, to subsequently isolate the strand of interest for pyrosequencing. Template preparation for pyrosequencing was done according to the manufacturer's protocol, using 10–15 μ l 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 DNA, which we attributed to degradation of ear notch DNA that was stored at 4° for up to 3 years. All promoter assays were replicated 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 deviations of 5%, 3%, and 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.

The University of California, Santa Cruz (UCSC) definition and designation of a CpG island (GC content of at least 50%, length >200 bp, observed_{CpG}/expected_{CpG} ratio >0.6) were used (Gardiner-Garden and Frommer 1987). The promoter of *PHB* was classified as a CpG island with intermediate CpG density, which was defined as having a GC content >50%, an observed_{CpG}/expected_{CpG} >0.48, length at least 200 bp, and no overlap with the UCSC CpG islands (Weber *et al.* 2007); the *PHB* promoter CpG island was located using the CpGIE

program (Wang and Leung 2004). Primers were designed using PSQ Assay Design software (QIAGEN). At least three CpGs were analyzed for each CpG island examined, and the distance of the analyzed CpGs from the transcription start sites can be found in Table S3.

Promoter identification of *PITX2* (Ple158) was based on the ENCODE chromatin states track downloaded from the UCSC genome browser (Ernst *et al.* 2011). Repeat masker (<http://www.repeatmasker.org>, Institute for Systems Biology) and Galaxy (Giardine *et al.* 2005; Blankenberg *et al.* 2010; Goecks *et al.* 2010) were used to determine the base pair composition of LINEs and Alu in the MaxiP constructs.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 5.02. An α -value of 0.05 was used for testing significance. A Mann–Whitney *t*-test was used to test for significant differences in DNA methylation levels between male and female mice. Mouse strains with modified Z-scores >3.5 in absolute value were considered outliers (Iglewicz and Hoaglin 1993). Spearman's correlation was performed to examine the relationship between DNA methylation levels of neighboring genes.

Results

DNA methylation reflects XCI of Pleiades constructs

The Pleiades Promoter Project generated reporter constructs with human promoters and targeted them to the *Hprt* locus on the mouse X chromosome by homologous recombination (Yang *et al.* 2009; Portales-Casamar *et al.* 2010; J.-F. Schmouth and E. M. Simpson, unpublished data). Integration of the Pleiades promoter constructs created a chimeric *HPRT/Hprt* locus that consisted of the human *HPRT* promoter and exon 1 and mouse *Hprt* exons 2–9 (Figure 1A). While MiniP constructs contained a human promoter of ≤ 4 kb in size that drives a reporter (*lacZ*, EGFP, or EGFP/cre), MaxiP constructs were derived from human BACs of up to 195 kb with the reporter (*lacZ* or EGFP) inserted at the translation start codon of the gene of interest on the human BAC. To predict whether the constructs were subject to the *cis*-regulation of XCI now that they were X linked, we generated female mice carrying a deletion at the *Xist* gene [*Xist*^{1lox} (Csankovszki *et al.* 1999)] on the X chromosome without the knock-in, thereby causing the Pleiades knock-in to always be on the Xi. The MaxiP constructs *AMOTL1* (Ple5), *MAOA* (Ple127), *NR2E1* (Ple142), and *NR2F2* (Ple143) were not expressed in the brains of females heterozygous for the *Xist* deletion, but were expressed in various parts of the brain in females without the *Xist* deletion (Figure 1B), indicating that these MaxiP constructs are expressed only when present on the Xa and are thus subject to XCI. DNA methylation analysis on DNA from ear notch samples of hemizygous male and heterozygous female mice transgenic for *AMOTL1* and *NR2E1* showed that CpG island promoters on the BACs were signif-

icantly DNA hypermethylated in females compared to males of the same strain ($P < 0.05$; Figure 1C), in agreement with the XCI statuses assessed by *lacZ* expression in the females with nonrandom XCI. DNA methylation was also examined in brain and/or liver samples for a few constructs, including the *NR2E1* BAC, and similar levels of DNA methylation to those in ear notches were observed (Table S2). We conclude that DNA methylation at CpG island promoters is a reliable predictor of XCI status for transgenes at *Hprt*. DNA methylation was therefore used to determine the XCI status of the remaining MaxiP constructs (*NOV*, *NGFR*, *PITX2*, and *LCT*). *PITX2* (CpG island 46 designation on UCSC) showed less female DNA methylation than the other MaxiP constructs, which could reflect either the presence or the absence of *cis*-acting regulators of XCI or a tendency to be preferentially located on the Xa. To examine the latter possibility we looked at DNA methylation of the flanking *Phf6* and *HPRT* promoters.

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

The Pleiades construct and the human *HPRT* promoter are located on the same chromosome; therefore, if substantial skewing of XCI were present, their DNA methylation levels would be correlated, reflecting the proportion of cells in which they are both on the Xi. In contrast, *Phf6* DNA methylation should not be affected by skewing since it is present on both X chromosomes. The *HPRT* promoter CpG island was truncated in the chimeric gene, but the chimeric gene complemented the *Hprt* deletion and provided resistance to HAT selection. Both the *HPRT* and the *Phf6* 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%, $P < 0.0001$), suggesting that both neighboring 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*), consistent with variability in levels of skewing of XCI in the samples analyzed. A correlation between the DNA methylation levels at the human promoter and at *HPRT*, but not with *Phf6* was observed (Figure 2), supporting the presence of skewing of XCI in the analyzed ear notch samples. Intriguingly, different MaxiP constructs showed different slopes in the correlation of their DNA methylation level with *HPRT* (Figure 2B), suggesting that 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 levels in females homozygous for the knock-in and in females heterozygous for the *Xist* deletion who carry the knock-in solely on the Xi. The *AMOTL1*, *NOV*, and *NR2E1* MaxiP constructs showed similar levels of *HPRT* DNA methylation on the Xi ($\sim 70\%$) but slightly different levels of promoter DNA methylation on the Xi (Figure

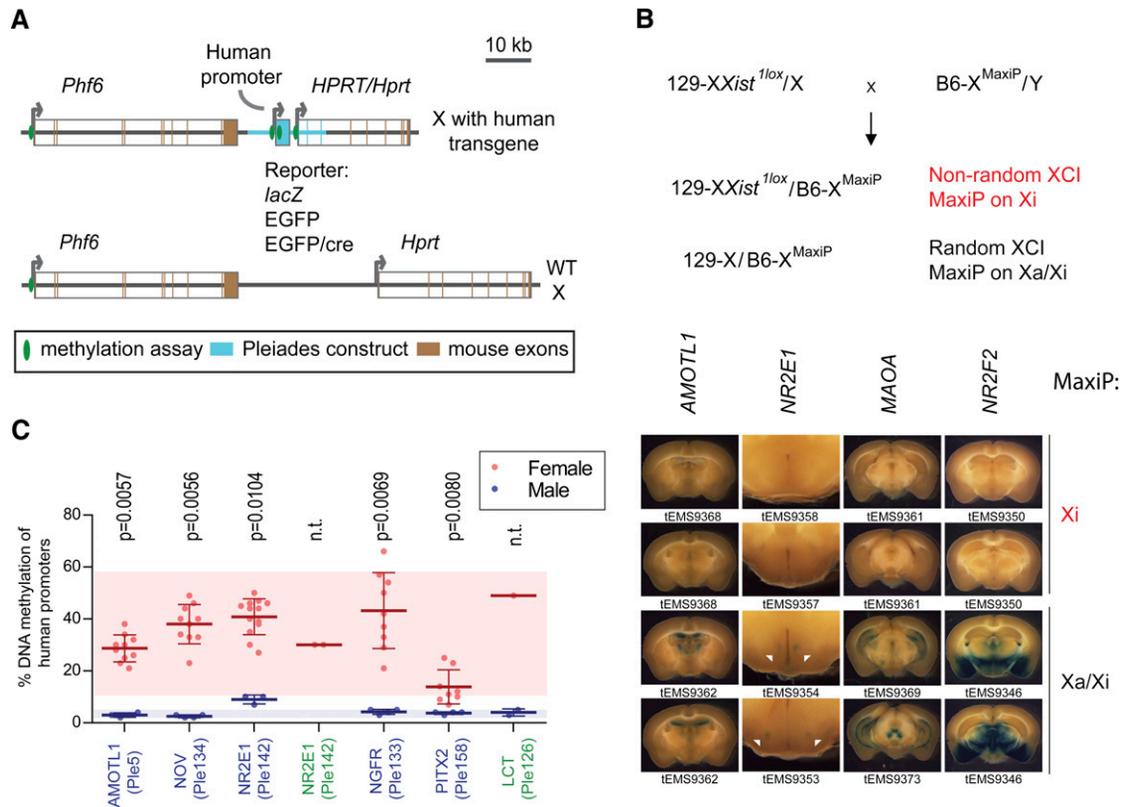


Figure 1 BAC-derived MaxiP constructs of 120–195 kb are subject to XCI and show DNA hypermethylation of CpG island promoters in females compared to males. (A) Experimental system in which human promoters driving a reporter (Pleiades constructs) were integrated at the *Hprt* locus on the mouse X chromosome, using homologous recombination. A chimeric *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, so the wild-type (WT) mouse locus is shown below the knock-in chromosome. The sizes of the Pleiades construct and the internal exons are not shown to scale. (B) Generation of mice carrying the Pleiades knock-in on the Xi. Top, the breeding scheme crossing *Xist*^{1lox} females with males carrying the Pleiades knock-in, to generate females carrying the *Xist* deletion and the knock-in on the Xi (Xi), and females with wild-type *Xist* and the knock-in, which could be on the Xa or Xi (Xa/Xi). Bottom, the *lacZ* staining in the brains of females with *AMOTL1*, *NR2E1*, *MAOA*, and *NR2F2*. Regions with *lacZ* staining in the brain sections of mice carrying the *NR2E1* transgene are indicated by the white arrowheads. Images labeled with the same tEMS number were obtained from the same mouse. (C) DNA hypermethylation of the MaxiP constructs predicts XCI status. Each construct is denoted with a Pleiades (Ple) number, along with the human gene from which the constructs originate. Constructs with *lacZ* and EGFP as the reporter are colored in blue and green, respectively. DNA methylation of *PITX2* was examined at CpG island 46 (UCSC). 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 a 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, ± 1 standard deviation between mice for the strain; shaded regions, 2 standard deviations from the average DNA methylation level.

2B). DNA methylation levels at *PITX2* and *NGFR* were strikingly different from those at the other MaxiP constructs.

PITX2 (CpG island 46) showed a much lower range of DNA methylation when compared to DNA methylation of *AMOTL1*, *NOV*, and *NR2E1*, and DNA hypermethylation of *HPRT* indicates that the low level of female DNA methylation at *PITX2* is not attributable to skewing of XCI but to its intrinsic resistance to accumulate DNA methylation (Figure 2B). In contrast, the *NGFR* MaxiP construct showed a lower *HPRT* DNA methylation range (13–33%) compared to the other MaxiP constructs, suggesting that the capacity of *HPRT* to accumulate DNA methylation is altered in this construct. We also 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 the *NGFR* BAC (Figure S1). 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 largest decrease in capacity to accumulate promoter DNA methylation and the *NGFR* BAC showed an impact on *HPRT* DNA methylation. To understand the *cis*-modulatory effects of the integrated DNA, we explored the *PITX2* and *NGFR* BACs in more detail.

