Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain

Christopher L. Keown[a], Joel B. Berletch[b], Rosa Castanon[c], Joseph R. Nery[c], Christine M. Disteche[b,d], Joseph R. Ecker[c,e], and Eran A. Mukamel[a,1]

[a]Department of Cognitive Science, University of California, San Diego, CA 92037; [b]Department of Pathology, University of Washington, Seattle, WA 98195; [c]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037; [d]Department of Medicine, University of Washington, Seattle, WA 98195; and [e]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037

Edited by Paul D. Soloway, Cornell University, Ithaca, NY, and accepted by Editorial Board Member Andrew G. Clark December 20, 2016 (received for review July 21, 2016)

DNA methylation at gene promoters in a CG context is associated with transcriptional repression, including at genes silenced on the inactive X chromosome in females. Non-CG methylation (mCH) is a distinct feature of the neuronal epigenome that is differentially distributed between males and females on the X chromosome. However, little is known about differences in mCH on the active (Xa) and inactive (Xi) X chromosomes because stochastic X-chromosome inactivation (XCI) confounds allele-specific epigenomic profiling. We used whole-genome bisulfite sequencing in a mouse model with nonrandom XCI to examine allele-specific DNA methylation in frontal cortex. Xi was largely devoid of mCH, whereas Xa contained abundant mCH similar to the male X chromosome and the autosomes. In contrast to the repressive association of DNA methylation at CG dinucleotides (mCG), mCH accumulates on Xi in domains with transcriptional activity, including the bodies of most genes that escape XCI and at the X-inactivation center, validating this epigenetic mark as a signature of transcriptional activity. Escape genes showing CH hypermethylation were the only genes with CG-hypomethylated promoters on Xi, a well-known mark of active transcription. Finally, we found extensive allele-specific mCH and mCG at autosomal imprinted regions, some with a negative correlation between methylation in the two contexts, further supporting their distinct functions. Our findings show that neuronal mCH functions independently of mCG and is a highly dynamic epigenomic correlate of allele-specific gene regulation.

DNA methylation | X-chromosome inactivation | imprinting | Non-CG | Bcor

In diploid mammals, the equivalence of the two parental alleles is violated by allele-specific epigenetic regulation in a small, but critical, subset of the genome. Genomic imprinting, or parent-of-origin–dependent gene regulation (1), is critical for embryonic development and plays a role in neuronal differentiation (2). In females, epigenetic inactivation of one X chromosome silences transcription of most genes to equalize gene expression with males (3). Both imprinting and X-chromosome inactivation (XCI) are critical to healthy brain development (4, 5). Despite the importance of allele-specific gene regulation in the brain, the epigenetic mechanisms controlling these patterns are not completely known, in part, due to the challenge of allele-specific epigenomic profiling. In particular, DNA methylation patterns can reflect allelic asymmetries in autosomal gene regulation (6), but their correlation with XCI has not been fully addressed.

XCI has unique advantages as a case study for the investigation of allele-specific epigenomic regulation. The inactivated allele is selected stochastically during early development and maintained through subsequent cell divisions (7), yielding a mosaic pattern of allelic expression in adult female tissues. Despite extensive inactivation of one X chromosome, some genes escape silencing and are expressed from the inactive X chromosome (Xi): ∼3% of X-linked genes in mice (8) and 15% in humans (9). Analysis of peripheral blood showed that XCI and escape from XCI are correlated with high or low levels of DNA methylation at CG dinucleotides (mCG) in promoter regions, respectively (10). However, different epigenetic profiles may be associated with XCI and escape from XCI in the brain because the DNA methylation landscape of neurons is distinct from other cell types. In particular, neurons accumulate methylation at millions of genomic cytosines in CA and CT dinucleotides during postnatal brain development beginning at 1 wk of age in mice (6, 11). This non-CG methylation correlates with reduced gene expression and inactivation of distal regulatory elements in a highly cell type-specific manner (12). Although the functional relevance of non-CG methylation (mCH) is unclear, it is bound by the transcriptional repressor methyl-CpG binding protein 2 (MeCP2) as neurons mature, and is enriched at genes that are up-regulated in Rett syndrome (13, 14).

Mosaic XCI prevents discrimination of methylation on the active X chromosome (Xa) and Xi alleles by conventional methylome profiling. We reasoned that understanding the allele-specific distribution of neuronal mCH in the context of X inactivation and imprinting could yield new insights into this unique aspect of the brain epigenome. Therefore, we profiled allele-specific DNA methylation, as well as transcription, in mouse frontal cortex using a *Xist* mouse mutant hybrid in which the paternal allele was deterministically inactivated in all cells (8). To assign sequencing reads to alleles, we used female

Significance

Mammalian cells contain two copies of the genome inherited from the two parents. Although most genes are expressed using both, a small but critical part of the genome has different levels of expression from each copy. These parts include the X chromosome in females and imprinted genes in both genders, which play key roles in brain development and cognition. We measured gene expression and DNA methylation, an epigenetic modification of the genome, in the brains of mice using a technique that allowed us to analyze the maternal and paternal copies of the genome separately. Our findings show that a brain-specific form of DNA methylation called non-CG methylation marks regions of active transcription within the inactive X chromosome.

Author contributions: C.L.K., C.M.D., J.R.E., and E.A.M. designed research; C.L.K., J.B.B., R.C., J.R.N., C.M.D., and E.A.M. performed research; C.L.K., J.B.B., C.M.D., and E.A.M. contributed new reagents/analytic tools; C.L.K. and E.A.M. analyzed data; and C.L.K. and E.A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.D.S. is a Guest Editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE83993).

[1]To whom correspondence should be addressed. Email: emukamel@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1611905114/-/DCSupplemental.

