Maternal Bias and Escape from X Chromosome Imprinting in the Midgestation Mouse Placenta

Elizabeth H Finn[*], **Cheryl L Smith**[*,§], **Jesse Rodriguez**[†,‡], **Arend Sidow**[*,§], and **Julie C Baker**[*]

[*]Department of Genetics, Stanford University School of Medicine, Stanford CA, 94305

[§]Department of Pathology, Stanford University School of Medicine, Stanford CA, 94305

[†]Biomedical Informatics Program, Stanford University School of Medicine, Stanford CA, 94035

[‡]Computer Science Department, Stanford University, Stanford CA, 94305

Abstract

To investigate the epigenetic landscape at the interface between mother and fetus, we provide a comprehensive analysis of parent-of-origin bias in the mouse placenta. Using F1 interspecies hybrids between *mus musculus* (C57BL/6J) and *mus musculus castaneus*, we sequenced RNA from 23 individual midgestation placentas, five late stage placentas, and two yolk sac samples and then used SNPs to determine whether transcripts were preferentially generated from the maternal or paternal allele. In the placenta, we find 103 genes that show significant and reproducible parent-of-origin bias, of which 78 are novel candidates. Most (96%) show a strong maternal bias which we demonstrate, via multiple mathematical models, pyrosequencing, and FISH, is not due to maternal decidual contamination. Analysis of the X chromosome also reveals paternal expression of *Xist* and several genes that escape inactivation, most significantly *Alas2, Fhl1*, and *Slc38a5*. Finally, sequencing individual placentas allowed us to reveal notable expression similarity between littermates. In all, we observe a striking preference for maternal transcription in the midgestation mouse placenta and a dynamic imprinting landscape in extraembryonic tissues, reflecting the complex nature of epigenetic pathways in the placenta.

Keywords

placenta; X-inactivation; imprinting; parent-of-origin bias; RNA sequencing

INTRODUCTION

In mammals, while most genes are transcribed equally from the maternal and paternal allele, over 100 genes have been identified which have a strong parent-of-origin bias. This bias,

Corresponding author: Julie C Baker, Associate Professor, Stanford University, 300 Pasteur Drive Room M337, MC5120, Stanford, CA 94305.

called genomic imprinting, was initially discovered through genetic complementation studies (Searle and Beechey, 1978), in which crossing a male with a deletion and a female with a duplication in the same locus failed to rescue the deficiency. This suggested that healthy development requires specific contributions from both the maternal and paternal genomes. Imprinting was further confirmed by the analysis of parthenogenetic and androgenetic embryos. Parthenogenones, with two maternal genomes, show substantial undergrowth of extraembryonic tissues (Surani and Barton, 1983; Surani et al., 1984). Conversely, androgenones, with two paternal genomes, have mostly normal placental tissue (Barton et al., 1984). These data lead to a hypothesis that the paternal genome is responsible for the development of extraembryonic tissues, while the maternal genome limits their growth.

The distinctive patterns of imprinting in the murine placenta are epitomized by the placenta-specific silencing of the entire paternal X chromosome. Mammals silence one copy of the X chromosome in females in order to ensure equivalent gene expression between XX females and XY males (Lyon, 1961). This silencing is controlled by *Xist* - an RNA transcribed exclusively from the inactive chromosome, which leads to its repression (Brockdorff et al., 1991; Brown et al., 1991; Johnston et al., 1998). *Xist* is expressed from the paternal allele in cleavage-stage embryos, silencing the paternal X chromosome in early embryos and in the placenta (Brockdorff et al., 1991; Brown et al., 1991; Harper et al., 1982; Monk and Harper, 1979; Takagi and Sasaki, 1975; West et al., 1977). Maternal *Xist* is reactivated in the inner cell mass, and during tissue development, stochastic variation causes one allele to take over and one chromosome to be silenced. In contrast, this reactivation is absent in the placenta and the maternal X chromosome continues to be expressed. The story of the X chromosome highlights that imprinting in the placenta occurs at different regions, and perhaps through different mechanisms, than imprinting in other cell lineages.