***PITX2* is DNA hypermethylated at transcription start sites as well as intragenic and intergenic CpG islands in females**

CpG island 46 on the *PITX2* BAC is not annotated as the start of the *PITX2* transcript, so we analyzed the DNA methylation levels at eight additional locations on the BAC including

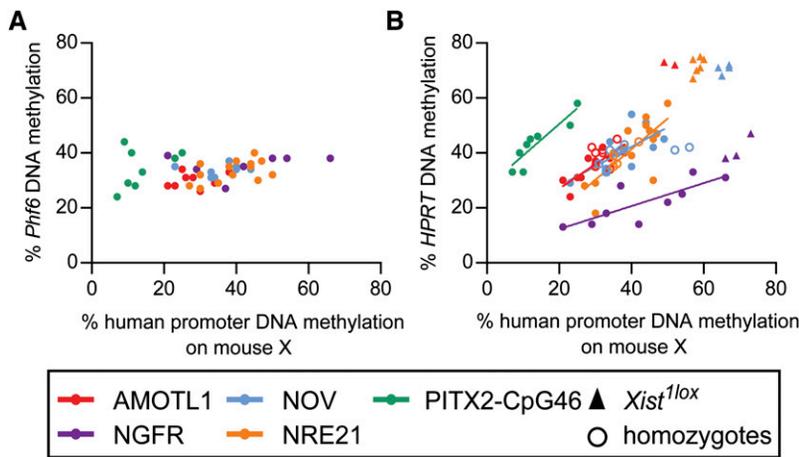


Figure 2 DNA methylation analysis of flanking *HPRT* and *Phf6* genes reveals skewing of XCI and differential susceptibility of MaxiP constructs to DNA methylation on the Xi. (A) Spearman's correlation analysis between *Phf6* and the MaxiP DNA methylation levels showed no significant correlations. *AMOTL1*, $r = 0.2532$, $P = 0.5206$; *NGFR*, $r = 0.1482$, $P = 0.7825$; *NOV*, $r = 0.2270$, $P = 0.5821$; *NR2E1*, $r = 0.4633$, $P = 0.0953$; *PITX2*-CpG46, $r = 0.2635$, $P = 0.5364$. (B) Spearman's 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*-CpG46, $r = 0.9581$, $P = 0.0011$. *Xist*^{1lox} females (triangles) carrying different MaxiP constructs on the Xi showed different levels of DNA methylation, for which homozygous mice showed intermediate DNA methylation (open circles). Each circle or triangle indicates DNA methylation levels from an individual mouse.

exon 1 and intron 2 (non-CpG island), three internal CpG islands, the promoter CpG island of the annotated alternative isoform, and two intergenic CpG islands (Figure 3). 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). 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. Although our DNA methylation assays are located in the gene body of *PITX2*, chromatin modifications associated with promoters were found to overlap the assays in intron 2 and CpG islands 46 and 196 (Figure 3) (Ernst *et al.* 2011), suggesting *PITX2* has additional internal promoters. Intergenic CpG islands 59 and 29 show no or very weak chromatin modifications associated with promoters or enhancers, yet both CpG islands showed female-specific DNA hypermethylation (Figure 3). Similarly, analysis of an intergenic and an intragenic CpG island on the *NR2E1* BAC demonstrated female-specific DNA hypermethylation (data not shown). Interestingly, *lacZ* showed a clear difference in DNA methylation levels between males and females (Figure 3), 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. Consistent with published data (Straussman *et al.* 2009), our assessment of male and female blood and lymphoblast lines suggests that in humans the promoter CpG island 196 is always unmethylated while other sites show variable DNA methylation that is not sex specific (data not shown), with the exception of CpG island 29, which appears to be DNA methylated in all tissues except sperm.

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

A distinguishing characteristic of the *NGFR* construct from the other MaxiP constructs was the presence of a truncated gene

at the end of the BAC that is adjacent to the *HPRT/Hprt* locus (Figure 4A). The *PHB* gene is truncated within the 3'-UTR ~200 bp from the end of the gene, and we hypothesized that *PHB* escaped from XCI and that the run-on transcription from *PHB* through the *HPRT/Hprt* locus positioned ~2.5 kb downstream could be the cause of the reduced *HPRT* DNA methylation on the Xi. 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 4B). By qPCR, we showed that *PHB* was not a highly expressed gene relative to *Pgk1*, but was expressed from the Xi in females heterozygous for the *Xist* deletion at levels up to 30% of the level of expression observed in males (Figure 4B), while females with random XCI showed a level of *PHB* expression close to 60% of that in males. Variability was observed in *PHB* expression levels from the Xi between females, perhaps reflecting the variable escape from XCI previously described for X-linked genes (Carrel and Willard 2005). However, while the expression level at ~1.4 kb downstream of the truncated *PHB* gene in the intergenic region was essentially the same as at the 3'-UTR of *PHB* (Figure 4B), this transcription had ceased by ~250 bp upstream of *HPRT/Hprt*, indicating that there is no substantial run-on transcription through the *HPRT/Hprt* locus. In addition, analysis of *HPRT* expression showed that *HPRT/Hprt* remained inactivated on the Xi despite its lower level of DNA methylation and proximity to a gene escaping from XCI. In agreement with the *PHB* expression analysis, the promoter of *PHB* has an island of intermediate CpG density (GC% = 52.9, observed/expected CpG = 0.57, length = 1823 bp) that showed relatively low DNA methylation in females with the *Xist* deletion, but the *PHB* DNA methylation level on the Xi was still distinct from the level of DNA methylation on the Xa in males (Figure 4C). 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.

MiniP constructs are generally subject to XCI

Since our MaxiP results agreed with previous reports that DNA methylation is an accurate marker for XCI status (Goldman

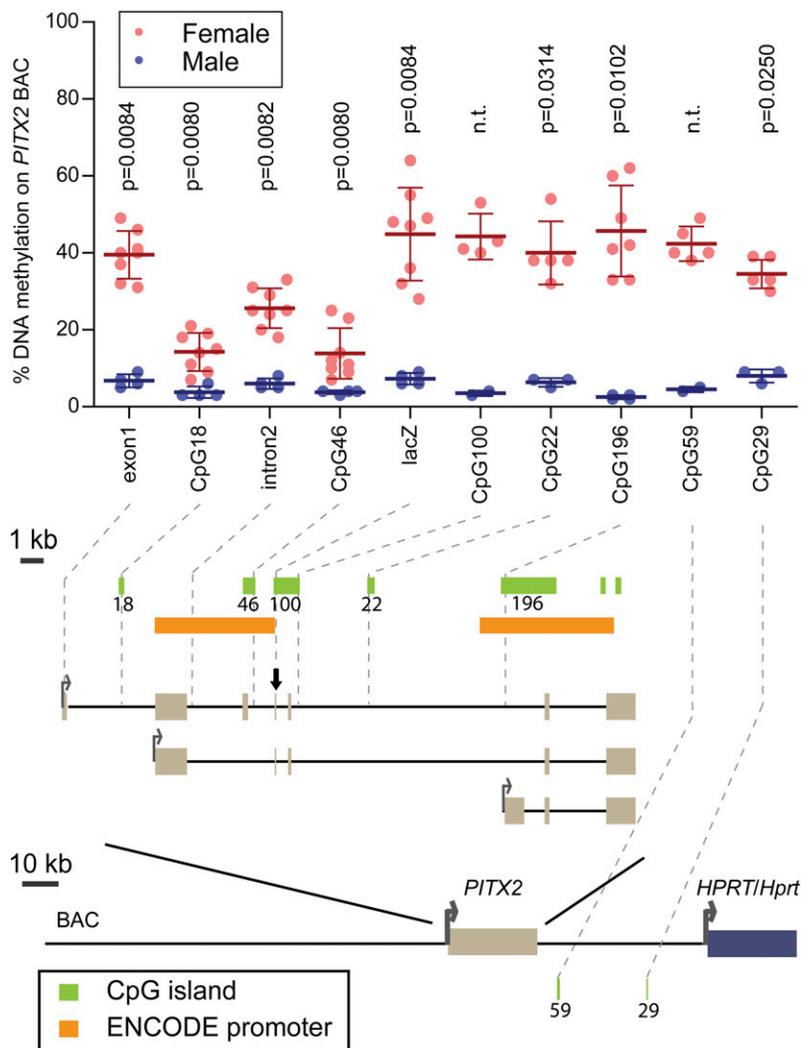


Figure 3 DNA methylation analyses of multiple regions in the *PITX2* BAC construct show varying extents of DNA hypermethylation in females. DNA methylation assays were designed in the gene body, within or outside of CpG islands, as well as in the intergenic CpG islands. The corresponding locations of the DNA methylation assays are shown with dashed lines to the sites on the construct. The gene is depicted with three transcript isoforms (Dreszer *et al.* 2012) and the BAC (RP11-26811) is shown below. Location of the *lacZ* insertion is indicated with a downward arrow. The promoter marks are based on the ENCODE study on chromatin states by Ernst *et al.* (2011). Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, ± 1 standard deviation between mice for the strain. Significance is tested using a Mann–Whitney *t*-test. n.t., not tested due to limited male samples.

et al. 1998; Weber *et al.* 2007), we analyzed promoter DNA methylation of the MiniP constructs to predict their XCI statuses. Heterozygous females overall showed significantly higher DNA methylation levels at promoter CpG islands compared to males (female average, 45%; male average, 12%; $P < 0.0001$). To determine whether there were MiniP promoters that might escape XCI, we analyzed the DNA methylation levels of the constructs separately. DNA methylation levels were analyzed at 46 island-containing MiniP constructs, which originated from 23 human genes. For MiniP constructs 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 MiniP constructs showed promoter DNA hypermethylation in females compared to males, with female and male averages of 44% and 4%, respectively, with the outliers removed in the analysis (Figure 5A). Of the Pleiades constructs that were also examined for expression in the ear notch samples, excluding the outliers, all showed female-specific DNA hypermethylation independent of whether the transgenes displayed expression (Figure S2). For three constructs we observed DNA methylation levels < 2 standard deviations below the average in a single

female, although still > 2 standard deviations above the average in males. These females thus might represent a gene with variable inactivation between mice; however, identification of variable escapees is confounded by skewing of XCI that could result in high standard deviations in DNA methylation levels among females of the same strain. The low DNA methylation level in single females for the three constructs was likely attributable to skewing of XCI, since *HPRT* DNA methylation levels were also lower (8–27%) in these females. Therefore, 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 average DNA methylation in females compared to males, we concluded that none of the MiniP constructs appeared to escape XCI. Promoter DNA hypermethylation was observed in males for the MiniP constructs derived from genes *CARTPT*, *GPX3*, *ICMT*, *OXT*, and *POGZ*, but did not appear to correlate with transgene silencing (Table S1 and Figure S2). It is unknown what DNA sequences in these elements generate exceptions to the DNA methylation patterns observed for the majority of the MiniPs, but interestingly for one of these