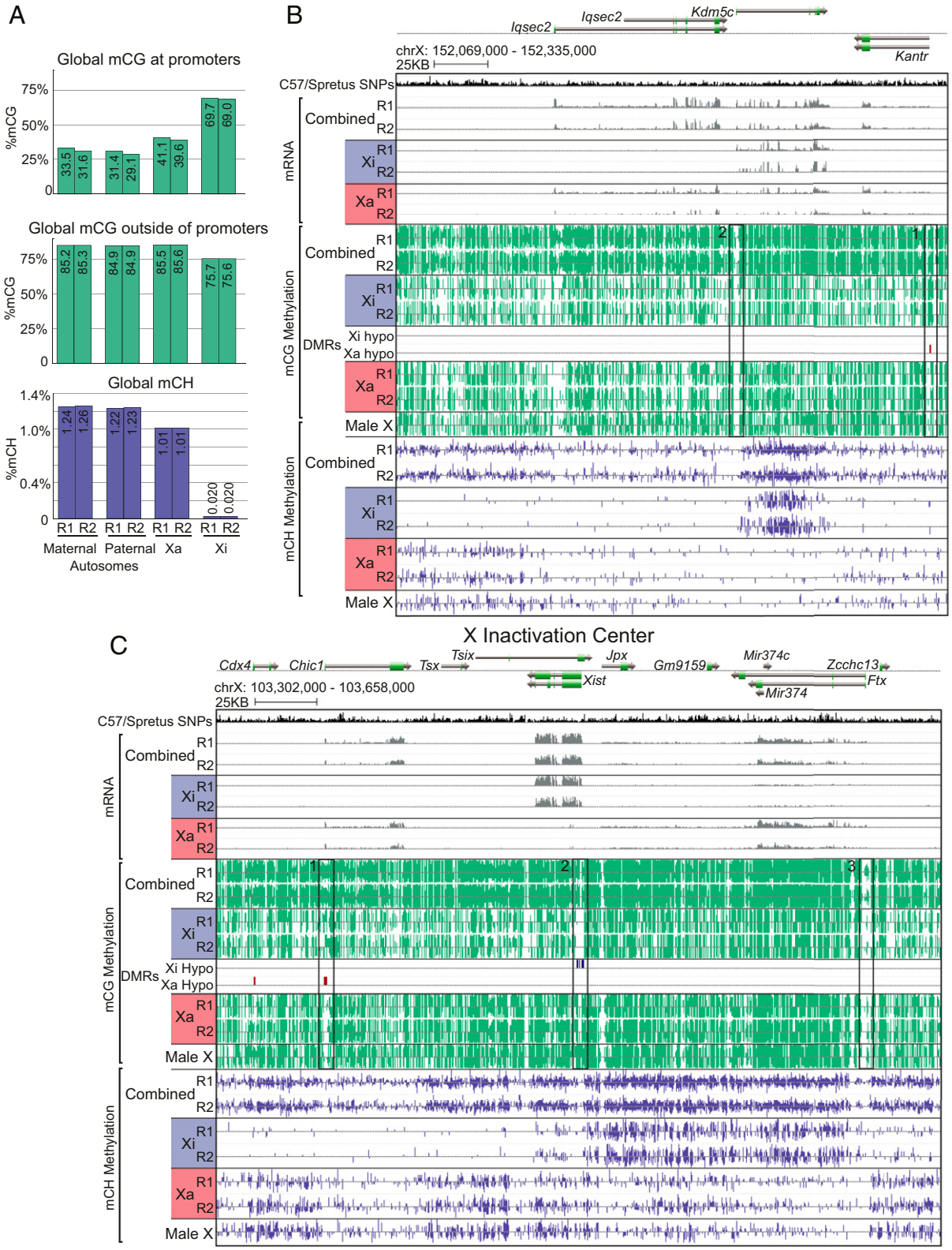

Fig. 1. Ultrasparse mCH on Xi correlates with escape domains. (*A*) Allele-specific mCG and mCH levels on autosomes and chromosome X. Browser view of methylation and expression for the *Kdm5c* locus (*B*) and the XIC (*C*). Ticks show the methylation level at individual cytosine positions (CG, green; CH, blue) on the forward (upward ticks) and reverse (downward ticks) strands. Combined tracks show both alleles, whereas the Xa and Xi tracks include only reads sorted using SNPs between C57 and *spretus*. Monoallelically expressed genes (*Iqsec2* and *Chic1*) and intergenic regions harbor mCH on Xa only, whereas diallelically expressed escape genes (*Kdm5c*) and the Xi-expressed noncoding RNA *Xist* contain dense mCH on Xi. Male X data are from 6-wk-old frontal cortex (11). chrX, chromosome X; R1, replicate 1; R2, replicate 2.

F1 mice from crosses between C57BL/6 *Xist* mutant and *Mus spretus* wild-type mice (8), and analyzed species-specific genetic variants (~42 million single-nucleotide polymorphisms (SNPs), including 1.95 million SNPs on the X chromosome). Our data reveal distinct allele-specific patterns of mCG and mCH at X-linked genes that reflect the accessibility of chromosomal domains during brain development. At autosomal imprinted regions, we found large domains of differential methylation that show a dissociation between mCG and mCH and point to independent regulation of these features of the neuronal epigenome.

Results

Allele-Specific Global Levels of CG and Non-CG DNA Methylation on Female X Chromosomes. Using female interspecific F1 transgenic mice with deterministic X inactivation (8, 15), we examined global levels of mCG and mCH on Xa and Xi in the adult frontal cortex. mCG is present throughout nonpromoter regions of the genome from the earliest stages of brain development, whereas mCH accumulates gradually during postnatal development starting at 1 wk of age in mice (11). We therefore reasoned that the silencing of Xi, established during random XCI in the embryonic inner cell mass (7), may block the accumulation of mCH on Xi. By contrast, nonpromoter mCG may be less affected by the chromosomal inactivation because it is established and epigenetically inherited from the early embryonic stage (16).

Consistent with this reasoning, we found significantly increased levels of mCG at promoters on Xi (69.4%) compared with Xa (40.4%; $P = 0.003$, paired t test) and autosomes (30–33%; $P = 0.006$) (Fig. 1*A*). By contrast, mCG at nonpromoter regions was lower on Xi (75.6%) compared with Xa (85.6%; $P = 0.013$) and autosomes (84–85%; $P = 0.012$). Strikingly, Xi is nearly devoid of mCH (0.02%) compared with Xa and autosomes (1.01%; $P < 0.001$). Xi thus resembles mCH deserts: large regions (median size of 1.8 Mbp) on autosomes that lack mCH, are transcriptionally silent, and are marked by inaccessible chromatin (11). The pattern of methylation is likewise less correlated across the two X alleles (mCG: $r = 0.27$, mCH: $r = -0.08$; correlation using 10-kb bins) compared with the autosomes (mCG: $r = 0.90$, mCH: $r = 0.91$). These results suggest that Xi is largely inaccessible to the de novo DNA methyltransferase *Dnmt3a*, which is responsible for establishing mCH in neuronal genomes starting in the second postnatal week (14).

Differential Methylation Patterns at Genes Known to Escape XCI. A subset of X-linked genes escapes from XCI, allowing expression from Xi. Comparison of male and female brain samples from both mice and humans showed a striking enrichment of mCH in females within the gene bodies of several escape genes (11, 17). Our allele-specific analyses show that this mCH signature of escape genes derives exclusively from the Xi. The pattern is exemplified by the allele-specific expression and methylation profiles of a known escape gene, *Kdm5c*, and two neighboring nonescape genes, *Iqsec2* and *Kantr* (Fig. 1*B*). By sorting reads based on the presence of SNPs that vary between C57 and *spretus* genomes (*Methods*), we identified sequencing reads originating from Xa and Xi for both expression and methylation data. As expected, *Iqsec2* and *Kantr* were monoallelically expressed from Xa, whereas *Kdm5c* escaped XCI and showed diallelic expression (Fig. 1*B*, mRNA tracks). These expression patterns correlated with a differentially methylated region (DMR) at the promoter of *Kantr* that is hypermethylated at CG sites (repressed) on Xi (Fig. 1*B*, box 1). In contrast, we observed CG hypomethylation on both Xi and Xa in the CpG island at the promoter of the escape gene *Kdm5c*, as expected (10) (Fig. 1*B*, box 2).