The exact mechanisms underlying the placenta's unique epigenetic state, including variation in imprinting and X inactivation, are unknown. However, it is likely that chromatin modifications play a role. Overall, the placenta appears to be depleted for many of the repressive marks seen in embryonic tissues. Placental DNA is distinctly hypomethylated (Razin et al., 1984) and trophoblast cells can survive without any DNA methylation (Sakaue et al., 2010). It has been shown at two regions on chromosome seven that histone marks, rather than DNA methylation, control allele-specific silencing at imprinted regions in the placenta (Lewis et al., 2004; Umlauf et al., 2004). However, even these histone marks are relatively low in the placenta: in particular, H3K27me3 and arginine methylation (Chuong et al., 2013; Rugg-Gunn et al., 2010; Torres-Padilla et al., 2007). Further, the placenta is tolerant of polyploidy, as is demonstrable by tetraploid complementation, in which tetraploid cells injected into a blastocyst are excluded from the embryo but contribute to the placenta (Nagy et al., 1990). Wild-type placental cells are frequently polyploid or polynucleate, such as trophoblast giant cells, syncytiotrophoblasts, and extravillous cytotrophoblasts (Hu and Cross, 2010). These observations suggest that the placenta is the site of marked flexibility in the biochemical pathways generally controlling gene expression.

The plasticity of the placental environment, illustrated both in epigenetic marks and expression of imprinted genes, raises the question of the extent of placental imprinting.

Marked improvements in sequencing technology allow for the discrimination between, not only how many imprinted genes exist, but to what extent these genes display bias. High throughput RNA sequencing of F1 hybrids has provided an unprecedented ability to identify allele-specific expression due to imprinting (Goncalves et al., 2012; Gregg et al., 2010a; Gregg et al., 2010b; Okae et al., 2012; Wang et al., 2011), although this technology and the associated statistical and bioinformatic approaches are still in development. While initial studies identified a small number of novel imprinted transcripts (Babak et al., 2008; Wang et al., 2008), recent studies have suggested imprinting at more than 1,300 transcripts in the mouse brain (Gregg et al., 2010a; Gregg et al., 2010b). Additionally, several recent reports have used RNA sequencing of F1 hybrids to identify an additional 200 to 1000 imprinted genes specific to the placenta (Okae et al., 2012; Wang et al., 2011). There is still controversy in the field as to whether this marked increase in the number of imprinted transcripts is a result of increased sequencing depth and sensitivity or noise and bias in the sequencing data (DeVeale et al., 2012). It is essential to refine methodologies for sample preparation and data analysis in order for these technologies to clearly provide new insights into genomic imprinting.

Here, we provide a new means to detect parent-of-origin bias in the placenta. We show that our methods improve identification of imprinted loci; we identify regions on the paternal X chromosome that escape imprinting; and we demonstrate a clear preference for placental expression of the maternal allele of hundreds of transcripts. Moreover, throughout this analysis, we use individual placental samples rather than pooled samples, allowing us to analyze inter-individual variation. In all, this study defines an improved approach for studying imprinting in the placenta, and sheds light on several features of the placental epigenetic environment.

MATERIAL AND METHODS

An overview of the general experimental procedure (covered in the following four sections) is provided in Figure S1.

Mouse Strains and Crosses

Strains used were C57BL/6J and CAST/EiJ. C57BL/6J females were crossed with CAST/EiJ males to generate B6 X CAST F1 hybrids, and CAST/EiJ females were crossed with C57BL/6J males to generate CAST X B6 F1 hybrids. Placentas and yolk sacs of F1 hybrids were dissected at stage e11.5 (12 placentas from 3 litters B6 X CAST, 11 placentas from 2 litters CAST X B6) and e17.5 (6 animals from 1 litter, B6 X CAST only). To minimize maternal contamination, we removed the decidual layer of the placenta during dissection.

Illumina sequencing of 3' ends of mRNA

RNA sequencing libraries were built for each individual placenta using an approach which targets the 3' end (3SEQ) as described previously (Beck et al., 2010). Samples were prepped in cohorts containing individuals from multiple litters dissected on different days. Total RNA was isolated from placenta using a phenol-chloroform purification with Trizol

(Invitrogen). Additionally, DNA was precipitated with ethanol and purified for genotyping to verify gender. Briefly, after purification, mRNA was heat sheared before reverse transcription with the P7_oligodT primer. The P5 linker was then ligated to the free end, sequences are amplified using primers to P5 and P7, and the resulting library was sequenced from P5 using the Illumina Genome Analyzer II. In total, we sequenced and mapped 225 million reads for the B6 X CAST cross and 226 million reads for the CAST X B6 cross, with an average of 20 million reads per library.

Alignment and quantification

A *Mus musculus castaneus* genome was created in silico by applying the high-quality Castaneus SNPs (as defined by the Sanger Mouse Genomes Project) (Yalcin et al., 2012) to the reference C57-Black6/J genome. For each reference genome, a transcriptome was constructed using RefSeq annotations, including spliced transcripts of all genes separated by 200 "N" nucleotides. This spliced transcriptome sequence was added to the reference genomic sequence to create a composite genome.