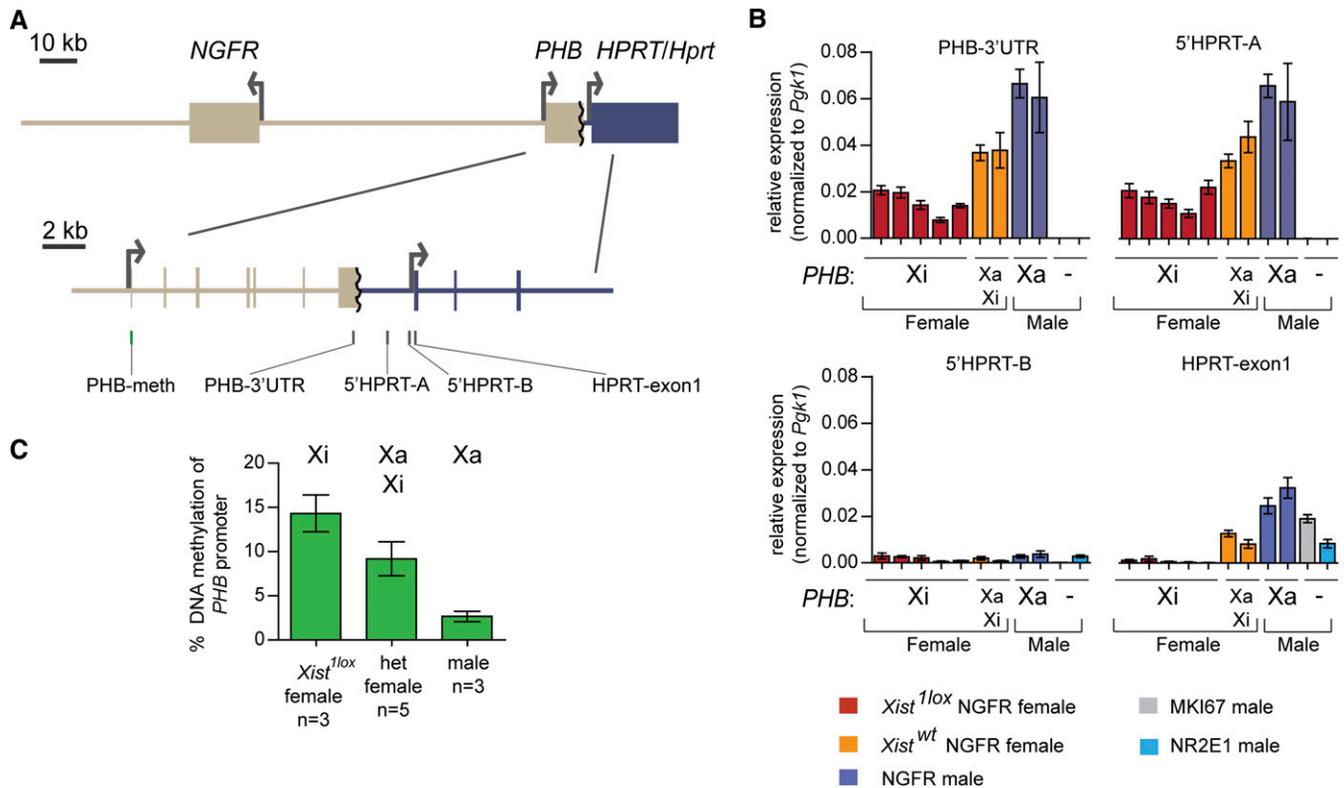


Figure 4 *PHB* on the *NGFR* BAC escapes from XCI. (A) Structure of the *NGFR* MaxiP construct. *PHB* is truncated in the 3'-UTR on the BAC, located ~2.5 kb from the *HPRT/Hprt* locus. Gray and green bars below the MaxiP indicate the locations of the expression and DNA methylation assays, respectively. Internal exons are not shown to scale. (B) Expression of *PHB*, the intergenic region between *PHB* and *HPRT/Hprt*, and *HPRT/Hprt* (exon 1), normalized to *Pgk1*. DNA from a mouse targeted at *Hprt* with BACs for *MKI67* (Ple131) and *NR2E1* (Ple142) served as negative controls (-) since they lack the *PHB* gene. The x-axis indicates whether *PHB* was present only on the Xi or on the Xa in a given mouse, or on either X chromosome (Xa Xi), as in the case for females with random XCI. Error bars indicate ± 1 standard deviation between two qPCR runs. (C) DNA methylation of *PHB* promoter in ear notches of mice carrying the *NGFR* MaxiP on the Xi, either the Xa or the Xi, or the Xa. Error bars indicate ± 1 standard deviation between mice for the strain.

(*CARTPT*) the endogenous island shows DNA hypermethylation at the endogenous promoter. In general, we analyzed fewer mice per construct for the MiniPs, but overall MiniP constructs showed higher levels of DNA methylation compared to the MaxiP constructs (average 45% and 33%, respectively, $P < 0.0001$), perhaps reflecting a closer association of the MiniPs with X-linked *cis*-acting elements or a protective influence of sequences within the large BAC constructs. DNA methylation levels at *HPRT* and *Phf6* were not significantly different between MiniPs and MaxiPs (Figure 5, B and C).

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

lacZ DNA methylation resembled the DNA methylation pattern of the promoter region in *PITX2*, leading us to test the utility of *lacZ* DNA methylation to predict XCI status. Similar to CpG island promoters, female mice showed significantly higher *lacZ* DNA methylation than males ($P < 0.0001$), even though males did have substantial DNA methylation (male and female average DNA methylation levels of 26% and 49%, respectively). Mice with an autosomal *lacZ* showed no significant difference in DNA methylation levels between males and females (Figure S3), indicating that the differ-

ence in the DNA methylation levels of the X-linked *lacZ* between the sexes is likely a consequence of the epigenetic regulation of XCI. The lower level of male (Xa) DNA methylation for X-linked *lacZ* may reflect the previously reported permissive nature of the *Hprt* integration sites (Bronson *et al.* 1996; Cvetkovic *et al.* 2000). 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 6A). 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 screened additional Pleiades constructs for which there was no assay for promoter DNA methylation (Figure 6B). Consistent with the lack of *lacZ* expression in *XXist*^{1lox}/*X*^{MAOA} and *XXist*^{1lox}/*X*^{NR2F2} mice (Figure 1B), *MAOA* and *NR2F2* showed female-specific *lacZ* DNA hypermethylation (Figure 6B), further supporting the usage of this locus as a surrogate to determine XCI status. However, compared to promoter DNA methylation (Figures 1C and 5A), males more often showed DNA hypermethylation of the *lacZ* reporter (Figure 6B and

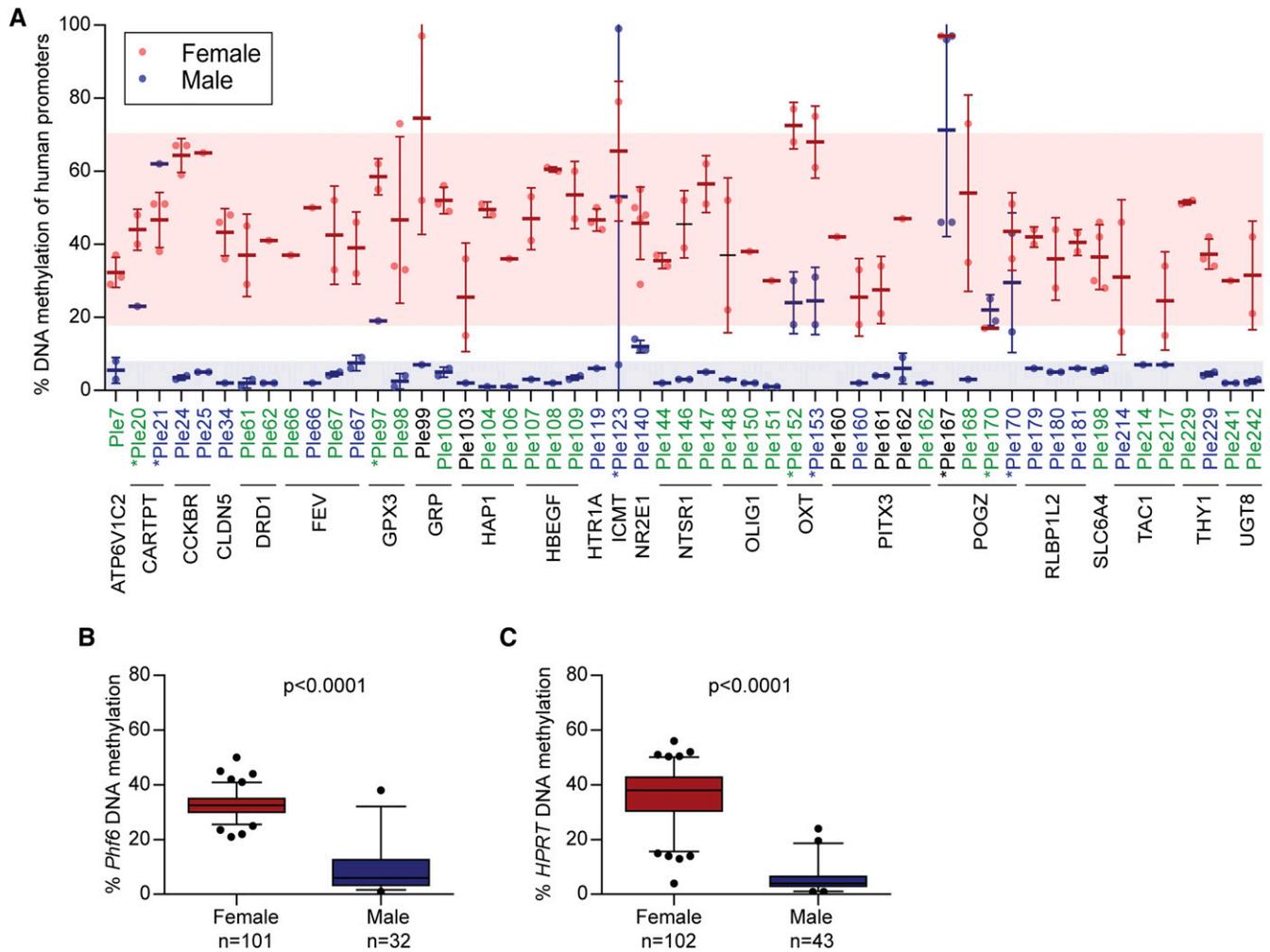


Figure 5 Promoter DNA methylation of the MiniP constructs is reflective of XCI. (A) DNA methylation levels of individual MiniP constructs. Constructs with *lacZ*, EGFP, and EGFP/cre as the reporter are colored in blue, green, and black, respectively. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, ± 1 standard deviation between mice for the strain; shaded regions, 2 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, Ple153), and *POGZ* (Ple167, Ple170). Modified Z-scores >3.5 in absolute values were marked as outliers (Iglewicz and Hoaglin 1993). (B and C) *Phf6* promoter (B) and *HPRT* promoter (C) both showed a significant difference in DNA methylation levels between males and females. A Mann–Whitney *t*-test was used to test for significance. Boxplot whiskers are 5th–95th percentiles. Circles, the average DNA methylation levels of each strain.

Figure S4). Using the criteria of nonoverlapping standard deviations of DNA methylation between the sexes and a male average DNA methylation level <2 standard deviations of the female average of all strains, we excluded seven strains, including two (*DCX* Ple53 and *VIP* Ple250) for which the single male analyzed showed higher DNA methylation than the female average. Thus, we predict an additional 11 constructs subject to XCI based on *lacZ* DNA methylation.