Gene body mCH has been associated with transcriptional repression in mammalian brain cell types (6, 11, 12). Consistent with this repressive association, mCH on Xa is highest in the intergenic region upstream of *Iqsec2* and relatively lower in gene bodies of expressed genes *Iqsec2*, *Kdm5c*, and *Kantr*. This pattern is similar to the distribution of mCH on the male X ($r = 0.94$). In contrast, mCH on Xi presents an opposite (positive) correlation with transcription: Xi is remarkably void of mCH except in the gene body of the escape gene, *Kdm5c*, where mCH is enriched (3.09%) and more abundant than on Xa (0.26%; $P = 0.001$).

Allele-specific mCH is also evident within the X-inactivation center (XIC), a 10- to 20-Mb region that controls the establishment and maintenance of XCI (7, 18). As expected, *Xist*, the long noncoding RNA (lncRNA) that triggers inactivation *in cis*, is monoallelically expressed from Xi; the escape gene, *Ftx*, is diallelically expressed; and *Chic1* is monoallelically expressed from Xa (Fig. 1*C*). Promoter mCG is consistent with this pattern of expression: *Chic1* is hypomethylated on Xa, *Xist* is hypomethylated on Xi, and *Ftx* is hypomethylated on both Xa and Xi (no DMR) (Fig. 1*C*, boxes 1–3). Gene body mCH is relatively high throughout the XIC on Xa, particularly in bodies of unexpressed genes (*Cdx4*, *Tsx*, *Tsix*, and *Xist*), and lowest in bodies of expressed genes (*Chic1* and *Ftx*). Once again, this pattern is very similar to male X ($r = 0.96$). By contrast, mCH on Xi is associated with transcriptional activity. It is enriched throughout a region upstream of *Xist* that includes the escape genes *Jpx* and *Ftx* and, to a lesser extent, within *Xist* itself.

At the XIC, allele-specific regulation of expression on the X chromosome is maintained, in part, through physical segregation of epigenetically defined chromatin regions called topologically associated domains (TADs) (19). We found that the start of the mCH-enriched region upstream of *Xist* aligns precisely with the boundary between two TADs identified by ~20-kb-resolution chromosome conformation capture (19). We further examined this correspondence throughout the ~5-Mb region surrounding *Xist* and found an additional block of enriched mCH on Xi (Fig. S1). This region coincides with the boundaries of a TAD comprising two

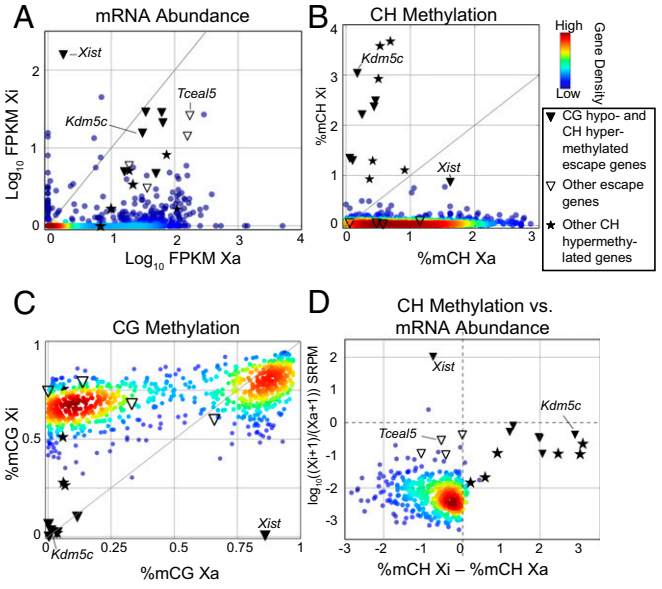

Fig. 2. Escape genes are marked by unique mCG and mCH patterns. (*A*) Allele-specific expression for each gene on chromosome X demonstrates nonrandom X inactivation and identifies escape genes. Each point corresponds to a gene, and the color of the point indicates the density of genes. Genes with a significant number of reads originating from Xi in both replicates are indicated as escape genes (Table 1). (*B*) mCH in gene bodies is virtually absent from Xi, except at a subset of escape genes, genes involved in X inactivation (*Jpx*), and a few additional loci. (*C*) Most gene promoters harbor dense mCG on Xi, whereas the bulk of escape genes have unmethylated promoters. A subset of escape genes (*Tceal5*, *Gpr34*, *Slc16a2*, and *Gpm6b*) violate this pattern and contain high levels of promoter mCG. (*D*) Comparison of gene body mCH (*x* axis) and expression (*y* axis) between Xa and Xi reveals the clustering of escape genes.

Table 1. Escape genes and CH-hypermethylated genes

Gene	Gene body, % mCH		Gene promoter, % mCG		Diallelic RNA abundance, FPKM	Allele-specific RNA abundance, SRPM	
	Xa	Xi	Xa	Xi		Xa	Xi
Escape genes with mCH on Xi							
5530601H04Rik	0.19	***2.40***	3.7	***2.1***	6.49	36.9	***12.3***
Ddx3x	0.17	***1.89***	0.2	***6.3***	44.8	84.9	***44.5***
Eif2s3x	0.10	***1.67***	3.4	***1.4***	22.5	30.5	***24.4***
Ftx	0.42	***2.47***	11.6	***10.1***	6.05	67.0	***7.65***
Kdm5c	0.28	***3.11***	1.2	***3.0***	7.40	57.9	***25.0***
Kdm6a	0.39	***2.22***	0.4	***0.02***	5.43	22.7	***8.25***
Xist	1.79	***0.79****	86.0	***0.5***	24.9	5.74	***546***
Firre	0.94	***1.16***	7.8	73.6	6.25	64.2	***1.91****
Pbdc1	0.52	***3.27***	3.7	***2.1***	4.40	17.2	***4.02****
Tmem29	0.43	***1.52***	6.1	27.8	10.2	28.8	***3.52****
Escape genes without mCH on Xi							
Gpm6b	0.57	0.04	13.7	79.1	121	252	***27.2***
Gpr34	1.15	0.04	65.9	59.8	3.62	13.9	***4.01***
Slc16a2	1.15	0.16	0.1	74.3	5.31	34.1	***3.91***
Tceal5	0.00	0.12	33.2	67.9	16.7	23.5	***9.93***
Other CH-hypermethylated genes							
Gm38020[†]	0.07	***3.85***	93.5	93.4	2.76	43.7	4.99
Jpx[†]	0.77	***3.46***	6.8	26.5	1.21	7.51	0.815
4933407K13Rik[‡]	0.25	***1.05***	5.9	51.2	0.565	8.40	0.10

Bold italic values indicate significant gene body mCH on Xi, significant CG promoter hypomethylation on Xi, or Xi RNA abundance (FDR < 0.05 in both replicates).

*Genes significant in one replicate only.

[†]Occurs within the XIC. *Gm38020* overlaps the escape gene *Ftx*, and *Jpx* has previously been reported to escape (42).

[‡]Located at the *Dxz4* macrosatellite, which is involved in Xi chromosome topology.

escape genes, *Pbdc1* and *5530601H04Rik* (19) (Fig. S1). This correspondence suggests that mCH accumulates within topologically defined domains of accessible chromatin (i.e., active TADs) on Xi.