BWA (Li and Durbin, 2009) was used to align all reads to both composite genomes separately, thus mapping each individual read to a genome. For each read, we retained its best alignments, but required that they all correspond to the same genomic location. This effectively disregarded reads mapping to multiple genomic locations but allowed reads to map to multiple transcriptomic positions that corresponded to the same exon or splice junction. The alignments for each read were then consolidated into a single genomic alignment, including intronic splice gaps. Since each read was aligned to both the B6 and Cast references, we chose the alignment from the reference with the fewest number of differences (selecting randomly in the case of a tie), and used this alignment for the remaining analyses.

For measuring allelic bias, only SNPs that were at least 50bp away from Sanger-annotated short Cast indels were used for counting allelic expression; this was done to reduce potential bias due to lower mapping sensitivity of reads containing indels.

Global expression patterns were exported from DNANexus using the 3'-seq/transcriptome-based quantification analysis. These data were subsequently normalized using DESeq (Anders and Huber, 2010).

Detection of Significant Effects and Identification of Candidates

For gene specific analyses, we compared the total number of reads mapped to the Castaneus allele to the total number of reads mapped to the B6 allele. For SNP specific analyses, we split each read "vote" evenly between the SNPs it contained, and used these numbers. We used a two-sample paired t-test and a Wilcoxon Mann Whitney test to determine when the maternal read counts were significantly different from the paternal read counts in our sampled biological replicates. A transcript was called as significantly biased if it had a p value less than 7E-07 in at least one test, which corresponds to a p value of 0.01 after Bonferroni correction.

Pyrosequencing

CDNA samples from both sides of the reciprocal cross as well as genomic DNA samples and primers were sent to Stanford's Protein And Nucleotide (PAN) facility, where sequences were analyzed using Qiagen's PyroMark Q24 system.

FISH

Wild type C57BL6/J e11.5 placentas (decidua still attached) were fixed in paraformaldehyde, embedded in OCT, and flash-frozen before being sectioned to 10 microns. Probe sets were designed by Biosearch Stellaris against the mRNA for candidate genes, and their protocol for fixed/frozen sections was followed.

Validation of Candidates using 3SEQ of a Decidual Sample

Maternal contamination was computationally modeled. Decidual expression patterns were determined using 3SEQ of a decidual sample. These sequencing results were mapped to the transcriptome to quantify global gene expression and compared to each fetal placental sample to derive a sample specific decidual/placental expression ratio for each gene in each fetal placental sample. Several mathematical models were used to determine percent decidual tissue:

1. The average maternal allele percent at known exclusively paternally expressed SNPs was calculated and used as an approximation for maternal tissue percent.

2. The slope of the line of best fit between maternal allele percent and relative decidual expression level at known paternally expressed genes was calculated and used as an approximation for maternal tissue percent, based on the model that maternal allele expression at these genes is proportional to the percent tissue as well as the relative expression level in decidua.

3. The slope of the line of best fit between maternal allele percent and relative decidual expression level at all genes was calculated and used as an approximation for maternal tissue percent, based on the above described model.

Extent of tissue contamination required to explain top candidate genes was calculated by modeling proportion maternal as a function of decidual tissue percent and relative decidual expression. The formula and derivation are below:

$$pMat = \frac{Maternal\ Reads}{Total\ Reads} = \frac{2*Tissue_D*Expression_D + Tissue_P*Expression_P}{2*Tissue_D*Expression_D + 2*Tissue_P*Expression_P}$$

$$pMat * \left(2*Tissue_D * \frac{Expression_D}{Expression_P} + 2*(1 - Tissue_D) \right) = 2*Tissue_D * \frac{Expression_D}{Expression_P} + 1 - Tissue_D\ Tissue_D * \left(1 + \frac{2*Expres}{Express} \right)$$

Correlation Analysis and Clustering

Heatmaps were drawn using variance-stabilized data (DESeq). Rows and columns were clustered using cluster 3.0 (distance metric: Kendall's tau, clustering method: average linkage).

Pearson correlation coefficients were calculated pairwise for each combination of samples. Variation between individuals in different litters versus variation between individuals in the same litter was determined using a two sample t-test for difference of mean of the Pearson correlation coefficients.

