Supplemental Methods

Opossum lines and crosses

Two random-bred stocks of *Monodelphis domestica*, designated LL1 and LL2 (derived from the Population 1 and Population 2 stocks described by (VandeBerg and Williams-Blangero 2010) were used to generate reciprocal F_1 embryos (Supplemental Figures S1-S3 and Table S1). The LL1 stock was derived as a subgroup from Population 1 ancestors, which is the same ancestral stock that furnished DNA for the opossum genome sequence (Mikkelsen et al. 2007); LL2 was derived by admixture of Population 1 and Population 2 animals and comprises an approximate 1:7 mixture of the two genetic backgrounds, respectively (John L. VandeBerg, personal communication). Embryos from the parental crosses of LL1 x LL1 and LL2 x LL2 were also collected to infer the direction of allelic transmission. To control for segregating polymorphisms in the two lines, the same males were used for the F_1 and parental crosses, and the females in F_1 and parental crosses were full siblings (Supplemental Figures S2). To collect prenatal stage animals of known gestational ages, the time of copulation was determined by videotaping paired animals with minor modifications of (Mate et al. 1994). All procedures involving opossums were approved by the Institutional Animal Care and Use Committee of Texas A&M University, College Station (TAMU Animal Use Protocols 2011-141 and 2011-191).

Tissue selection, dissection, and sex typing

Based on video evidence of mating, 13 days post-copulation (d.p.c.) females were euthanized and the fetuses and respective extra-embryonic membranes (EEM) were collected by dissection and placed in either RNAlater (Ambion) or phosphate buffered saline (PBS) and stored at -80 ^{o}C. Head/brain and EEM tissues were dissected from each fetus (Figure 1A, B). At this early stage of development, brain size was very small; and despite identical gestation times,

fetuses within a litter varied in size. For smaller fetuses, the head was further dissected to remove jaws, muzzle, and other non-cranial structures. The remaining cranial region and the corresponding EEM tissues were used for RNA extraction, whereas the limbs were used for genomic DNA extraction. For the two largest fetuses (sample_ID: A0571_b4 and A0579_b4), we were able to identify and isolate the frontal cortices and other brain structures. To avoid maternal cell contamination during the collection of EEM, we followed the umbilical cord to the uterine/placenta interface and retreated the scissors ~1-2 mm before cutting. The fetuses were sex-typed using Y-chromosome-specific primers (Supplemental Figures S3: developed independently by the Opossum Y Chromosome Mapping Project (see Acknowledgements) using a Y-specific BAC sequence containing the opossum sex-determining region on the Y (SRY) gene (AC239615.3, JN086997.1). Two female fetuses from each reciprocal cross, and one male and one female fetus from each stock-specific cross (total of eight individuals) were used for RNA extraction.

Total RNA and genomic DNA extractions, and QC

After collection, the tissues were homogenized in TriReagent (Invitrogen) and total RNA was extracted using BCP (1-bromo-2 chloropropane), precipitated with isopropanol, and resuspended in RNase-free water. Potential DNA contamination was removed by both DNase I treatment and Qiagen RNeasy Plus Mini kit (Qiagen, CA). RNA concentrations and A260nm/A280nm ratios were checked with a NanoDrop ND-1000 Spectrophotometer. The RNA quality was validated on both 1% agarose gels and an Agilent 2100 Bioanalyzer. To extract genomic DNA, tissue was minced and digested with proteinase K overnight followed by phenol:chloroform extraction, ethanol precipitation overnight, and elution in TE.

Illumina mRNA-seq and sequence alignment

mRNA-seq libraries were made from the brain and EEM RNA samples (1-3 μg total RNA input) of the four individuals in reciprocal F_1 crosses and four individuals in the parental crosses described above, using the Illumina TruSeq RNA Sample Prep Kit (Illumina Inc., CA). The 16 libraries (eight from the F1 crosses and eight from the parental crosses) were multiplexed and run on four 51 bp single-end lanes each, on an Illumina HiSeq 2000 instrument (Illumina Inc., CA). Image analysis and base calling were performed using Illumina software. In total, 1.5 billion short reads (76.5 billion bps) were generated from the 16 samples (Supplemental Table S2). The q-score and nucleotide distribution QC indicated good RNA-seq data quality. The reads were aligned to the opossum reference genome assembly (monDom5, http://genome.ucsc.edu/) using TopHat v1.4.1 (Trapnell et al. 2009) with three mismatches allowed. Eighty-two percent of the reads were uniquely mapped to the opossum reference genome (Supplemental Table S2).

Quantification of X-linked gene total and parent-of-origin allele-specific expression from RNA-seq data

The metric for total expression level, FPKM (Fragments Per Kilobase pair of exon model per Million fragments mapped), was calculated for all samples using Cufflinks v1.3.0 (Trapnell et al. 2010) based on all mapped reads from the TopHat alignments. The multiple mapped reads were weighted using the "-u" parameter in Cufflinks. The expression level was normalized across brain and EEM samples using quartile normalization. The RNA-seq read coverage and the mapping percentages were homogeneous across all the samples (Supplemental Table S2). We covered 11,465 Ensembl opossum gene models with FPKM≥1 in all eight brain samples and 10,518 gene models in the eight EEM samples. The RNA-seq data *were deposited* in the Gene Expression Omnibus (GEO) database under accession number GSE45211.