These results demonstrate mCH at escape genes, and the XIC is positively correlated with expression from Xi, in contrast to the repressive association observed for both mCG and mCH on Xa and autosomes.

Differential mCH and mCG Between Xa and Xi Predict Escape Genes. To examine the relationship between escape genes and DNA methylation more broadly, we profiled our MethylC-Sequencing (Seq) and RNA-Seq data across all X-linked genes. The distribution of RNA, mCG, and mCH was dramatically different on Xi and Xa (Fig. S2). First, we identified escape genes in mouse frontal cortex using a binomial model that detects genes with a significant proportion of mRNA-Seq reads from Xi (*Methods*). In all, we found 11 genes that escaped inactivation (Fig. 2*A* and Table 1). Nine of these escape genes (*Xist*, *Ddx3x*, *Kdm6a*, *Kdm5c*, *Eif2s3x*, *5530601H04Rik*, *Ftx*, *Slc16a2*, and *Gpm6b*) are consistent with a previous survey in whole mouse brain samples (8). In addition, we detected diallelic expression of *Tceal5* and *Gpr34*, suggesting they may be novel escape genes in frontal cortex.

We then compared gene body mCH between Xa and Xi for all X-linked genes. On Xa, the median mCH level of gene bodies was 0.88% (range: 0.00–4.53%), and the pattern was similar to the male X ($r = 0.93$; Fig. S3). In contrast, mCH was statistically undetectable on Xi within the majority of gene bodies covered by our data (980 genes; Fig. 2*B*). In all, we identified 13 genes with statistically significant gene body mCH on Xi (Table 1). These genes included seven known escape genes (*Ddx3x*, *Xist*, *Eif2s3x*, *Kdm5c*, *Kdm6a*, *5530601H04Rik*, and *Ftx*), representing a statistically significant overlap ($P < 10^{-19}$, hypergeometric test). In addition, three genes with gene body mCH on Xi had significant expression from Xi in one of our two replicates and have been previously reported as escape genes in whole brain (*Firre* and *Pbdc1*) (8) or eye (*Tmem29*) (20). One other gene, *Jpx*, with mCH enrichment is located within the XIC and plays a direct role in XCI but is not significantly expressed from Xi. Finally, we identified significant mCH at the lncRNA *4933407K13Rik*, which is expressed from the macrosatellite locus *Dxz4*, a region that binds CTCF on Xi and plays a role in organizing the topology of Xi (21, 22).

In contrast to the enrichment of mCH on Xi at gene bodies of escape genes, we found strong depletion of mCG at the promoters of many of these genes (Fig. 2*C*) as previously observed for human escape genes in blood (10). Whereas the promoters of most X-linked genes are marked by increased mCG on Xi, seven of the escape genes showed significant hypomethylation on Xi. *Pbdc1* was also significantly hypomethylated on Xi, consistent with being an escape gene as previously reported (8). *Xist* was the only escape gene hypermethylated at its promoter on Xa, which is consistent with its silencing on Xa.

Integrating our findings for mCG, mCH, and gene expression (Fig. 2*D*), we observe the following pattern: Seven escape genes (*Ddx3x*, *Xist*, *Eif2s3x*, *Kdm5c*, *Kdm6a*, *5530601H04Rik*, and *Ftx*), and possibly *Pbdc1*, have distinct methylation patterns with hypomethylated CG promoters and hypermethylated CH gene bodies, and four other escape genes (*Slc16a2*, *Gpm6b*, *Gpr34*, and *Tceal5*) have Xi methylation patterns similar to nonescape genes with CG hypermethylation and CH hypomethylation. Considering prior surveys of escape from XCI (8), we observe that all CG-hypomethylated and CH-hypermethylated genes escape XCI across multiple tissues. *Gpm6b* was reported to escape XCI only in brain and lacks the unique DNA methylation signatures we observed at genes that ubiquitously escape XCI. Comparisons of DNA methylation and expression levels show that escape genes form a highly distinctive

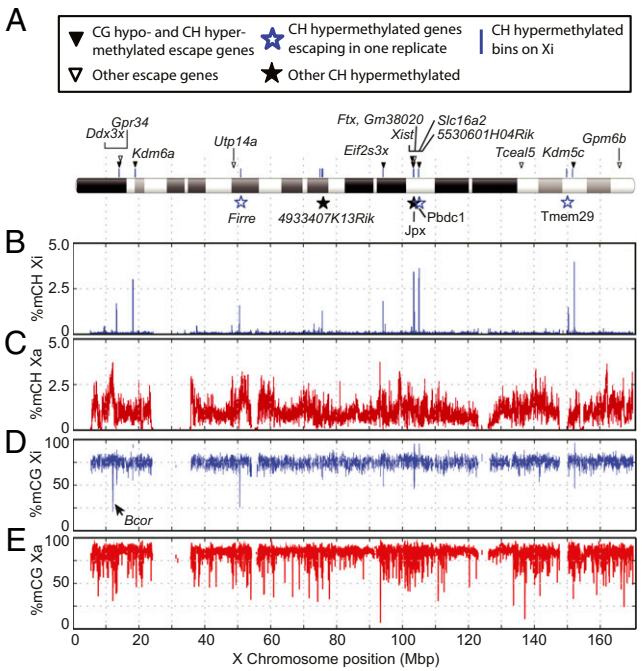

Fig. 3. X-chromosome–wide landscape of allele-specific DNA methylation. (*A*) Locations of genes escaping X inactivation, mCH-enriched regions (blue ticks), and mCH-enriched genes. (*B* and *C*) mCH is abundant across Xa but sparsely distributed on Xi. (*D* and *E*) mCG is abundant across both Xi and Xa but locally depleted at promoter regions of nonescape genes on Xa.

compartment in which a relative increase in mCH on Xi compared with Xa marks genes that escape X inactivation (Fig. 2*D*).

Analyses of Intergenic Regions. To include intergenic regions in our analysis, we next examined DNA methylation in 2.5-kb bins across the X chromosome. Fig. 3*A* shows the location of escape genes on the X chromosome (triangles), of significantly methylated bins on Xi (blue ticks), and of genes identified as CH-hypermethylated (stars), highlighting the chromosome-wide distribution of these genes. On Xi, mCH is absent throughout most intergenic regions, punctuated by 12 significant peaks of enriched mCH corresponding to the previously identified CH-methylated genes (Fig. 3*B*). In contrast, mCG is high throughout Xi, with a few exceptions corresponding to escape genes and the XIC (Fig. 3*D*). To quantify these patterns, we used MethylSeekR (23) to call CG unmethylated regions (CG-UMRs), which typically correspond to promoters of expressed genes. We identified 437 significant CG-UMRs on Xa but only 37 on Xi [false discovery rate (FDR) < 0.05, coverage by five or more reads in ≥30% of CG sites; Dataset S1]. Thirty-two CG-UMRs on Xi correspond to genes located in the XIC and to escape genes.