RESULTS

Sequencing Individual Placentas Reduces Error and Allows Internal Validation

We sought to examine allele specific variation during placental development using 3SEQ on individual placentas (Beck et al., 2010). Several studies have examined allelic bias within the placenta, but these had little consensus due to the use of different samples and methodologies. 3SEQ yields one read per transcript molecule, and thereby allows for more quantitative results with fewer total reads, allowing us to sequence individuals rather than pool samples. We used single nucleotide polymorphisms (SNPs) to identify parental contributions for each transcript in individual placentas from cross-species hybrids using *Castaneus* (Cast or C) and *C57-Black6* (B6 or B) parents. In total we sequenced mRNA from twelve F1 placentas at E11.5 from a B6 female X Cast male (BxC, litters B.1, B.3, and B.4) and eleven individuals from the reciprocal Cast female X B6 male (CxB, litters C.1 and C.2). These 23 samples provide a wealth of data with which to study imprinting within the mouse placenta. However, because there is no consensus on methodologies for sample preparation and statistical analysis, such datasets have been historically difficult to replicate due to internal biases and noise (DeVeale et al., 2012). Therefore, our first priority was to confirm the quality of the datasets.

The first test to validate the datasets was to examine how many transcripts contain a SNP in their 3' end that could distinguish parent of origin. While 3SEQ is useful because it provides more quantitative results, it achieves this by generating only one read per transcript, which is selected from the 3' end. This has the potential to limit the number of assayable genes because genes without a SNP in the 3' end cannot be tested for allelic bias. To measure the effect of sampling bias in this data, we determined the percentage of genes containing a useful SNP. We found at least one SNP in 67% of genes, two in 56% of genes, and three in 48% of genes (Fig S2A). A UCSC Genome Browser screen shot shows representative coverage of gene-rich regions (Fig S2B,C). This confirms our ability to sample a wide range of genes and effectively characterize allelic bias.

As a second validation step, we examined the frequency at which two SNPs in the same transcript agree on direction of bias (maternal or paternal). We expect that SNPs in the same coding region would be expressed with similar ratios, and that disagreements would likely be caused by errors and biases due to sequencing chemistry or alignment. When we examined all significantly biased SNPs with a p-value threshold of 0.0001, we observed 4,161 SNP pairs, of which 101 (2.43%) disagreed with respect to direction of bias (Fig 1A). When we examined all SNPs within significantly biased transcripts, we observed 1,693 SNP pairs of which 81 (4.87%) disagreed with respect to direction of bias (Fig 1B). This is significantly lower than previous studies which found approximately 20% discordance (DeVeale et al., 2012; Gregg et al., 2010a; Gregg et al., 2010b). As expected, more severe bias reflected by higher p-values is associated with lower levels of discordance (Fig 1C),

regardless of the test used to determine p-values. To ensure that the observed reduction in percent discordant SNPs was due to improvements in the sequencing methodology and not interdependencies between nearby SNPs, we repeated the analysis considering only those SNPs which were at least 100 base pairs apart, and thus separated by more than our maximum read length. With this methodology, we saw no discordant SNP pairs even at a much lower significance threshold: of 6,503 SNP pairs with bias significant to the $p < 0.05$ level, within the same transcript and separated by at least 100 base pairs, none were discordant. This indicates that our methods reduce the sequencing biases that likely cause SNP discordance.

Finally, we examined whether we could detect bias at known imprinted genes. We collected a list of 127 imprinted genes from studies in all mouse tissues ((Wang et al., 2011), with additions from www.otago.ac.nz/IGC; list is Table S1). Of these 127 genes, 43 showed multiple expressed SNPs within coding regions in the datasets. Considering these 43 "well-assayed" imprinted genes, 25 show significant bias for $p < 7E-7$, and 36 show significant bias for $p < 0.01$, in this study. Thus, the majority of known imprinted genes assayed by our methods are called as significant by these stringent criteria, highlighting the sensitivity of the method and the ability to identify true allelic bias.

Identification of 78 Novel Candidate Imprinted Genes in the Placenta

Confident in the quality of the datasets for identifying allelic bias, we next identified novel candidate imprinted genes. Because of the known difficulties in validating and reproducing novel candidate imprinted genes identified by RNA sequencing (DeVeale et al., 2012), we chose a stringent cutoff which we knew would fail to identify every known imprinted gene in the interest of producing fewer false positives; in particular we chose a p-value cut off of 7E-7, which corresponds to a 1% false discovery rate, or a p-value of 0.01 after Bon Feronni correction. We used both a t-test and a Wilcoxon Mann-Whitney test to identify genes for which maternal read counts significantly outnumbered paternal read counts across all 23 samples (Table 1; full results Table S2). 103 genes were determined to have significant parent-of-origin bias ($p < 7e-7$) in at least one test; 25 were previously characterized as imprinted, leaving 78 novel candidates for imprinting. All 103 genes are listed in table S2, and the top ten hits are in table 1.