SNP positions were called in combined RNA-seq data only from reads that mapped uniquely to the opossum reference sequence with ≥40 matching reads, using SAMtools software (Li et al. 2009). Problematic SNPs, such as those with a third allele, near an indel position, or at the exon-intron junctions, were removed. In total, 68,000 SNPs were called in the transcriptome datasets. The reference and alternative allele counts were summarized at high quality X-linked SNP positions. We further selected high-coverage SNPs with ≥8 X coverage in at least one of two female F_1 individuals in both reciprocal crosses. Nineteen additional X-linked genes assigned to MonDom5 ChrUn were also included (Supplemental Table S3). Retrotransposed X-linked genes were also excluded from the analysis (Supplemental Table S4). To quantify allele-specific expression in brain and EEM tissues from the reciprocal crosses, we calculated the ratio of the number of reference allele-containing reads divided by the total coverage at each identified SNP position (Wang et al. 2008) (Supplemental Data S1). The transmission directions were inferred from the parental crosses and supplemented using additional information from other LL1 individuals for which RNA-seq data was available (data not shown).

X-linked SNP genotyping by Sanger sequencing

To confirm the parental origins of the two alleles at escaper loci, we genotyped the F_1 individuals and their parents at informative SNP positions using Sanger sequencing (Supplemental Data S2A). Primers targeting informative SNPs were designed using Primer3 (http://frodo.wi.mit.edu/). DNA was PCR amplified using TaqGold® polymerase, purified, and Sanger sequenced at Beckman Coulter Genomics (Danvers, MA). Gel purification was necessary for some samples due to the presence of non-specific PCR products. Sequences were viewed, aligned, and analyzed using Sequencher 4.10®. For non-escaper genes, in which

expression of only a single allele was observed, the possibility that any particular informative SNP might not be heterozygous in all F_1 individuals had to be considered. This was necessary because the LL1 and LL2 lines are not 100% inbred and some alleles are shared between them at segregating loci. To check whether the SNPs were heterozygous in the female F_1 individuals, we classified the informative SNPs into six classes (Supplemental Table S5), and randomly selected 20 SNPs (one SNP per gene) for genotyping by Sanger sequencing. All 20 of these were verified as heterozygous in at least two of the four female samples (Supplemental Data S2B).

Validation of X-linked escaper and non-escaper gene expression by allele-specific pyrosequencing

Fifteen of twenty-four escaper genes possessed more than one informative SNP in the RNA-seq dataset. Judged from the abundances of RNA-seq reads containing these linked SNPs, the paternal allelic expression levels are consistent for SNP sites within a gene (Supplemental Data S1A and S1B). This agreement between multiple SNPs within the same gene provided an internal validation for the allele-specific expression quantification. To verify the escaping status of these genes and obtain an estimation of the paternal allelic expression using an independent method, we performed allele-specific pyrosequencing on all 24 escaper genes (Supplemental Figure S6-S28), one non-escaper gene (*HPRT1*) (Supplemental Figure S4), and one autosomal control gene (*GPM6B*) (Supplemental Figure S5). Pyrosequencing PCR and sequencing primers were designed to target informative exonic SNP positions within selected genes using PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA). To eliminate potential biased amplification, all primers were checked to guarantee that they did not overlap SNP positions shared between the LL1 and LL2 parents. Pyrosequencing PCR amplification was carried out in

40 µl system using Ampli-Taq Gold polymerase (Life Technologies) under the following cycling conditions: 1 cycle of 95^{o} C for 5 min, 45 cycles of 95^{o} C-45 sec, 50-57^{o} C-30 sec, 72^{o} C-20 sec, followed by 1 cycle of 72^{o} C for 10 min. PCR products were prepared according to the manufacturer's protocol and then loaded on the PSQ 96MA Pyrosequencer (Qiagen, CA) with the PyroMark Gold Reagents (Qiagen, CA) using the Allele Quantification method (AQ). Two technical replicates were done for each gene in each sample.

Native ChIP-seq and data analysis

Native-ChIP was conducted on a primary fibroblast cell line (derived from adult ear pinna), fetal brain, and EEM using a method modified from (Dindot et al. 2009). Briefly, total tissue samples of fetal brain and EEM were washed in PBS and homogenized in 500 µl of Buffer I (0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15mM Tris, 0.5 mM DTT, 0.1 mM PMSF) using a battery-powered homogenizer. The sample was centrifuged for 5 min. at 3000g, the supernatant was removed, and the pellet was re-suspended in 200 µl of Buffer I. Cells were lysed for 5 minutes on ice by adding 200 µl of Buffer II (Buffer I + 4 µl of NP40), and nuclei were isolated by centrifugation of lysed cells for 20 min at 10,000Xg through 1.5 ml of Buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris, 0.5 mM DTT, 0.1 mM PMSF). The pellet was washed with Buffer I, centrifuged, and re-suspended in 350 µl of MNase (micrococcal nuclease) Digestion Buffer (0.32 sucrose, 4 mM $MgCl_2$, 50 mM Tris, 0.1 mM PMSF). Chromatin was digested using 10 units of MNase (Sigma, N5386) for 10 min at 37°C. The reaction was stopped using 50 µl of 0.5 M EDTA. For input control, 100 µl of digested chromatin were removed and stored at -20° C.