We noted that 14% of CG-UMRs on Xi fall within or proximal to *Bcor*, which does not appear to escape in brain based on our RNA-Seq analyses but was listed as an escape gene in a cell line in a previous study (8) (Fig. 3*D*, arrowhead and Fig. S4). *Bcor*, a gene in which mutations can lead to oculofaciocardiodental (OFCD) syndrome, was previously shown to be half as methylated in females (Xa + Xi) as in males in human blood and buccal tissue (24). This pattern runs counter to the pattern at other nonescape genes, where males show lower mCG compared with females, suggesting a unique pattern at *Bcor*. Our results reveal that *Bcor* contains multiple CG clusters hypomethylated on Xi on the paternal allele. This finding is consistent with the previous finding in humans, suggesting a conserved epigenomic pattern. Wamstad et al. (25) suggested that *Bcor* is unlikely to be a maternally expressed imprinted gene because the observed mother-to-daughter transmission of *Bcor* mutations in OFCD is not lethal. Here, we further reason that *Bcor* is unlikely to be imprinted to express only the paternal allele because we should not observe a phenotype in mother-to-daughter transmission if the maternal allele is not expressed. Therefore, the allele-specific methylation observed at *Bcor* is most likely specific to the activation state of the chromosome rather than the parental origin.

Allele-Specific Methylation and Imprinting. In addition to its role in XCI, allele-specific DNA methylation plays a key role in regulating autosomal imprinted regions. A previous study profiled DNA methylation using MethylC-Seq in brain samples from male Cast/129 F1 hybrid mice and identified imprinted autosomal methylation in both CG and CH contexts (6). The C57/*spretus* F1 female mice in our study have twice as many SNPs (41.7 million compared with around 20 million for Cast/129 hybrids). Although our *Xist* mutant mouse line could not be used to produce a reciprocal cross (i.e., maternal *spretus* × paternal C57) to distinguish species-of-origin vs. parent-of-origin effects, we nevertheless analyzed maternal vs. paternal differences in methylation at known imprinted regions to confirm and extend prior observations using a different genetic background. We compared maternal and paternal mCG levels at promoters of all autosomal genes (Fig. 4*A*). As expected, most genes have equal mCG levels on the two alleles. We then identified genes with allelic differences in mCG (allele1 > 75% and allele2 < 25%) and a significant DMR in the promoter. Our results recapitulate the imprinted loci previously identified (6). In addition, we found maternal mCG at the imprinted *Nnat* promoter, a gene Xie et al. (6) could not examine due to a lack of SNPs in their cross.

We next sought to connect allelic differences in methylation with expression. We found 77 autosomal genes that were differentially expressed between alleles in both replicates (FDR < 0.05 and $\log_2$-fold change > 1.5). Differential expression of these genes was significantly correlated with allelic differences in both promoter mCG ($r = -0.558$, $P <$ 1e-6) and gene body mCH ($r = -0.312$, $P = 0.0027$) (Fig. 4*B*). Focusing on DMRs previously reported to be imprinted in a parent-of-origin–dependent manner using reciprocal crosses (6), we found largely consistent mCG differences. Several imprinted DMRs identified in the Cast/129 F1 hybrids (*Casc1* intragenic, *6330408a02Rik* 3′ end, *FR149454* promoter, *FR085584* promoter, *Myo10* intragenic, *Vwde* promoter, and *Pvt1* promoter) fail to show allele-specific CG methylation in our data, suggesting they might not be conserved across mouse species (Fig. 4*C* and Dataset S2).

Whereas CG DMRs were localized to discrete regions (900-bp median size), we found substantial allele-specific differences in autosomal mCH that extended over much larger domains encompassing one or more gene bodies, as observed previously (6). Surprisingly, we found that allele-specific mCH could exhibit either the same asymmetry as allele-specific mCG or the reverse asymmetry. For example, the imprinting control region for the *Kcnq1* gene, located at the promoter of the antisense transcript *Kcnq1ot1*, is a 2.5-kb DMR with allele-specific mCG on the maternal allele. However, there is a much larger mCH DMR, spanning the entire *Kcnq1ot1* transcript (87.6 kb), that is also hypermethylated on the maternal allele (Fig. 4*D*). A reciprocal example is the maternally imprinted locus on chromosome 12 containing *Meg3*, *Rian*, and *Mirg*, where the paternal allele is marked by discrete allele-specific mCG and more diffuse mCH (Fig. 4*E*).

In contrast, we observed the reverse asymmetry (i.e., lower mCG on the paternal allele and lower mCH on the maternal allele) at the ∼3.6-MB region of chromosome 7 containing imprinted *Snrpn*, *Snurf*, and *Magel2* (Fig. 4*F*). Genetic variants in this region can cause Prader–Willi or Angelman syndrome, depending on which allele is affected. This locus contains a large CH DMR spanning ∼3.6 MB that is hypomethylated on the maternal allele, whereas CG hypomethylation is restricted to the paternal allele and occurs