We compared these 103 candidate imprinted genes with those found in two recent studies that examined allelic bias in the placenta. These studies used different methodologies, including pooling placental samples, examining different gestational ages and employing different statistical approaches. However, true bias should be relatively insensitive to methodology. In fact, we observed a striking overlap among all three studies in the previously known imprinted genes identified: out of 42 known imprinted genes identified in at least one study, 25 were seen in at least two, and eight in all three (Fig 1E), (Okae et al., 2012; Wang et al., 2011). The consistency of bias at known imprinted genes provides a metric for judging the quality of our gene list. Prior studies, exemplified by the gene lists gathered by Okae et al and Wang et al, have been very long but have had few overlaps: between 2 and 3% of total genes. In contrast, 50% of our novel candidate genes were replicated in a prior screen for imprinted genes (Fig 1F). To determine whether this

improved replicability was statistically significant, we used a hypergeometric test to compare the reproducibility of our gene list (the probability that a gene is identified in at least two studies, given that it is on our gene list) with the overall reproducibility of all three gene lists (the probability that a gene is identified in at least two studies, given that it is in one). The probability that our observed reproducibility occurred simply by chance was exceptionally small: $p = 8.3 \times 10^{-53}$. Hence, the observed reproducibility we see is significantly greater than in prior studies. These genes, which are identified as highly biased in multiple studies, are likely good candidate genes. Thus, the datasets presented in this paper are notably enriched for biased transcripts that have been replicated in other studies.

To confirm that our results were not caused by biases in sequencing chemistry, we used pyrosequencing at 15 of the 103 genes called for parent-of-origin bias (*Crabp2, Cryab, Degs2, Gzme, Gzmf, Gzmg, Htra3, Lifr, Mrgpre, Pik3ip1, Zdhhc14, Efemp2, Ggt1, Ascl2* and *Itpk1*). Pyrosequencing uses sequence-specific primers and unmodified bases, which means that the intrinsic biases of the method are different from next-generation sequencing techniques. Samples with true bias will show biased expression by pyrosequencing as well as 3SEQ. Each of our fifteen genes tested showed more than 50% maternal expression by pyrosequencing, and ten of these showed maternal bias greater than the known imprinted gene, *Ascl2* (Fig 1G). In addition, degree of bias determined via pyrosequencing appears highly correlated with degree of bias determined via 3SEQ, (Pearson's r: 0.92; p-value: 9.6E-7), which strongly validates the accuracy of our 3SEQ quantification results.

Identification of Genes on the X-Chromosome that Escape Paternal Silencing

Consistent with the fact that allelic bias is particularly prominent on the X chromosome due to the silencing of the paternal copy, we observe significant bias at most X-linked transcripts and no large-scale regions escaping X chromosome imprinting (Huynh and Lee, 2003). As expected, we observe maternal allelic bias essentially unbroken across the length of the X chromosome in female F1 placentas, and absent from autosomes (Fig2 A,B). In all, our data confirms imprinting across the X chromosome in female mouse placentas.

While we did not find any broad regions of biallelic expression, we examined all transcripts in order to identify individual genes that escaped paternal X inactivation. As a confirmation of the sensitivity of the method, we observed paternally-biased expression of *Xist*. This is reflected not only in significantly biased paternal expression when considering *Xist* alone, but also as a sharp spike in the plot (Table 2, red bar in Fig 2B). Of roughly 300 assayable X-linked genes with sufficient coverage, 17 did not show parent-of-origin bias as determined by a Student's t-test (p > 0.05; Table 2). Nine of these show significant bias with a Wilcoxon Mann-Whitney test (p < 0.05). Three genes, *Fhl1, Slc38a5, and Alas2*, are particularly good candidates for escape of imprinting. They show consistent expression from both alleles at SNPs with good coverage within the coding region, without neighboring repetitive regions or indels. At these loci a relative lack of maternal bias can be observed in the plot (purple bars, Fig 2B). To verify the accuracy of our sequencing results in detecting paternal transcription from the X chromosome, we used pyrosequencing to examine the three candidates identified (*Fhl1, Slc38a5*, and *Alas2*) as well as *Eif2s3x*, which has been shown to escape X-chromosome silencing in other tissues (Yang et al, 2010). We found

significant (>15%) expression from the non-dominant paternal allele in every case (Fig 2C). Thus, our detection of paternal expression from the X chromosome at four loci in female midgestation placentas is likely reflective of true transcription from the paternal allele. Overall, we demonstrated lack of parent-of-origin bias at three genes on the X-chromosome which have not previously been shown to escape X inactivation, and shown placental escape from X-chromosome inactivation for Eif2s3x.