Discussion

Arguably the most dramatic example of *cis*-regulation in the mammalian genome is silencing of one X chromosome in females. However, the *cis*-acting elements involved in spreading heterochromatin along the ~ 155 -Mb chromosome from the initiating elements in the X inactivation center remain

unknown. Having 74 different human transgenes integrated into the mouse X chromosome presented us with an opportunity to assess *cis*-regulation of ~ 1.5 Mb of DNA at the identical genomic location. Analysis of female mice heterozygous for an *Xist* deletion causing nonrandom inactivation of the knock-in-bearing X chromosome provided clear evidence for XCI of four of the knock-ins (Figure 1). As a more rapid approach to assess the XCI status of multiple transgenes, we used DNA methylation as a surrogate measure of inactivation, since promoter DNA hypermethylation in females relative to males can be attributed to DNA methylation of the Xi and thus reflects inactivation of the gene (Yasukochi *et al.* 2010; Cotton *et al.* 2011). We demonstrated that in addition to such DNA hypermethylation for the genes subject to XCI, the escaping *PHB* gene in our system exhibited low promoter DNA methylation in both sexes, validating the usage of DNA

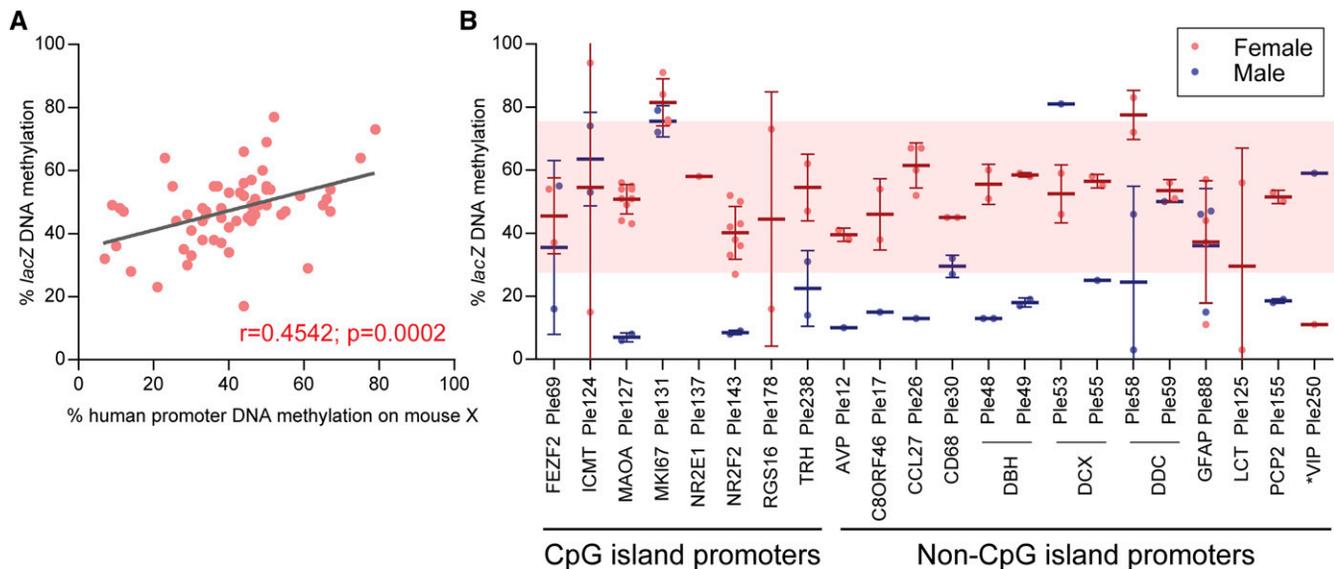


Figure 6 *lacZ* DNA methylation can be used as a surrogate for promoter DNA methylation. (A) Spearman's correlation between *lacZ* and promoter DNA methylation levels in females. Circles, DNA methylation levels from an individual mouse. (B) *lacZ* DNA methylation levels of the *Pleiad*es constructs that 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, ± 1 standard deviation between mice for the strain; shaded regions are the 2 standard deviations from the female average DNA methylation level with the outlier strains removed. Outliers are marked with asterisks (*). Female outlier: *VIP* (*Ple250*).

methylation to detect genes subject to, and escaping from, XCI. As the determination of whether a gene is subject to XCI could be confounded by skewing of XCI and there can be variability between females for whether a gene escapes XCI (Carrel and Willard 2005), we included an assessment of *HPRT* DNA methylation to detect samples with skewed XCI and we required consistently high promoter DNA methylation in multiple females to call a gene subject to XCI. Overall, 92% of the constructs analyzed showed DNA hypermethylation of the human promoters in female mice compared to males (Figures 1C and 5A), indicative of XCI of the knock-in gene.

Our goal in determining the XCI status of the 74 knock-in constructs was to characterize *cis*-acting elements involved in the spread of epigenetic silencing. In 1983 Gartler and Riggs proposed waystations as elements enriched upon the X chromosome that aid in spreading the silencing signal along the chromosome, based on the limited spread of XCI into autosomes that is seen for X;autosome translocations (Gartler and Riggs 1983). Because MiniP constructs are small transgenes, it is not surprising that all were shown to be 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 examined transgenes on the Xi. However, the ~187-kb chicken transferrin transgene is one of the few exceptions that consistently escaped from XCI (Goldman *et al.* 1987, 1998). As the MaxiP constructs were of a similar size (120–195 kb), we anticipated that the MaxiPs originating from autosomes would have a high probability to lack waystations and escape from XCI. However, our results dem-

onstrated that mouse XCI is consistently capable of inactivating foreign transgenes up to 195 kb. Thus, escape from XCI of the chicken transgene may reflect integration into a waystation-poor region relative to *Hprt*. In addition, however, there is now evidence for a different type of *cis*-regulatory element. Recently four different integrations of a *Kdm5c* BAC recapitulated both escape from XCI for *Kdm5c* and silencing for the flanking *Tsypyl2* and *Iqsec2* genes, strongly supporting the existence of a *cis*-acting element on the BAC that controlled escape from XCI (Li and Carrel 2008). Therefore, we hypothesize that there are both waystations and escape elements regulating the spread of XCI.

Escape elements are presumably outside the promoter as none of the 46 MiniPs examined in this study nor the majority of previously examined transgenes escape from XCI. We determined that the *PHB* gene escapes from XCI, and as waystations are reduced in abundance on autosomes and *NGFR* is subject to XCI while being farther from the mouse X-linked DNA than *PHB* (see Table S2), we conclude that the *PHB* region likely carries an escape element to escape from XCI. The observed reduction of *HPRT* DNA methylation to 41% on the Xi when adjacent to *PHB*, from an average of 70% DNA methylation on the Xi for other MaxiP constructs, suggests that the dominant escape element also influenced the *HPRT* locus. The *PHB* gene is truncated in the 3'-UTR of the gene; however, we demonstrated that transcription ceased before the *HPRT* promoter, and since the promoter, splice junctions, and coding sequences are intact, we do not believe that this truncation *per se* affects either the propensity for *PHB* to escape from XCI or the loss of

DNA methylation at *HPRT*. Interestingly, the reduced DNA methylation at *HPRT* was still sufficient to maintain XCI, while the 15% DNA methylation of *PHB* was insufficient for silencing, although there was not full expression from the Xi relative to the Xa. Given that we observed only one such escape element in 47 genes and 1.5 Mb of DNA, our data support the existence of relatively rare dominant escape elements. As escape from XCI is frequent in X;autosome translocations [$\sim 30\%$ (reviewed in Yang *et al.* 2011)], it is likely that, as previously proposed, such escape more generally reflects a lack of waystations rather than the presence of escape elements. We therefore decided to examine the large MaxiP constructs to determine whether there was evidence for particular elements that might be functioning as waystations.

Waystations have been proposed to be repetitive elements, and we calculated the base coverage in the MaxiP for several repeat elements previously positively or negatively correlated with genes escaping XCI: LINE-1, LINE-2, and Alu (Bailey *et al.* 2000; Carrel *et al.* 2006; Wang *et al.* 2006). Only Ple142 (*NR2E1*) and Ple126 (*LCT*) appeared to possess an environment that resembles the genomic context of escapes based on LINE-1, LINE-2, and Alu base coverage on the BAC (Figure S5). However, promoter DNA hypermethylation of *NR2E1* and *LCT* in females suggests that the genes were subject to XCI (Figure 1C). Although only eight BACs were examined, our analysis of LINEs and Alu content suggests that these three repetitive elements are insufficient to determine whether a gene is subject to XCI and that additional repeats or a combination of other factors may be required to provide a better prediction of the XCI status.

In addition to our search for *cis*-acting regulatory elements, our analysis of the Pleiades human knock-in constructs into the X-linked *Hprt* docking site revealed several other insights into the relationship of DNA methylation with XCI. First, we demonstrated that constructs have an intrinsic differential capacity for DNA methylation. Through analyzing MaxiP DNA methylation of female mice with complete non-random XCI due to an *Xist* deletion, we showed that different MaxiP constructs could accumulate DNA methylation at the promoters to different extents on the Xi (Figure 2B). However, the *HPRT* promoter and the *lacZ* reporter, which are shared among the MaxiP constructs, generally exhibited similar levels of DNA methylation (Figure S1), 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 states, 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.

Second, the differential female:male DNA methylation observed on the X is found beyond CpG-island promoters. Xa-specific DNA methylation has been reported in gene bodies (Hellman and Chess 2007). Our results demonstrated that the influence of XCI on DNA methylation of transgenes

applies not only to promoters, but also to gene body and intergenic CpG islands, since all analyzed locations on the *PITX2* BAC showed female-specific DNA hypermethylation that was not observed on the endogenous human chromosome (Figure 3), including regions such as CpG island 29 for which there is no evidence for promoter activity (Ernst *et al.* 2011) or overlap with conserved transcription factor binding sites (TRANSFAC Biobase, <http://www.gene-regulation.com/pub/databases.html>). Therefore, it is possible that the default state of CpG islands on the X chromosome is to acquire DNA methylation on the Xi and this is independent of whether the transgenes were expressed (Figure S2), consistent with the majority of the CpG islands being DNA hypermethylated on the Xi (Cotton *et al.* 2011; Sharp *et al.* 2011), and tissue-specific genes such as the human X-linked androgen receptor showing female-specific DNA hypermethylation independent of expression (Allen *et al.* 1992).

The recognition of CpG islands for DNA methylation on the Xi could explain the DNA hypermethylation in females compared to males on the X chromosome for the *lacZ* reporter (Figure 6), which is essentially an ~ 3000 -bp CpG island. Promoter-less artificial CpG islands inserted into the 3'-UTR of an autosomal and an X-linked gene have been shown to recruit the unmethylated CpG-binding protein Cfp1 and the promoter histone mark H3K4me3 even in the absence of RNA polymerase II binding (Thomson *et al.* 2010), 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 XCI status as the promoter DNA methylation since a higher frequency of males with DNA hypermethylation was observed.