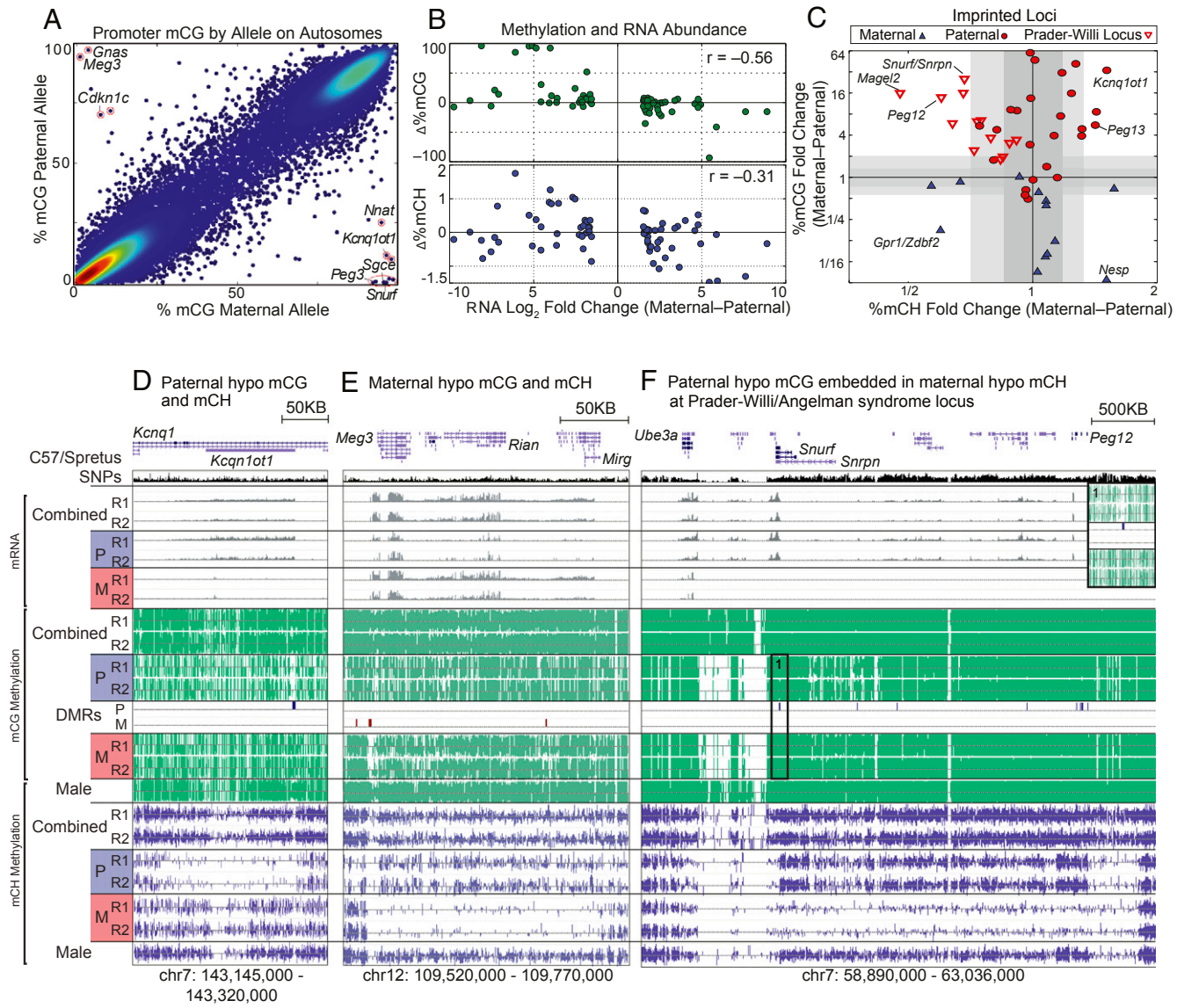

Fig. 4. Imprinted genes marked by allele-specific mCG and mCH. (*A*) Allelic differences in promoter mCG in autosomal genes recapitulate previously identified imprinted genes (6), indicated by red circles, including maternally imprinted *Peg13* and paternally imprinted *Gnas, Meg3*, and *Cdkn1c*, and suggest that *Nnat* has allele-specific methylation or is imprinted. (*B*) Eighty autosomal genes with allelic differential expression (>1.5 $\log_2$-fold change) are plotted against their allelic differences in mCG and mCH. (*C*) Differential CG methylation for imprinted DMRs identified by Xie et al. (6) is compared with differential mCH in the surrounding region (±10 kb). Shaded regions indicate 95% (dark shade) and 99% (light shade) confidence intervals for a null distribution obtained by comparing allelic differences in methylation across chromosomes in all autosomes in mCH using 20-kb bins (*x* axis) and in mCG using 1-kb bins (*y* axis). Browser views of imprinted loci *Kcnq1ot1* (*D*); *Meg3, Rian*, and *Mirg* (*E*); and *Snrpn*, *Snurf*, and *Magel2* (*F*). (*Inset*) Zoomed-in view of paternal hypo-mCG DMR at the promoter of *Snurf*. M, maternal; P, paternal.

mainly at the promoters of imprinted genes within the locus. Another example of a reverse asymmetry between allele-specific mCG and mCH occurs within the *Nesp*/*Gnas*/*Nespas* locus (Fig. S5). Together, these patterns of allele-specific autosomal mCG and mCH suggest a complex relationship between the two types of methylation, with both positive and negative correlations.

As with X inactivation, our analysis of methylation at imprinted autosomal loci reveals that mCG and mCH have contrasting allele-specific distributions indicating at least partly independent roles in gene regulation.

Discussion

Allele-specific regulation of domains of active and inactive chromatin is critical for healthy brain development in mammals, yet the landscape of DNA methylation within these domains has largely been studied without allele-specific resolution. Using MethylC-Seq and RNA-Seq in the frontal cortex of female transgenic mice with deterministic XCI, we obtained allele-specific, base-resolution DNA methylation and transcription profiles. In all, we identified 11 genes escaping XCI. Methylation profiling showed that the Xi chromosome was largely devoid of CH methylation, whereas most gene promoters showed CG hypermethylation. Seven escape genes (plus *Pbdc1*, a previously reported escape gene that escaped in one of our replicates) showed a pattern of CH hypermethylation in gene bodies and CG hypomethylation at promoters on Xi. Findings of hypo-mCG at promoters of escape genes are consistent with previous studies that analyzed CG methylation across multiple tissues in humans (10, 26). However, mCH had not been previously examined on Xi and Xa. Although the only genes with hypo-mCG on Xi were

also CH-hypermethylated, there were three additional genes with only hyper-mCH, suggesting distinct roles for the DNA methylation in these contexts.

mCH accumulates during postnatal development of frontal cortical neurons, reaching high levels in the adult mouse and human brain (6, 11). Indeed, the abundance of mCH is comparable to the abundance of mCG in adult neurons, and mCH is found in both excitatory and inhibitory neuron types (12). Our findings show that mCH is a high-fidelity epigenomic marker of allele-specific active chromatin domains, such as genes escaping X inactivation, which can be used for functional genomic annotation. However, the functions of mCH, if any, are unknown (27). Promoter mCG and gene body mCH are associated with transcriptional repression and are generally correlated. Our findings in genomic regions affected by XCI and parental imprinting demonstrate a partial dissociation between the CG and CH contexts of DNA methylation.

First, we identified hypo-mCG and hyper-mCH at a subset of escape genes. Most of these escape genes have been reported to escape across multiple tissues, so it is unclear if the presence of mCH, which is specific to the brain, is necessary for escaping inactivation or if it is a consequence of chromatin accessibility. Second, we identified three genes with hyper-mCH on Xi that lacked allele-specific CG hypomethylation. Because these cells are postmitotic and there is no known active mechanism for removing mCH, the presence of mCH at these regions may serve as a marker of previously active chromatin. Tissue type-specific or cell type-specific differential mCG in adult cells has been shown to reflect early developmental processes at so-called vestigial DMRs (12, 28). Third, we examined numerous imprinted genes and identified a complex and variable relationship between mCG and mCH. Although CG and CH sites are often hypomethylated on the same allele, as in the case of *Kcqn1ot1* and *Meg3*, we also identified a striking pattern in the Prader–Willi/Angelmann syndrome-associated region of chromosome 7, where allele-specific mCG and mCH were oppositely regulated.

A previous study that identified escape genes in whole brain reported 17 escape genes (8), and nine of the escape genes we found in frontal cortex overlap with these results. Our results demonstrate the presence of a DNA methylation signature at the large majority of escape genes. The absence of a methylation signature at other escape genes suggests there may be more than one pathway to escape from XCI. For example, *Gpm6b* has been previously reported to escape inactivation only in the brain. The characteristic escape gene hypo-mCG, which presumably occurs early in development in a precursor cell type, would be inconsistent with its escape only in brain. Therefore, there may exist a brain-specific mechanism to support the later escape of *Gpm6b*. In addition to tissue specificity, our results may suggest temporal dynamics to escape genes in the brain, as has been previously reported in mouse embryo (29). If mCH on Xi indeed marks active chromatin, then the nonescape genes with mCH may escape earlier in development and be downregulated in adults. Alternatively, it is possible these genes are expressed from Xi continuously over brain development at levels too low to achieve significance in our analysis.