While our sequencing data was not robust enough to demonstrate conclusive loss of X chromosome imprinting at all of the original 17 identified candidates, it was nonetheless possible that some of these loci were also escaping X chromosome inactivation. Therefore, we examined the relative expression ratio at all 17 genes initially identified as candidates for loss of X chromosome imprinting in male and female placentas. X chromosome inactivation is a dosage compensation mechanism, responsible for maintaining similar levels of X chromosome transcripts despite the haploidy in male samples. Thus, if a gene escapes X chromosome inactivation, it will generally be expressed at higher levels in female samples than in male samples. We excluded *Xist* from this analysis, due to the fact that it is only expressed in females, and in addition two genes – *Rps4x* and *Gm9112* – which were on our candidate list but did not have robust quantification. Many of the remaining 14 candidates indeed show higher expression levels in female placentas than in male placentas (Fig 2D). Of the strongest hits identified by SNP analysis, *Slc38a5* and *Alas2* have higher expression in female samples than in males. In addition, many of our other candidates show expression differences between males and females and are likely to have escaped X-chromosome inactivation, although the SNP data at these genes is not as conclusive due to the absence of multiple internal SNPs and low coverage at SNP sites. In particular, *Eif2s3x* has been shown to escape X chromosome inactivation in previous studies, shows paternal expression via pyrosequencing, and is likely a true hit. Expression analysis confirms loss of X chromosome inactivation at many of our candidate genes, including those with inconclusive SNP coverage data. Not only does this validate our list of genes which escape X chromosome inactivation, but it confirms that the loss of bias which we observed is correlated with a loss of silencing.

Candidate Imprinted Genes Show Allelic Bias in Other Tissues

We sought confirmation of the novel autosomal candidate imprinted genes by examining bias in other tissues. Although some imprinted genes show tissue-specific patterns of bias, most bias is maintained between tissue types and developmental stages. Therefore, we examined whether the 103 identified autosomal candidate imprinted genes maintain their bias in yolk sac and late stage placenta. To this end, we sequenced RNA from two E11.5 yolk sac samples (YS-1, YS-2), and five placental samples from an E17.5 litter (litter LS) and examined allelic ratios at the 103 transcripts with significant parent-of-origin bias identified in the E11.5 placenta. Of 103 genes, 83 showed bias in at least one other tissue or time point, and many (43 out of 103) showed bias in both (Fig 3A). While novel candidate genes showed more discordance between groups (Fig 3C), previously known imprinted genes were especially likely to show bias in all tissues (20 out of 25, Fig 3B). Hence, the allelic biases detected at E11.5 in the placenta are conserved among extra-embryonic tissues.

Samples Show Limited Decidual Contamination

As the maternal decidua and fetal placenta are tightly intercalated, it was crucial to determine the extent of decidual contamination of our samples. We determined the relative expression levels in maternal decidua versus fetal placenta at each transcript by sequencing RNA from one E11.5 maternal decidual sample, which we then compared to the averaged E11.5 fetal placental samples. If maternal bias were entirely explained by decidual contamination, the proportion of reads from the maternal allele should be determined by the amount of tissue derived from the mother (a constant within each sample) and the amount of expression from the decidua. However, we observed very limited correlation between maternal allele proportion and relative expression ratio when we examined all SNPs (Representative plot in Fig 4A; distribution of Pearson's R coefficients in bottom right corner; Fig S3 for all scatter plots). Thus, the majority of variation in maternal allele proportion across all SNPs cannot be explained by decidual contamination.

However, while decidual contamination was not sufficient to explain general trends across all SNPs, it is possible that some of the novel candidate genes showed particularly strong bias due to decidual contamination. To examine whether this was the case, we first needed an approximation of how much maternal tissue contamination was present in our samples. To this end, we determined maternal allele percent at a list of well-characterized genes known to be highly biased towards the paternal allele: *Peg10, Igf2, Peg3, Slc38a4, Plagl1, Kncq1ot1, Mest, Sfmbt2, Airn*, and *Sgce*. Unfortunately, these genes show inconsistent expression across our samples and previous studies have observed significant loss of imprinting at highly biased sites in placental samples (Lambertini et al., 2008), which introduce significant variation between samples and confounding variables of their own. Most striking was the variability in correlation coefficients (Fig 4B; distribution in bottom right corner): some samples had only two points and thus a perfect correlation, while others showed no correlation at all or a negative correlation. Given that each method for estimating decidual contamination levels was imperfect, we used an average of three different models. First, we used maternal read proportion of paternally expressed genes, by which we predicted our samples had a decidual contamination level between 0 and 20%. Second, we considered the slope of the linear regression model correlating maternal read proportion with decidual expression level: here, our samples show between 0 and 27% decidual contamination. Third, to ameliorate the problems introduced by inconsistent expression at known paternally expressed genes, we calculated the slope of the linear regression model correlating proportion maternal at all SNPs to relative decidual to placental expression level. This was between 0 and 2% in all samples. For the purposes of further analysis, we used an average between all three models; in this combined model our samples have a contamination level between 0 and 12%. In order to see whether decidual contamination could explain the presence of most genes on our gene list, we further used a mathematical model to determine the minimal amount of decidual tissue present in our samples that would explain our observed proportion maternal and decidual:placental expression ratio (Fig 4C). The formula used, and a summary of its derivation, are included in the methods section. While this modelling does show some genes which could be decidually derived (26 novel candidate genes require less than 20% dedicual contamination to explain observed levels of transcriptional bias), many novel candidate genes would require almost half of the tissue