A third insight was the unexpected observation that the differential female:male DNA methylation may reflect not only gain of DNA methylation on the Xi, but also protection from DNA methylation on the Xa. In general, promoter CpG islands are unmethylated on the autosomes; however, $\sim 4\%$ are reported to show DNA methylation, often with variability between tissues (Shen *et al.* 2007). Four of the 35 autosomal CpG islands analyzed (*CARTPT*, *OXT*, *THY1*, and *PITX2*: CpG island 29) showed an average of $>20\%$ DNA methylation in male and female cell lines and/or blood samples (data not shown; Straussman *et al.* 2009). However, when *THY1* and the *PITX2* BAC were present on the X chromosome, they became unmethylated on the Xa; this loss of DNA methylation on the Xa compared to the autosomal locus was also observed for a non-CpG island site (exon 1 of *PITX2*) and the *lacZ* reporter. In the knock-in mice, *CARTPT* continued to show DNA methylation; however, *OXT* dropped from 60% DNA methylation on the autosome to only 20% DNA methylation in males, again showing decreased DNA methylation in males. In general we observed a dominant regulation of XCI on promoter DNA methylation,

with female-specific gain of DNA hypermethylation and in several cases a male-specific loss of DNA methylation.

Overall, our analysis of the Pleiades human promoter constructs integrated into the mouse *Hprt* locus identified 1 of 47 genes in >1.5 Mb of human DNA that escaped inactivation. We propose that there is a dominant *cis*-acting escape element near the *PHB* gene that allows it to escape from XCI in an otherwise inactivated region on the X chromosome. This element exerts an influence on the DNA methylation of the flanking *HPRT* locus, but does not lower DNA methylation to a level that allows expression from the Xi. That eight autosomal BACs ranging in size from 120 to 195 kb contained a gene subject to XCI when integrated at the *Hprt* site suggests that waystations are likely able to act over a distance of at least 100 kb, in the absence of dominantly acting escape elements. Further analyses of BAC integrations at the *Hprt* locus will be useful to identify the nature of these escape elements as well as the boundaries that prevent their influence on adjacent genes.

Acknowledgments

We thank members of the Pleiades Promoter and CanEuCre Projects for the generation and breeding of mice carrying the Pleiades constructs and A. M. Devlin for the use of the pyrosequencing machine. This work was funded by the Canadian Institutes of Health Research (MOP-13690) (to C.J.B.). The Pleiades Promoter Project was funded by Genome Canada, Genome British Columbia, GlaxoSmithKline R&D, British Columbia Mental Health and Addiction Services, the Child and Family Research Institute, University of British Columbia (UBC) Institute of Mental Health, and UBC Office of the Vice President Research (to E.M.S.). The CanEuCre Project was funded by Genome British Columbia (Applied Genomics Consortium Program-CanEuCre-01) (to E.M.S.).

Literature Cited

- Allen, R. C., H. Y. Zoghbi, A. B. Moseley, H. M. Rosenblatt, and J. W. Belmont, 1992 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.* 51: 1229–1239.
- Bailey, J. A., L. Carrel, A. Chakravarti, and E. E. Eichler, 2000 Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. USA* 97: 6634–6639.
- Blankenberg, D., G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus *et al.*, 2010 Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chap. 19: Unit 19.10. 1–21.
- Bronson, S. K., E. G. Plaehn, K. D. Kluckman, J. R. Hagaman, N. Maeda *et al.*, 1996 Single-copy transgenic mice with chosensite integration. *Proc. Natl. Acad. Sci. USA* 93: 9067–9072.
- Carrel, L., and H. F. Willard, 2005 X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434: 400–404.
- Carrel, L., C. Park, S. Tyekucheva, J. Dunn, F. Chiaromonte *et al.*, 2006 Genomic environment predicts expression patterns on the human inactive X chromosome. *PLoS Genet.* 2: e151.
- Chong, S., J. Kontaraki, C. Bonifer, and A. D. Riggs, 2002 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.* 22: 4667–4676.
- Cotton, A. M., L. Avila, M. S. Penaherrera, J. G. Affleck, W. P. Robinson *et al.*, 2009 Inactive X chromosome-specific reduction in placental DNA methylation. *Hum. Mol. Genet.* 18: 3544–3552.
- Cotton, A. M., L. Lam, J. G. Affleck, I. M. Wilson, M. S. Penaherrera *et al.*, 2011 Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum. Genet.* 130: 187–201.
- Csankovszki, G., B. Panning, B. Bates, J. R. Pehrson, and R. Jaenisch, 1999 Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat. Genet.* 22: 323–324.
- Cvetkovic, B., B. Yang, R. A. Williamson, and C. D. Sigmund, 2000 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.* 275: 1073–1078.
- Dreszer, T. R., D. Karolchik, A. S. Zweig, A. S. Hinrichs, B. J. Raney *et al.*, 2012 The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* 40: D918–D923.
- Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shores, L. D. Ward *et al.*, 2011 Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43–49.
- Filippova, G. N., M. K. Cheng, J. M. Moore, J. P. Truong, Y. J. Hu *et al.*, 2005 Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev. Cell* 8: 31–42.
- Gardiner-Garden, M., and M. Frommer, 1987 CpG islands in vertebrate genomes. *J. Mol. Biol.* 196: 261–282.
- Gartler, S. M., and A. D. Riggs, 1983 Mammalian X-chromosome inactivation. *Annu. Rev. Genet.* 17: 155–190.
- Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski *et al.*, 2005 Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15: 1451–1455.
- Goecks, J., A. Nekrutenko, and J. Taylor, 2010 Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11: R86.
- Goldman, M. A., K. R. Stokes, R. L. Idzerda, G. S. Mcknight, R. E. Hammer *et al.*, 1987 A chicken transferrin gene in transgenic mice escapes X-chromosome inactivation. *Science* 236: 593–595.
- Goldman, M. A., P. S. Reeves, C. M. Wirth, W. J. Zupko, M. A. Wong *et al.*, 1998 Comparative methylation analysis of murine transgenes that undergo or escape X-chromosome inactivation. *Chromosome Res.* 6: 397–404.
- Goodfellow, P. J., C. Mondello, S. M. Darling, B. Pym, P. Little *et al.*, 1988 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. USA* 85: 5605–5609.
- Goto, Y., and H. Kimura, 2009 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.* 37: 7416–7428.
- Gribnau, J., S. Luikenhuis, K. Hochedlinger, K. Monkhorst, and R. Jaenisch, 2005 X chromosome choice occurs independently of asynchronous replication timing. *J. Cell Biol.* 168: 365–373.
- Hellman, A., and A. Chess, 2007 Gene body-specific methylation on the active X chromosome. *Science* 315: 1141–1143.
- Iglewicz, B., and D. Hoaglin, 1993 *How to Detect and Handle Outliers*. ASQC Quality Press, Milwaukee.

- Lewandoski, M., E. N. Meyers, and G. R. Martin, 1997 Analysis of Fgf8 gene function in vertebrate development. *Cold Spring Harb. Symp. Quant. Biol.* 62: 159–168.
- Li, N., and L. Carrel, 2008 Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. *Proc. Natl. Acad. Sci. USA* 105: 17055–17060.
- Lopes, A. M., P. S. Burgoyne, A. Ojarikre, J. Bauer, C. A. Sargent *et al.*, 2010 Transcriptional changes in response to X chromosome dosage in the mouse: implications for X inactivation and the molecular basis of Turner Syndrome. *BMC Genomics* 11: 82.
- Lyon, M. F., 2006 Do LINES have a role in X-chromosome inactivation? *J. Biomed. Biotechnol.* 2006: 59746.
- Magness, S. T., H. Jijon, N. Van Houten Fisher, N. E. Sharpless, D. A. Brenner *et al.*, 2004 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.* 173: 1561–1570.
- Matarazzo, M. R., M. L. De Bonis, R. I. Gregory, M. Vacca, R. S. Hansen *et al.*, 2002 Allelic inactivation of the pseudoautosomal gene SYBL1 is controlled by epigenetic mechanisms common to the X and Y chromosomes. *Hum. Mol. Genet.* 11: 3191–3198.
- Migeon, B. R., 2011 The single active X in human cells: evolutionary tinkering personified. *Hum. Genet.* 130: 281–293.
- Nguyen, D. K., F. Yang, R. Kaul, C. Alkan, A. Antonellis *et al.*, 2011 Cln4–2 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.* 21: 402–409.
- Portales-Casamar, E., D. J. Swanson, L. Liu, C. N. De Leeuw, K. G. Banks *et al.*, 2010 A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc. Natl. Acad. Sci. USA* 107: 16589–16594.
- Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. Mclay *et al.*, 2005 The DNA sequence of the human X chromosome. *Nature* 434: 325–337.
- Schmouh, J. F., R. J. Bonaguro, X. Corso-Diaz, and E. M. Simpson, 2012 Modelling human regulatory variation in mouse: finding the function in genome-wide association studies and whole-genome sequencing. *PLoS Genet.* 8: e1002544.
- Sharp, A. J., E. Stathaki, E. Migliavacca, M. Brahmachary, S. B. Montgomery *et al.*, 2011 DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* 21: 1592–1600.
- Shen, L., Y. Kondo, Y. Guo, J. Zhang, L. Zhang *et al.*, 2007 Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* 3: 2023–2036.
- Simpson, E. M., C. C. Linder, E. E. Sargent, M. T. Davisson, L. E. Mobraaten *et al.*, 1997 Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat. Genet.* 16: 19–27.
- Soriano, P., 1999 Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* 21: 70–71.
- Splinter, E., E. De Wit, E. P. Nora, P. Klous, H. J. Van De Werken *et al.*, 2011 The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* 25: 1371–1383.
- Straussman, R., D. Nejman, D. Roberts, I. Steinfeld, B. Blum *et al.*, 2009 Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* 16: 564–571.
- Thomson, J. P., P. J. Skene, J. Selfridge, T. Clouaire, J. Guy *et al.*, 2010 CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464: 1082–1086.
- Wang, Y., and F. C. Leung, 2004 An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* 20: 1170–1177.
- Wang, Z., H. F. Willard, S. Mukherjee, and T. S. Furey, 2006 Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Comput. Biol.* 2: e113.
- Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo *et al.*, 2007 Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39: 457–466.
- Wu, H., R. Fassler, A. Schnieke, D. Barker, K. H. Lee *et al.*, 1992 An X-linked human collagen transgene escapes X inactivation in a subset of cells. *Development* 116: 687–695.
- Wutz, A., 2011 Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* 12: 542–553.
- Yang, C., A. G. Chapman, A. D. Kelsey, J. Minks, A. M. Cotton *et al.*, 2011 X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum. Genet.* 130: 175–185.
- Yang, F., T. Babak, J. Shendure, and C. M. Disteche, 2010 Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* 20: 614–622.
- Yang, G. S., K. G. Banks, R. J. Bonaguro, G. Wilson, L. Dreolini *et al.*, 2009 Next generation tools for high-throughput promoter and expression analysis employing single-copy knock-ins at the Hprt1 locus. *Genomics* 93: 196–204.
- Yasukochi, Y., O. Maruyama, M. C. Mahajan, C. Padden, G. M. Euskirchen *et al.*, 2010 X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proc. Natl. Acad. Sci. USA* 107: 3704–3709.