Equal aliquots of the remaining digested chromatin (EEM = 2.0 µg/rxn; fetal brain = 11 µg/rxn) were incubated at 4° C overnight with anti-H3K4me3 (Millipore #07-473), anti-H3K9Ac (Millipore #CS200583), anti-H3K9me3 (Millipore #07-442), anti-H3K27me3 (Millipore #07-449), or non-specific, rabbit IgG (Millipore #12-370). Antibody-bound chromatin was isolated using Dynabeads® Protein A (Invitrogen), washed, and eluted. N-ChIP and input DNA were purified using the Qiagen MiniElute Spin Columns (Qiagen) and enrichment was verified using real-time PCR. Illumina libraries were constructed at Global Biologics, LLC, and sequenced at the University of Missouri – Columbia DNA Core Facility (Columbia, MO) and Genomics Resources Core Facility at Weill Cornell Medical College (New York, NY). Raw reads were quality filtered, trimmed, and aligned using Bowtie in the Galaxy suite (Giardine et al. 2005; Blankenberg et al. 2010; Goecks et al. 2010). Aligned reads were visualized on the UCSC genome browser (Kent et al. 2002) and IGV (Robinson et al. 2011; Thorvaldsdottir et al. 2012) and significant peaks were called using the MACS algorithm (Zhang et al. 2008) (Table S3). The ChIP-seq data *were deposited* in the Gene Expression Omnibus (GEO) database under accession number GSE45186.

Re-annotation of X-linked promoters

Due to the incompleteness of the annotation of the *Monodelphis* genome, it was necessary to re-annotate the X-linked genes to ensure that the 5' exons, UTRs and putative promoters were correctly annotated. We used predicted RNA structure from TopHat, the presence of CpG islands (both currently and newly annotated), and the presence of H3K4me3 peak to annotate new 5' exons and putative promoters for 312 X-linked genes.

Bisulfite-sequencing of promoter DNA

Two µg of genomic DNA were treated with bisulfite using the EpiTech Bisulfite Kit from Qiagen, Inc. Treated DNA was PCR amplified using primers designed by Methyl Primer Express v 1.0 (Applied Biosystems). One µl of bisulfite treated gDNA was used in PCR amplification in 50 µl reaction using Ampli-Taq Gold polymerase (Life Technologies) under the following cycling conditions: 1 cycle of 95^{o} C for 5 min, 35 cycles of 95^{o} C-15 sec, 50 or 55^{o} C-30 sec, 72^{o} C-20 sec, followed by 1 cycle of 72^{o} C for 10 min. PCR products were cloned using the TopoTA Cloning® Kit (Life Technologies). At least 16 transformed colonies were selected and Sanger sequenced using the M13 forward primer at Beckman Coulter Genomics (Danvers, MA). Sequences were viewed, aligned, and analyzed using Sequencher 4.10.

Quantification of DNA methylation percentage using PyroMark assays

Quantification of methylation percentages in individual consecutive CpG sites was achieved with high reproducibility by pyrosequencing of bisulfite-treated DNA using the PyroMark Assay method. Bisulfite conversion was carried out on 500 ng genomic DNA of fetal brain and EEM samples in both sexes with the Qiagen EpiTect Bisulfite Kit (Qiagen, CA). PyroMark primers were designed to target the CpG islands with PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA). PCR products were prepared, run and analyzed on the PSQ 96MA Pyrosequencer (Qiagen, CA) with PyroMark CpG software 1.0.11. Background subtraction was done using the "control peak heights" feature. Each sample was repeated twice and the average was used for the analysis.

References

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 19**: Unit 19 10 11-21.

Dindot SV, Person R, Strivens M, Garcia R, Beaudet AL. 2009. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* **19**: 1374-1383.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**: 1451-1455.

Goecks J, Nekrutenko A, Taylor J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**: R86.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Mate KE, Robinson ES, Vandeberg JL, Pedersen RA. 1994. Timetable of in vivo embryonic development in the grey short-tailed opossum (Monodelphis domestica). *Molec Reprod Dev* **39**: 365-374.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A et al. 2007. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447**: 167-177.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24-26.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178-192.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511-515.

VandeBerg JL, Williams-Blangero S. 2010. The laboratory opossum (*Monodelphis domestica*). Chapter 19. pp 246-261. In *UFAW Handbook on the Care and Management of Laboratory and Other Research Animals 8th edition*, (ed. R Hubrecht, J Kirkwood). John Wiley & Sons, Ltd., Chichester.

Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG. 2008. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS ONE* **3**: e3839.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.