dissected to be decidual in origin, and as such it is unlikely that the majority of our candidate genes are due to decidual contamination (Fig 4C).

A List of 24 High Value Candidates

To determine a top list of candidates for further testing, we classified the 78 novel candidates into likely contaminants or likely hits based on required percent decidual contamination as well as literature and database searches (Table S3). As suggested above, we find that 26 genes could indeed be due to maternal contamination (cut off at 20%). In addition, examination of literature and online databases revealed that 24 genes on our candidate list could be associated with blood and a further 2 were completely uncharacterized. Lastly, we found only one internal SNP in 2 of the candidates, suggesting that these may not be as robust. Therefore, we conservatively suggest that the parent-of-origin effect shown from sequencing in these 54 genes could be due to contamination. However, from our original list of 78 novel candidate genes, this provides 24 high-quality candidate genes (see Figure S1 for selection flowchart). Indeed, we show by pyrosequencing that 10 of these show significant parent-of-origin bias; namely *Crabp2, Cryab, Degs2, Gzme, Gzmf, Gzmg, Htra3, Mrgpre, Pik3ip1*, and *Zdhhc14* (Fig 1G). Of note is the fact that the only high-value candidate tested which did not show maternal bias greater than control via pyrosequencing was *Lifr*, which shows relatively low but nonetheless highly significant bias via 3SEQ as well, suggesting that this could be due to a highly replicable but overall slight effect at this locus. Overall, this suggests that these 24 are high value novel imprinting candidates that warrant deeper investigation.

To verify that the bulk of placental expression is from broad domains of mostly fetally derived cells rather than a few dispersed maternally derived cells, we used FISH to determine the spatial localization of several candidate genes within the placenta (Fig S4). As the analysis performed uses overall average expression to examine levels of contaminating cells – blood or decidua – it is important to determine whether this overall average expression level is made artificially high by a few, dispersed, decidual cells which highly express a transcript which could lower the D/P ratio and could create error in our mathematical model. Thus we examined the candidate genes, *Lifr, Degs2, Ggt1, Pik3ip1*, and two controls: *Fkbp6*, which shows predominantly paternal expression, and *Bpgm*, which is a gene known to be highly expressed in red blood cells. Low-magnification imaging with a decidual marker control revealed no particular bias towards decidual expression for any of these genes (Fig S4B, top rows). Higher magnification imaging revealed specific staining patterns to blood vessels, or common expression in the cytoplasm of many placental cells in all cases, confirming that at least these candidate genes are expressed in even levels in placental cells and support the assumptions in the mathematical model (Fig S4B, bottom rows).

Evidence for a Placental Maternal Transcription Bias

After extensive validation, a striking 75 out of 78 of all candidate genes (96%), and 22 out of 24 high-quality candidate genes (93%) are expressed from the maternal allele. We determined whether the significant maternal expression bias was due to a genome-wide effect. A genome-wide activation would be reflected as a slight bias at every gene, and a

global skew of maternal allele proportions across all autosomal transcripts. Therefore, we examined the distribution of maternal allele proportions across all autosomal genes in all of the samples. Most samples showed no skew, and no sample showed a median higher than 0.6 (Fig S5). This indicates that while there is a preference for highly biased sites to be transcribed from the maternal allele in the midgestation placenta, it is not due to a global transcriptional preference for the maternal genome, but rather is reflective of a specific effect at distinct loci.

To further investigate this surprising result, we examined whether this bias occurred only at our identified highly significant genes or – more generally – at many less significantly biased genes across the genome. Therefore, we broadened the analysis to include SNPs and transcripts that were highly biased (with 95 or 100% of reads coming from a single allele) but possibly insignificant due to low coverage (Fig 4D,E). Each placental sample showed a marked enrichment of 100% maternally expressed SNPs as compared to 100% or at least 95% paternally expressed SNPs (Fig 4D). To exclude the possibility that this result was due to exclusively decidual expression at thousands of genes, we modeled maternal decidual read count. Briefly, we multiplied our measured decidual read count at each SNP with our modeled decidual tissue contamination level for each placental sample in order to calculate an estimated number of decidually derived reads at each SNP in each sample. (Modeled Decidual Reads = Observed Reads in Decidual Sample * Modeled Percent Decidual Tissue.) When we subtracted out modeled decidual reads, we still observed that there was a significant, if small, enrichment of 100% maternally expressed SNPs as compared to 100% paternally expressed SNPs (Fig 4E; $p = 4.657e-06$ in a paired t-test). This is consistent with a generalized bias for the maternal allele, across thousands of SNPs and hundreds of transcripts, in all midgestation (E11.5) placental samples.