Communicating editor: J. A. Birchler

GENETICS

Supporting Information

<http://www.genetics.org/content/suppl/2012/09/28/genetics.112.143743.DC1>

Targeting of >1.5 Mb of Human DNA into the Mouse X Chromosome Reveals Presence of *cis*-Acting Regulators of Epigenetic Silencing

Christine Yang, Andrea J. McLeod, Allison M. Cotton, Charles N. de Leeuw, Stéphanie Laprise,
Kathleen G. Banks, Elizabeth M. Simpson, and Carolyn J. Brown

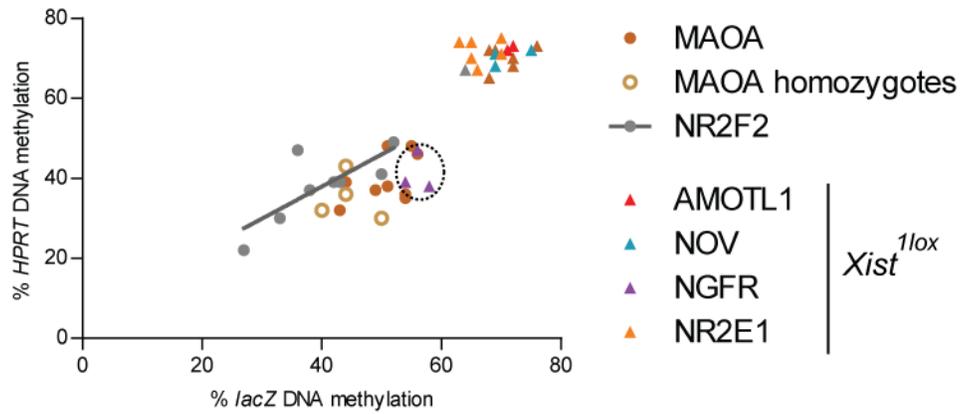


Figure S1 DNA methylation of *HPRT* and *lacZ* was high on the Xi except for *NGFR* (Ple133). Spearman correlation between *HPRT* and *lacZ* DNA methylation levels was significant for *NR2F2* (Ple143: $r=0.7545$; $p=0.0368$), but not for *MAOA* (Ple127). DNA methylation levels of *NGFR* in $Xist^{1lox}$ females (encircled on the graph) were substantially lower than *AMOTL1*, *NOV*, and *NR2E1* in $Xist^{1lox}$ females. Filled circles, DNA methylation in individual female mice heterozygous for the knock-in and carrying 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 circles).

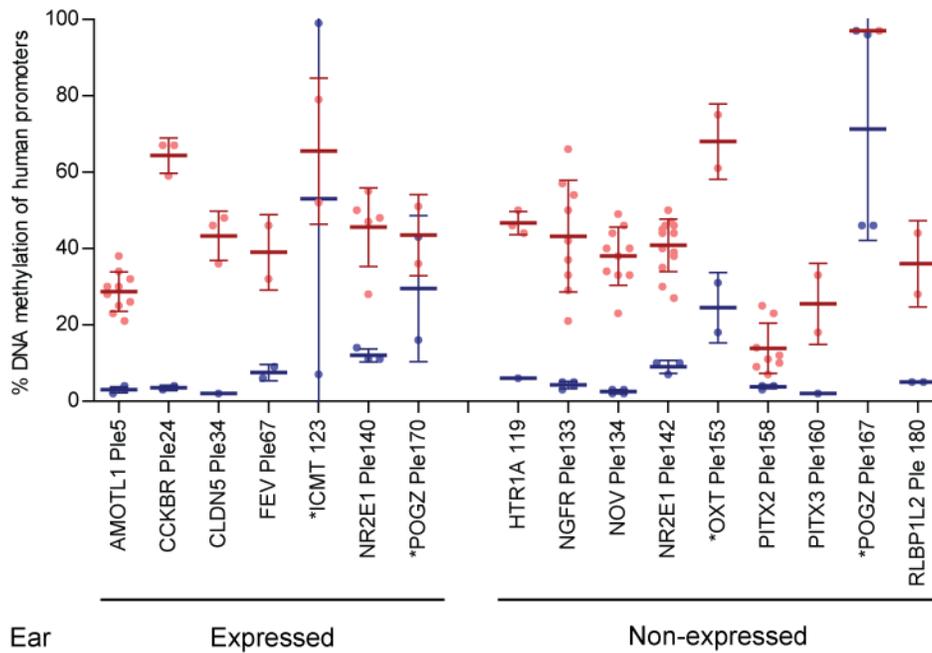


Figure S2 Promoter DNA methylation of the Pleiades constructs is reflective of XCI and independent of expression. Females generally showed DNA hypermethylation compared to males regardless of expression status of the transgene. Expression status (expressed and non-expressed) was based on *lacZ* staining in the ear notches of mice. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, \pm one standard deviation between mice for the strain. Outliers in DNA methylation are marked with asterisks (*). Female outlier: *POGZ* (Ple167). Male outliers: *ICMT* (Ple123), *OXT* (Ple152, Ple153), *POGZ* (Ple167, Ple170). Modified Z-score greater than 3.5 in absolute values were marked as outliers (Iglewicz and Hoaglin 1993).

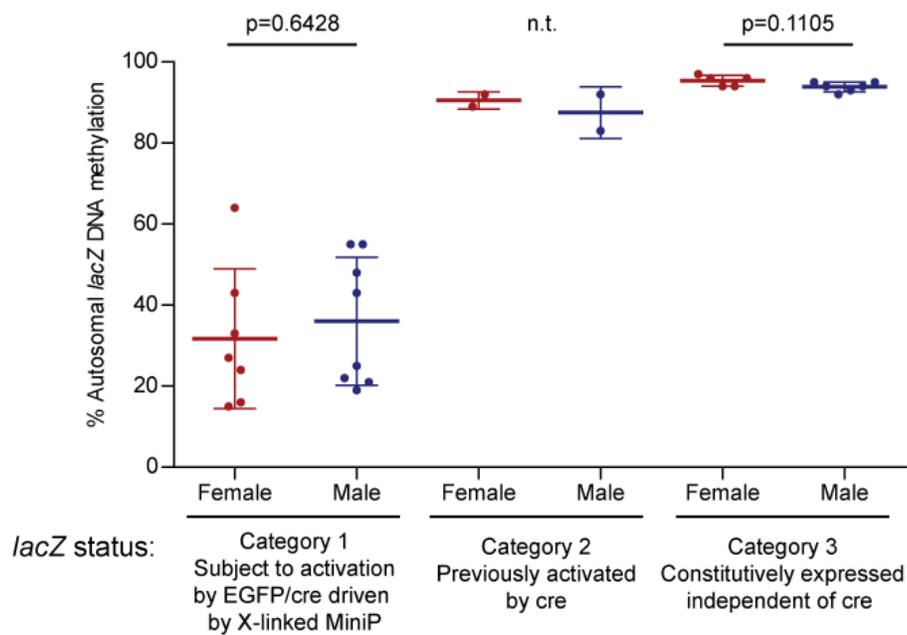


Figure S3 DNA methylation of autosomal *lacZ* reporter was not significantly different between the sexes. Three categories of mice with an autosomal *lacZ* reporter at the *ROSA26* locus (*Gt(ROSA)26Sor^{tm1Sor}/J*; Soriano 1999) were assessed. Category 1: *lacZ* is activated by EGFP/cre driven by different X-linked MiniPs. Category 2: *lacZ* had been previously activated by cre. Category 3: *lacZ* expression did not required activation by cre (Friedrich and Soriano 1991). Each circle represents the level of DNA methylation in an individual mouse. In all categories the autosomal *lacZ* was driven by the same promoter. Bar, average; error bars, \pm one standard deviation between mice for the strain. Significance was tested using Mann-Whitney t-test; n.t., not tested.

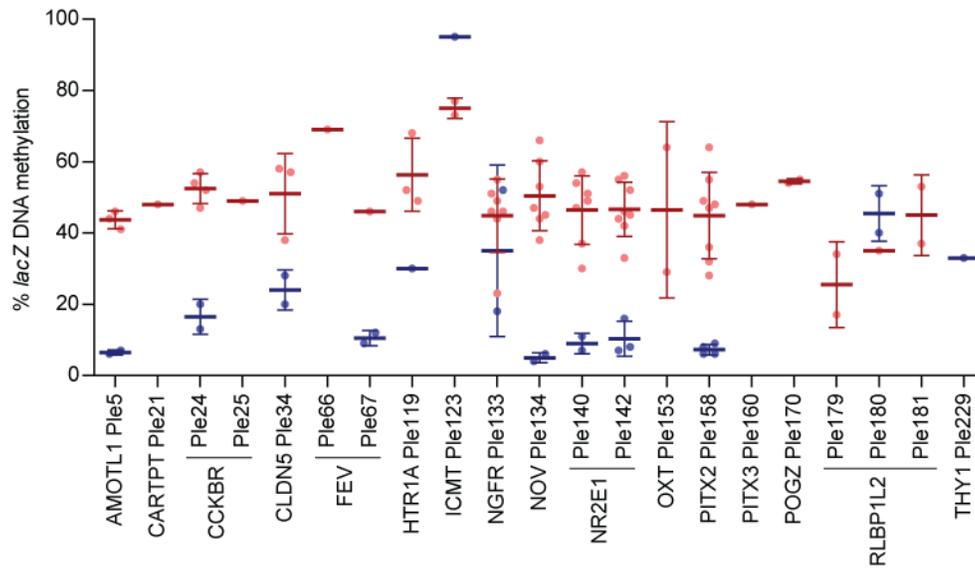


Figure S4 *lacZ* reporter in constructs that were analysed for promoter DNA methylation generally showed DNA hypermethylation in females compared to males. Circles, DNA methylation of the individual sample; bar in the center of the error bars, average DNA methylation for the strain; error bars, \pm one standard deviation between mice for the strain.

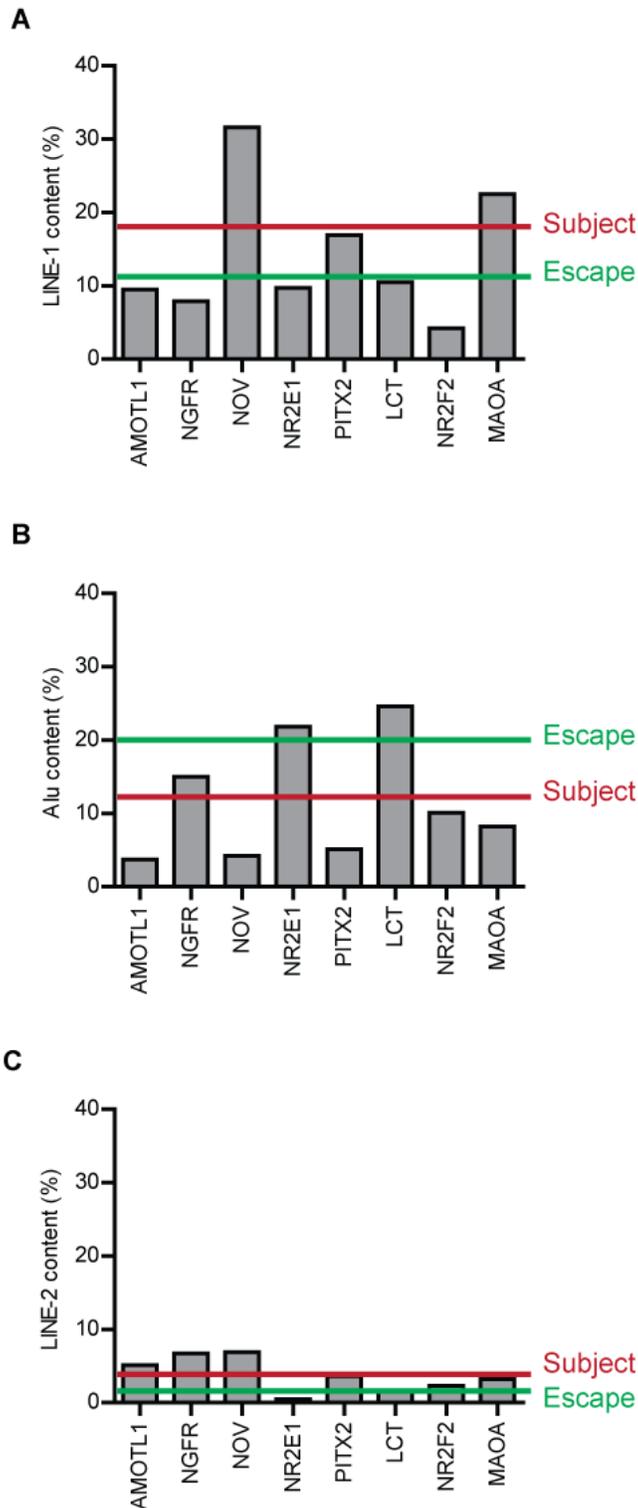


Figure S5 MaxiPs, while all subject to XCI, showed no correlation between the XCI status and repeat contents. Repeat content for LINE-1 (A), Alu (B), and LINE-2 (C) did not correlate strongly with the XCI statuses of the MaxiPs. Genes that are subject to XCI are predicted to be enriched in LINE-1 and LINE-2 and depleted in Alu, while genes that escape from XCI are predicted to be depleted in LINE-1 and LINE-2 and enriched in Alu. The median levels of repeat content in 100-kb windows surrounding genes subject to XCI (red line) and genes escaping from XCI (green line) were estimated from Wang *et al.* (2006).