Finally, our findings also contribute to the evidence of sex differences in the epigenome (30, 31). Sex chromosomes and XCI are genetic drivers of sex differences that, together with the effect of sex hormones, result in sexually dimorphic brain structure and cognition. Sex-specific DNA methylation can help us to understand how genes are regulated differentially between sexes and to develop efficacious treatments for disease in both sexes, an important objective set forth by the NIH (32). To this point, our analysis shows that DNA methylation on female Xa and male X is largely similar in both CG and CH contexts (Fig. S3). The differences we identified here are specific to Xi and support a role for escape genes in the development of sex differences, particularly in humans, where escape genes are far more numerous than in mice (33). The results may also shed light on the nonverbal learning disabilities and attention deficit hyperactivity disorder-like symptoms exhibited in Turner syndrome (45,X) females missing the Xi (4). If differences between typical females and those females with Turner syndrome are restricted to escape genes and the nonescape genes with hyper-mCH, then these genes likely play a critical role in healthy female brain development.

Materials and Methods

Mouse Model/Animals. *Xist* is an lncRNA that initiates inactivation *in cis*. Previous work has shown that deletion of a proximal A repeat inhibits *Xist* transcription and prevents inactivation (15). To profile allele-specific DNA methylation and transcription in a deterministic model of XCI, we used 14-wk-old female F1 progeny of C57BL/6 *Xist* mutant female mice and *M. spretus* wild-type male mice (The Jackson Laboratories) (8). Due to the deletion, the maternal X chromosome (C57) failed to inactivate and was the Xa in all cells, whereas the paternal chromosome (*spretus*) was ubiquitously inactivated (Xi). Furthermore, the genetic variability between the two mouse species allowed us to assign sequencing reads to the parent of origin. We collected samples from four biological replicates at 14 wk of age. Animals were weaned in groups of three to five per cage. Female F1 pups were genotyped at weaning to confirm the presence of the mutant allele.

Animals were anesthetized with CO_2, followed by cervical dislocation. Brains were removed and rinsed in cold PBS. For dissection, whole brains were placed in cold DMEM supplemented with 10% (vol/vol) FBS. The prefrontal cortex (PFC) was obtained by first removing the cerebellum, followed by slicing coronally 1 mm at the bregma and carefully isolating the frontal cortical tissue under a dissecting microscope. PFC samples were rapidly frozen on dry ice until processing. DNA and RNA were isolated from pooled PFC samples obtained from two individuals from separate litters. All protocols were approved by the University of Washington's Institutional Animal Care and Use Committee.

MethylC-Seq. Libraries were sequenced as single-end reads and prepared using the following procedure: Genomic DNA was extracted from ground, frozen tissue using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). Two micrograms of genomic DNA was spiked with 10 ng of unmethylated cl857Sam7 Lambda DNA (Promega). The DNA was fragmented with a Covaris S2 instrument to 150–200 bp, followed by end repair and addition of a 3′ A base. Cytosine-methylated adapters provided by Illumina were religated to the sonicated DNA at 16 °C for 16 h with T4 DNA ligase (New England Biolabs). Adapter-ligated DNA was isolated by two rounds of purification with AMPure X P beads (Beckman Coulter Genomics). Adapter-ligated DNA (≤450 ng) was subjected to sodium bisulfite conversion using the Methyl Code Kit (Life Technologies) as per the manufacturer's instructions. The bisulfite-converted, adapter-ligated DNA molecules were enriched by four cycles of PCR with the following reaction composition: 25 μL of Kapa Hi Fi Hotstart Uracil+Readymix (Kapa Biosystems) and 5 μL of TruSeq PCR Primer Mix (Illumina) (50 μL final). The thermocycling parameters were 95 °C for 2 min; 98 °C for 30 s; and then four cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 4 min, ending with one 72 °C 10-min step. The reaction products were purified using AMPure X P beads. Two separate PCR reactions were performed on subsets of the adapter-ligated, bisulfite-converted DNA, yielding two independent libraries from the same biological sample for subsequent sequencing using a HiSeq 2500 system (Illumina).

mRNA-Seq Library Preparation. Ribosomal RNA was removed from samples using a Ribo-Zero rRNA Removal Kit (Illumina). mRNA-Seq libraries were then generated using the TruSeq Stranded RNA LT Kit (Illumina) according to the manufacturer's instructions. Samples were sequenced using the HiSeq 2500 system.

Reference Genomes. The mm10 reference genome is the reference for the C57BL/6J strain. For *M. spretus*, we created a pseudo-reference genome by updating the mm10 reference with known C57-*spretus* SNPs as reported by the Sanger Institute (34) (www.sanger.ac.uk/science/data/mouse-genomes-project). We only retained high-confidence SNPs that passed all quality filters (denoted in the file as FI = 1), resulting in ∼1.95 million SNPs on chromosome X. Before allele sorting, our reads covered 89.1% of the genome. We were able to assign 68.6% of reads to one of the alleles, yielding broad and deep coverage for C57 (78.2% covered, 11.93× average read depth) and *spretus* alleles (70.1% covered, 9.79× average read depth). High coverage (at least five reads) was achieved at 67.0% of the genome in C57 and at 51.6% of the genome in *spretus*.

Mapping of MethylC-Seq Data. Sequencing reads were mapped separately to both the C57 and *spretus* reference genomes using Methylpy (11).

Unmethylated phage lambda DNA was spiked into each sequencing run, allowing us to estimate the bisulfite nonconversion rate directly (0.36% and 0.40% for the two replicates, respectively). Reads that mapped to one or both of the reference genomes were then pooled and assigned to the parent of origin, corresponding to Xa (C57) and Xi (*spretus*). Only reads containing one or more SNPs that matched 100% to one parental reference or the other were retained and used in the analysis. We noted that a small proportion (~2.5%) of CH sites were covered by reads that contained sequence mismatches potentially consistent with a CG position. To prevent contamination from these ambiguous sites, we excluded them from our analysis.

Mapping of mRNA-Seq Data. The mRNA-Seq data were mapped as previously reported (8). The mRNA-Seq reads were mapped separately to both the C57 and *spretus* reference genomes using TopHat2 (35) with default parameters. The transcriptome included only exons as defined in the GENCODE release M7 (level 3) comprehensive gene annotation file (www.gencodegenes.org/). High-quality reads (mapping quality score MAPQ $\geq$ 30) that mapped to one or both of the reference genomes were then pooled and assigned to the parent of origin, corresponding to Xa (C57) and Xi (*spretus*). Only reads containing SNPs that matched 100% to one parental reference or the other were retained for analysis.

Data Analysis. Browser representations were created using Anno-J (www.annoj.org) (36). Pearson correlations were used except where stated otherwise. *P* values were <0.01 unless otherwise stated.