References

- Friedrich, G., and P. Soriano, 1991 Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-1523.
- Iglewicz, B., and D. Hoaglin, 1993 *How to Detect and Handle Outliers*. ASQC Quality Press, Milwaukee.
- Soriano, P., 1999 Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**: 70-71.
- Wang, Z., H. F. Willard, S. Mukherjee and T. S. Furey, 2006 Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Comput Biol* **2**: e113.

Table S1 Summary of MiniP constructs

Gene name	Ple#	Reporter	Promoter size (kb)	Average female promoter methylation (%)	Average male promoter methylation (%)	lacZ expression in ear	Promoter CpG island
ATP6V1C2	Ple7	EGFP	2.4	32	6	N/A	119
CARTPT	Ple20	EGFP	3.6	44	23*	N/A	63
	Ple21	lacZ	2.4	47	62*	NT	
CCKBR	Ple24	lacZ	3.7	64	4	+	26
	Ple25	lacZ	3.6	65	5	NT	
CLDN5	Ple34	lacZ	3.5	43	2	+	147
DRD1	Ple61	EGFP	3.7	37	2	N/A	145
	Ple62	EGFP	3.4	41	N/A	N/A	
FEV	Ple66	EGFP	1.5	37	N/A	N/A	126
		lacZ		50	2	NT	
	Ple67	EGFP	2.2	43	5	N/A	
		lacZ		39	8	+	
GPX3	Ple97	EGFP	3.4	59	19*	N/A	43
	Ple98	EGFP	2.9	47	3	N/A	
GRP	Ple99	EGFP/cre	1.4	75	7	NT	57
	Ple100	EGFP	2.4	52	5	N/A	
HAP1	Ple103	EGFP/cre	0.7	26	2	NT	130
	Ple104	EGFP	2.1	50	1	N/A	
	Ple106	EGFP	1.1	36	1	N/A	
HBEGF	Ple107	EGFP	3.4	47	3	N/A	64
	Ple108	EGFP	3.6	61	2	N/A	
	Ple109	EGFP	3.5	54	4	N/A	
HTR1A	Ple119	lacZ	3.6	47	6	-	100
ICMT	Ple123	lacZ	2.2	66	53*	+	74
NR2E1	Ple140	lacZ	3.9	46 (ear) 45 (brain) 32 (liver)	12 (ear) 10 (brain) 9 (liver)	+	318
NTSR1	Ple144	EGFP	3.0	36	2	N/A	158
	Ple146	EGFP	3.8	46	3	N/A	
	Ple147	EGFP	2.1	57	5	N/A	
OLIG1	Ple148	EGFP	0.9	37	3	N/A	202
	Ple150	EGFP	2.9	38	2	N/A	
	Ple151	EGFP	2.6	30	1	N/A	
OXT	Ple152	EGFP	1.3	73	24*	N/A	105
	Ple153	lacZ	2.7	68	25*	-	
PITX3	Ple160	EGFP/cre	2.5	42	N/A	NT	104
		lacZ		26	2	-	
	Ple161	EGFP/cre	3.9	28	4	NT	
	Ple162	EGFP/cre	3.6	47	6	NT	
EGFP		N/A		2	N/A		
POGZ	Ple167	EGFP/cre	1.2	97*	71*	+	48
	Ple168	EGFP	4.0	54	3	N/A	
	Ple170	EGFP	3.2	17	22*	N/A	
		lacZ		44	30*	+	
RLBP1L2	Ple179	lacZ	3.5	42	6	NT	59 ⁺
	Ple180	lacZ	3.8	36	5	-	
	Ple181	lacZ	3.7	41	6	NT	

SLC6A4	Ple198	EGFP	2.6	37	6	N/A	81
TAC1	Ple214	EGFP	2.7	N/A	7	N/A	137
		lacZ		31	N/A	NT	
	Ple217	EGFP	0.9	25	7	N/A	
THY1	Ple229	EGFP	1.2	52	N/A	N/A	48
		lacZ		37	5	NT	
UGT8	Ple241	EGFP	3.1	30	2	N/A	121
	Ple242	EGFP	2.5	32	3	N/A	

All average DNA methylation levels refer to DNA methylation examined in the ear notch samples only, except for *NR2E1* where DNA methylation was also examined in brain and liver. The numbers indicated for the promoter CpG island refer to the UCSC CpG island designations. †full-length CpG island; * detected as outliers in levels of DNA methylation in males and/or females; NT, not tested; N/A, not applicable.

Table S2 Summary of MaxiP constructs

Gene name	Ple#	BAC size (kb)	Minimal distance from TSS to X boundary ^a (kb)	Average <i>Xist</i> ^{1lox} female promoter methylation (%)	Average female promoter methylation (%)	Average male promoter methylation (%)	<i>lacZ</i> expression in ear	Promoter CpG island ^b	Other CpG islands ^c	Other transcripts on BAC ^d
<i>AMOTL1</i>	Ple 5	193.7	80.5	50	29	3	+	114	+	-
<i>MCM6 (LCT)</i>	Ple126	138.6	24.2	ND	49	4	N/A	91	-	+
<i>NGFR</i>	Ple133	157.9	78.1	69 #	43	4	-	255	+	+
<i>NOV</i>	Ple134	184.3	59.8	66	38	3	-	57	-	-
<i>NR2E1</i>	Ple142	120.2	55.2	58	39 #	9 #	-	318	+	+
<i>PHB</i>	Ple133	157.9	20.3	14 #	9	3	N/A	^	+	+
<i>PITX2</i> -exon1	Ple158	182.2	82.1	NT	40	7	-	-	+	-
<i>PITX2</i> -CpG18					14	4				
<i>PITX2</i> -intron2					26	6				
<i>PITX2</i> -CpG46					14	4				
<i>PITX2</i> -CpG100					44	4				
<i>PITX2</i> -CpG22					40	6				
<i>PITX2</i> -CpG196					46	3				
<i>PITX2</i> -CpG59					42	5				
<i>PITX2</i> -CpG29					35	8				

MCM6 is a gene on the *LCT* BAC and the results of *MCM6* are shown in the table. The table shows DNA methylation levels in ear notch samples only. The constructs indicated with # showed similar DNA methylation levels in brain and/or liver to ear notch samples. TSS, transcription start site; ^a shortest distance from TSS to mouse X chromosomal DNA, which could be from the TSS to mouse *Hprt* locus or from the TSS to the intergenic region between *Hprt* and *Phf6*; ^b presence of CpG island overlapping TSS; ^c presence of non-promoter associated CpG island(s) on the BAC, as defined by UCSC; ^d presence of other TSS that is not associated with the gene of interest on the BAC, as defined by UCSC; ND, not done; NT, not tested; N/A, not applicable. The numbers indicated for the promoter CpG island refer to the UCSC designation for CpG islands. ^, no CpG island found at *PHB* promoter according to UCSC definition of islands, but there is an island of intermediate CpG density.

Table S3 Primer table

Assay	Construct	Distance of assay from TSS (bp)	Sequence (5'-3')	Ta (°C)	Position of CpGs analysed
AMOTL1_F1	Ple 5	72 (5')	GGGATAAAGGAAGGGATGTTG	55	8-13
AMOTL1_R1	(RP11-936P10)		*TCACTAAAACCCTACACTCCACC		
AMOTL1_S2			GGAGGGTGTGTTGTAGA		
ATP6V1C2_F1	Ple 7	544 (5')	AGGTGGGAGTTTTTTGGGTAAT	53.9	1-5
ATP6V1C2_R1			*CAAAAAAATCACCTACTCCCAAATATCT		
ATP6V1C2_S1			GGGAGTTTTTTGGGTAA		
CARTPT_F1	Ple 20-21	296 (5')	GTAATGTGGTTGTTTGGAGGTAATA	55	1-3
CARTPT_R1			*TCCCAACACCTAACAATAATAACAACT		
CARTPT_S1			TGGTTGTTTGGAGGTAAT		
CCKBR_F1	Ple 24-25	424 (5')	GAGGAGTTGTAGGGAATTA	55	1-5
CCKBR_R1			*AATACTTTAATCTAAACCTAAAACC		
CCKBR_S1			GAGGAGTTGTAGGGAAT		
CLDN5_F1	Ple 34	820 (3')	AGTTGTTAGAGGTTTTGTGATTG	53.3	1-5
CLDN5_R1			*AAAAATACCCTCTTTAAAAATTC		
CLDN5_S1			GTTGTTAGAGTTTTGTGA		
DRD1_F1	Ple 61-62	38 (3')	TATTGTTATAGGTTTTTGAGAGGT	53.3	1-5
DRD1_R1			*CCTTCAACCCTACAAAACAAA		
DRD1_S1			ATTGTTATAGTTTTTTGAGA		
FEV_F1	Ple 66-67	1 (3')	*GGAGGGGGAGGAGAGTGA	53.9	1-4
FEV_R1			CCCTCCCTAAAACCTTCTTC		
FEV_S1			AAAACCTTCTTCCAA		
GPX3_F1	Ple 97-98	39 (3')	TGGGGAGTTGAGGGTAAGT	55	1-5
GPX3_R1			*CCCAACCACCTTCAAAC		
GPX3_S2			GGGAGTTGAGGGTAAGT		
GRP_F1	Ple 99-100	216 (5')	AGAGGGAGGAGTTTATTAATTGTGTT	55	1-5
GRP_R1			*CATTACCCCTCTTTTTCTCCT		
GRP_S1			AAATTGTGTTGGATGGA		
HAP1_F1	Ple 103-104,	262 (5')	GGAGGGGTTGTTTTAGTTAGGG	53.9	1-4
HAP1_R1	106		*ATTTTTTCTACCCTCTCCATCTCC		
HAP1_S1			GTTGTTTTTAGTTAGGGATT		
HBEGFc_F1	Ple 107-109	209 (5')	GTTTGGGGAAAGGTAGGAAT	55	1-5
HBEGFc_R1			*TCACAATTTTTAAAACCAAACC		
HBEGFc_S1			GTTTGGGGAAAGGTA		
HPRTb_F1	All constructs	94 (5')	GGAATTAGGGAGTTTTTTGAATAGG	55	1-3
HPRTb_R1			*CCTACCAATTTACAACTCACTAAATA		
HPRTb_S1			GGGAGGGAAAGGGT		
HTR1A_F1	Ple 119	266 (5')	*TTTGGGATTGGAGATTGTTTGT	55	5-8