MethylC-Seq Analysis. Methylation was analyzed separately for the CG and CH (i.e., CA, CC, CT) contexts. We examined mCG at promoters and mCH in gene bodies, both of which correlate with transcriptional repression in the brain (6, 11, 12). Gene transcription start and end sites were taken from GENCODE release M7 (level 3), and promoters were defined as ±1,000 bases from the transcription start site. Methylation was quantified as the number of methylated cytosine base calls (*m*) divided by total cytosine base calls (*c*), and was corrected for the nonconversion rate (*NCR*; calibrated using spike-in lambda DNA) using the maximum likelihood formula:

$$mC = g\left[\frac{m/c - NCR}{1 - NCR}\right],$$

$$g[x] = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}.$$

Importantly, MethylC-Seq cannot distinguish between methylcytosine and hydroxymethylcytosine, which is present at significant levels in brain in the CG context (11, 37). However, prior studies using Tet-assisted bisulfite sequencing have shown that there is no detectable hydroxymethylation at non-CG sites (11, 38). Therefore, although differences in CG methylation could be driven by changes in one or more types of methylation, our analyses regarding CH methylation are not affected by this ambiguity.

mRNA-Seq Analysis. Identification of escape genes using mRNA-Seq adhered to previously published methods (8). First, reads that mapped to C57, *spretus*, or both reference genomes were aggregated and used to quantify diploid expression in fragments per kilobase of transcript per million mapped reads (FPKM) with Cufflinks (39). Next, reads were assigned to the Xa or Xi only if all SNPs within a read corresponded to either the C57 (Xa) or *spretus* (Xi) reference. Reads that did not meet these criteria or that contained no SNPs were discarded. We then quantified haploid expression as allele-specific reads per 10 million mapped reads (SRPM). Finally, a binomial model was used to compute a confidence interval for the expression of each gene on Xi (8). A gene was said to escape inactivation significantly if diploid expression FPKMs were ≥1, Xi-SRPM was ≥1, and the lower bound of the 99% confidence interval from the binomial model was >0.

Definition of CH-Hypermethylated and CG-Hypomethylated Genes. We modeled the methylation of genes on Xi using a mixture distribution that we fit using an iterative procedure. We first fit a beta-binomial distribution for the apparent CH methylation levels of all gene bodies on the inactive Xi by maximum likelihood. We then used this beta-binomial distribution to compute a *P* value for each gene (i.e., the likelihood of observing that gene's mCH level, given the null distribution) and marked any genes with significantly greater mCH (FDR < 0.05 using the Benjamini–Yekutieli correction) as "hyper-mCH" genes. We then repeated our fitting procedure using only genes that were not marked as hyper-mCH. This procedure was repeated until it converged on a set of hyper-mCH genes. Only genes with significant mCH in both replicates were reported in our results. The same analysis was applied on CG methylation in promoters to identify significantly CG-hypomethylated genes on Xi.

Additional Datasets. MethylC-Seq data for 6-wk-old male mouse frontal cortex and fetal brain tissue (embryonic day 13.5, mixed male and female) were from a previously published study (11). These datasets were mapped to mm10 and processed using the same methods described above.

Data Access and Browser. Data are accessible in the Gene Expression Omnibus (GEO) database (accession no. GSE83993). Data are also displayed via a web-based browser at brainome.ucsd.edu/mm_xist_hybrid.

ACKNOWLEDGMENTS. We thank M. Margarita Behrens, Yupeng He, and Miles Wilkinson for helpful discussions. This work was supported by NIH Grants R00NS080911 (to E.A.M.), MH105768 (to J.B.B.), and GM046883 (to C.M.D.). The work in J.R.E.'s laboratory was supported by the Howard Hughes Medical Institute. J.R.E. is an Investigator of the Howard Hughes Medical Institute.

1. Reik W, Walter J (2001) Genomic imprinting: Parental influence on the genome. *Nat Rev Genet* 2(1):21–32.
2. McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37(1):179–183.
3. Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* 190:372–373.
4. Rovet JF (1993) The psychoeducational characteristics of children with Turner syndrome. *J Learn Disabil* 26(5):333–341.
5. Buiting K, et al. (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat Genet* 9(4):395–400.
6. Xie W, et al. (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148(4):816–831.
7. Augui S, Nora EP, Heard E (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* 12(6):429–442.
8. Berletch JB, et al. (2015) Escape from X inactivation varies in mouse tissues. *PLoS Genet* 11(3):e1005079.
9. Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434(7031):400–404.
10. Sharp AJ, et al. (2011) DNA methylation profiles of human active and inactive X chromosomes. *Genome Res* 21(10):1592–1600.
11. Lister R, et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341(6146):1237905.
12. Mo A, et al. (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 86(6):1369–1384.
13. Chen L, et al. (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc Natl Acad Sci USA* 112(17):5509–5514.
14. Gabel HW, et al. (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* 522(7554):89–93.
15. Hoki Y, et al. (2009) A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* 136(1):139–146.
16. Wang L, et al. (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157(4):979–991.
17. Schultz MD, et al. (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523(7559):212–216.
18. Lee JT, Strauss WM, Dausman JA, Jaenisch R (1996) A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* 86(1):83–94.
19. Nora EP, et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485(7398):381–385.
20. Reinius B, et al. (2012) Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome. *BMC Genomics* 13(1):607.
21. Horakova AH, et al. (2012) The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol* 13(8):R70.
22. Deng X, et al. (2015) Bipartite structure of the inactive mouse X chromosome. *Genome Biol* 16:152.
23. Burger L, Gaidatzis D, Schübeler D, Stadler MB (2013) Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res* 41(16):e155.
24. Joo JE, et al. (2014) Human active X-specific DNA methylation events showing stability across time and tissues. *Eur J Hum Genet* 22(12):1376–1381.
25. Wamstad JA, Corcoran CM, Keating AM, Bardwell VJ (2008) Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development. *PLoS One* 3(7):e2814.
26. Cotton AM, et al. (2015) Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet* 24(6):1528–1539.

27. Kinde B, Gabel HW, Gilbert CS, Griffith EC, Greenberg ME (2015) Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci USA* 112(22):6800–6806.
28. Hon GC, et al. (2013) Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* 45(10):1198–1206.
29. Lingenfelter PA, et al. (1998) Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nat Genet* 18(3):212–213.
30. Nugent BM, et al. (2015) Brain feminization requires active repression of masculinization via DNA methylation. *Nat Neurosci* 18(5):690–697.
31. McCarthy MM, et al. (2009) The epigenetics of sex differences in the brain. *J Neurosci* 29(41):12815–12823.
32. Clayton JA, Collins FS (2014) Policy: NIH to balance sex in cell and animal studies. *Nature* 509(7500):282–283.
33. Yang F, Babak T, Shendure J, Disteche CM (2010) Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* 20(5):614–622.
34. Keane TM, et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.
35. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
36. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322.
37. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324(5929):929–930.
38. Yu M, et al. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149(6):1368–1380.
39. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
40. Karolchik D, et al. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32(Database issue):D493–D496.
41. Williamson CM, Skinner JA, Kelsey G, Peters J (2002) Alternative non-coding splice variants of Nespas, an imprinted gene antisense to Nesp in the Gnas imprinting cluster. *Mamm Genome* 13(2):74–79.
42. Tian D, Sun S, Lee JT (2010) The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143(3):390–403.