HTR1A_R1			ACTCCAACCTAAAAAATAAAATTAACCT		
HTR1A_S1			CTAAAAAATAAAATTAACC		
ICMT_F1	Ple 123-124	206 (5')	GGAATTTTTTGAGTTTGGGATTAA	58.3	1-4
ICMT_R1			*CATCCCAACTCTAAACCAAACCTCTATA		
ICMT_S1			TGGGATTAAGTTTGGATA		
LacZ-meth_F1	Reporter	~724 (3' of	TGTATTGGAGGTTGAAGTTTAGATGT	55	1-4
LacZ-meth_R1		<i>lacZ</i> seq)	*TTTCACCTACCATAAAAAAATAATTAC		
LacZ-meth_S1			TGGAGGTTGAAGTTTAGAT		
MCM6_F1	Ple 126	190 (5')	*GTGGAATGATTTAAAGAATATTTGAAAA	55	3-7
MCM6_R1	(RP11-406M16)		CCTTCTAAAAAAACCCATCTACCTT		
MCM6_S1			CTTCTAAAAAAACCCATC		
mPhf6_F1	Endogenous	305 (5')	GTAAGGGTTAAGTTTGTGTATTTGT	55	1-3
mPhf6_R1			*CCAAAAAACCTAAACCAAATCCT		
mPhf6_S1			GTTAAGGTTTGTGTATTTGTT		
NGFR_F1	Ple 133	279 (5')	AGGAAGATGGGTAAAGAGAGTGAATT	55	1-4
NGFR_R1	(RP11-158L10)		*TCCCTACCTTATCCCTTAAACCT		
NGFR_S1			GGTAAGAGAGTGAATTTTGT		
NOV_F33	Ple 134	26 (5')	GTTTTTTATTTTTGGGAAAAGTT	48	1-6
NOV_R33	(RP11-840I14)		*ACAATTAATAAATACTACTCTCCTTAAA		
NOV_S1			TTTTTTGGGAAAAGTTAG		
NR2E1_F1	Ple 140, 142	22 (3')	*TTAGGAGTTGGGGAAAAGTTAA	55	2-4
NR2E1_R1	(RP11-144P8)		AACTAAATCCCCTATAATATCTCCAAAA		
NR2E1_S1			ATCCCCTATAATATCTCCA		
NR2E1-CpG40_F1	Ple 140, 142	Intragenic	*GAGTGTTTTTTGGGTTTGGAGTA	52	7
NR2E1-CpG40_R1	(RP11-144P8)		ATATCCTCCAACCCATTACCC		
NR2E1-CpG40_S1			ATCCTCCAACCCATT		
NR2E1-CpG75_F1	Ple 140, 142	Intergenic	*GGGTTTAGATTATTGTATTTGTTGAG	52	5
NR2E1-CpG75_R30	(RP11-144P8)		ACTAATAAAACCAAACCTCTTAAACTT		
NR2E1-CpG75_S1			CTAATAAAACCAAACCTCTT		
NTSR1_F1	Ple 144,	368 (5')	GTTGGGGGAGGTGTATAGTT	58.3	1-3
NTSR1_R1	146-147		*TACCACCCTCTCCCTATT		
NTSR1_S1			TTGGGGGAGGTGTAT		
OLIG1_F1	Ple 148,	36 (5')	GAGGGAGGTTGTTTTGAGTAGA	55	1-11
OLIG1_R1	150-151		*CCCTACCCTTTAAACCC		
OLIG1_S1			GGTATAAGTAGTTAATGAATA		
OXT_F1	Ple 152-153	107 (5')	GTTTTGTTAATGAAGAGGAAAGTT	55	1-3
OXT_R1			*ACCTAACCTTTTATACCTAAACAT		
OXT_S1			TTGTTAATGAAGAGGAAAGT		
PHB-IC_F1	Ple 133	1 (5')	GAATTAGGGTGAGGTTTAAAGTTATTTT	58.3	1-5
PHB-IC_R1	(RP11-158L10)		*ACATAAATTCCTCAACACACA		
PHB-IC_S1			GGGTGAGGTTTAAAGTTAT		

PITX2-CpG18_F1	Ple 158	Gene body	GGGATTGGGGTTAATTAGTTTTTGG	58.3	1-4
PITX2-CpG18_R1	(RP11-26811)	1385 (5') of	*AACTCCCTCCCCTTTCAAATTC		
PITX2-CpG18_S1		Alt-TSS	AGGGATTGGGGTTAA		
PITX2-CpG22_F1	Ple 158	Gene body	*AAATTTGTAGTTTATTTGAAAGGTGTTT	58.3	1-3
PITX2-CpG22_R1	(RP11-26811)		ACAATAATACAATTTCCCTAAAAATA		
PITX2-CpG22_S1			AACTAATAACAATTTCCCTA		
PITX2-CpG29_F1	Ple 158	Intergenic	*GTTTTGATTGGAGGAGGTATTAGT	58.3	1-5
PITX2-CpG29_R1	(RP11-26811)		AACCCTAACCCACCAATACTCC		
PITX2-CpG29_S1			AACCCTAACCCACCA		
PITX2-CpG46_F92	Ple 158	Gene body	GTATTTTTTTAGGTTTGTGGTAGAG	48	1-5
PITX2-CpG46_R92	(RP11-26811)		*CCCCAACCAACCAATCTTTTT		
PITX2-CpG46_S1			TGGTAGAGAAGGGGA		
PITX2-CpG59_F1	Ple 158	Intergenic	TGATTAGGATTTTTGGATTATGAATT	55	1-7
PITX2-CpG59_R1	(RP11-26811)		*CCATATCATTAAACAAAACTAACATT		
PITX2-CpG59_S1			GGATTTTTGGATTATGA		
PITX2-CpG100b_F1	Ple 158	Gene body	TGGAGTGGAAAAGTGGTTAATA	56.3	1-6
PITX2-CpG100b_R1	(RP11-26811)		*AACCTAAATAACTAAATAAACCTAAT		
PITX2-CpG100b_S1			GTGGAAAAGTGGTTAATA		
PITX2-CpG196b_F1	Ple 158	Gene body	TGGTTTTAAGATGTTAGGTTAATAGGG	55	1-6
PITX2-CpG196b_R1	(RP11-26811)	18 (3') of	*ACTCAACTCCAAACACCCAAA		
PITX2-CpG196b_S1		Alt-TSS	GATGTTAGGTTAATAGGGAA		
PITX2-exon1_F1	Ple 158	146 (3')	*AAAGGTTAGAGGGATTAATATAGGT	58.3	1-3
PITX2-exon1_R1	(RP11-26811)		ACTTCCCTTCTACAACAATTTCT		
PITX2-exon1_S1			ACTTCCCTTCTACAACAAT		
PITX2-intron2_F1	Ple 158	Gene body	AGATATTAATAATTTATAGGGTGTGAA	53.3	1-3
PITX2-intron2_R1	(RP11-26811)	1580 (3') of	*AAACTTTATACCCAACCCTTTATCT		
PITX2-intron2_S1		Alt-TSS	TAATTTATAGGGTGTGAAG		
PITX3_F1	Ple 160-162	78 (5')	GAGTTTTAGTAGGGTAGTTGGAAAGG	55	1-10
PITX3_R20			*CCATTCACCTTATAACAAACAAAA		
PITX3_S1			GTAGGGTAGTTGGAAAGG		
POGZ_F1	Ple 167-168,	84 (5')	GTAGGGGTTTGATGAGTTTATGA	55	1-4
POGZ_R1	170		*CTTTTTCACCACCTCCCAATTA		
POGZ_S1			GGGTTTGATGAGTTTA		
RLBP1L2b_F1	Ple 179-181	396 (5')	TGGGGAGGTTGGAAAGTATG	58.3	1-5
RLBP1L2b_R1			*CCCCTCCTCAACAACTACT		
RLBP1L2b_S1			GGGGAGGTTGGAAAG		
SLC6A4_F1	Ple 198	134 (5')	*TGTTAGGTTTTAGGAAGAAAGAGAGA	58.3	6-10
SLC6A4_R68			CATCCTAACTTCTACTCTTAACTTTA		
SLC6A4_S1			AACTACACAAAAAACAAAT		
TAC1_F1	Ple 214, 217	2 (3')	GAATTTAATTGGGTTTAGATGTTATGGG	55	1-6
TAC1_R1			*TTAATTAACCCCTCCTCCTTT		

TAC1_S1			GGGTTTAGATGTTATGGGTA		
THY1_F1	Ple 229	24 (5')	GGAGGTGGGTTTTAGTTGAAA	58.3	1-3
THY1_R1			*AAAAAACATTATCCTCCTCCCTAAA		
THY1_S1			TGAAAAGGAAATGTGGA		
UGT8_F1	Ple 241-242	88 (5')	GTGGGTGGTGGTAGAAAG	58.3	1-4
UGT8_R1			*CCCACTCTTCCCTCTTTA		
UGT8_S1			TGGGTGGTGGTAGAA		
qhHPRT_F2	All constructs	7 (3')	CCTGCTTCTCCTCAGCTTCAG	60	Expression
qhHPRT_R2			CGGGAAAGCCGAGAGGTT		
qHPRT-5'A_F1	All constructs	1121 (5')	CAAATCTCCTGCCATCACATACC	60	Expression
qHPRT-5'A_R1			AGTGCCAGCACATAGTTGGT		
qHPRT-5'B_F1	All constructs	258 (5')	GCCACAGGTAGTGCAAGGTCTT	60	Expression
qHPRT-5'B_R1			CCAGTCATCGCGTGAATCCT		
qPgk1-e1_F1	Endogenous	64 (3')	CGTCTGCCGCGCTGTT	60	Expression
qPgk1-e1_R1			AACACCGTGAGGTCGAAAGG		
qPHB-3'UTR_F1	Ple 133	3'UTR	CTGTCACTGATGGAAGGTTTGC	60	Expression
qPHB-3'UTR_R1	(RP11-158L10)		AGGCCTGCCTTCTCAGTTCA		
mFln_F6	Endogenous	Intragenic	*CCAGCTTCCCTAGTCCAAATGC	58.3	SNP
mFln_R3			TGCATACAGTCAGTGCAAGTACAAG		
mFln_S1			CCTAGAGAGGGCTGAA		

Primers biotinylated at the 5' end are indicated with an asterisk (*). Position of analysed CpGs is relative to the sequencing primer. Distance of pyrosequencing assays from transcription start site (TSS) is the distance of the closest CpG to the TSS. Ta, annealing temperature for PCR; Alt-TSS, alternative transcription start site; SNP, single-nucleotide polymorphism.