To determine whether this generalized preference for the maternal allele is maintained across extraembryonic tissues, we examined all highly biased SNPs and transcripts in two yolk sac and five late stage placental samples. We did not observe a general preference for the maternal allele in either yolk sacs or E17.5 placentas (Fig 4D,E). This is in marked contrast to our highly-conserved novel candidate imprinted genes which generally replicate in yolk sac and late stage placenta (Fig 3C). Not only does this suggest that the general preference for the maternal allele is tissue- and time point- specific, but it is inconsistent with the notion that this bias originates due to tissue contamination, as tissue contamination would be expected to play a role in bias in late stage as well as midgestation placental samples. These observations indicate some global maternal expression bias specific to the midgestation placenta.

Variation in Transcription Patterns is Highly Litter-Specific

Sequencing individual placentas allowed for an examination of variation between individuals. As the placenta shows a great deal of epigenetic plasticity, variation, and environmental response (Cross and Mickelson, 2006), we sought to examine the extent and patterns of this variation within and between litters. To this end, we used Cluster 3.0 to perform hierarchical clustering on expression data at all genes, expression data at imprinted genes, and maternal/paternal allele ratio at imprinted genes (see methods). We observed

tissue-specific (yolk sac, placenta, or decidua), time-point specific (E17.5 versus E11.5), and litter-specific patterns in all cluster diagrams (Fig 5A–C). These patterns were especially striking when expression data at all genes were examined (Fig 5A). As expected, yolk sac, maternal decidua, and late stage samples form three individual clusters in the hierarchical clustering. Yolk sac samples in particular cluster separately from all placental samples (Fig 5A–C, yellow bars). In addition, within the placental cluster, the E17.5 and E11.5 stages are distinct. The presence and prominence of these patterns in these datasets serve as a validation of the methods.

In addition to clustering by tissue type and stage, we observed a surprising tendency for E11.5 placentas to cluster by litter (Fig 5A–C). All samples came from genetically identical individuals and were collected at approximately the same time point. Random environmental fluctuations in laboratory mice would be expected to be very small. Nonetheless, samples from littermates show a high degree of correlation not seen in other comparisons. To further analyze this general trend, we examined pairwise correlations between individual E11.5 placental samples. We determined the Pearson's R correlation coefficient for global expression patterns for each pair of E11.5 fetal placental samples. Again, we observed a higher degree of correlation when comparing littermates than when comparing individuals from different litters (Fig 5D). A t-test confirms that correlations between littermates are greater than correlations between non-littermates with a p-value of 1.3E-7.

We sought to identify particular genes or classes of genes responsible for the observed patterns of variation between litters. To this end, we performed principal component analysis on the global expression data (Fig S6). We were able to identify principal components linked to the variation between mid-gestation and late stage placenta (PC1), the tissue specific differences between yolk sac and placenta (PC2), and the differences between maternal decidua and placental samples (PC5). While litters often formed non-overlapping or minimally overlapping groups, no specific principal component or pair of principal components separated all mid-gestation placental samples according to litter. We also examined gender differences, but were unable to find a principal component that segregated the mid-gestation samples by gender.

In order to determine whether there was an underlying biological distinction between principal components, we used gene ontology (GO) to find enriched terms in the top 100 weighted genes for each principal component (Fig S6 G). Top results for PC1 were development related. PC2 was enriched for proteins found in the basement membrane. PC3 and PC4 were both enriched for nuclear factors. PC5 was enriched for extracellular factors. These gene ontology results were unsurprising given the classes of samples the principal components differentiated.

DISCUSSION

The rapid expansion of sequencing technology has created a growing need for novel methods to refine and interpret complex datasets. The improved methodology highlighted in this paper allows four main conclusions: first, 3SEQ analysis of individual placental samples

generates a robust data set for identifying candidate novel imprinted genes. Second, certain genes on the X chromosome escape paternal silencing in female placentas. Third, there is a genome-wide preference in the midgestation placenta towards expressing the maternal allele, which cannot be explained by maternal contamination. Finally, there is a large amount of variation between litters in gene expression and degree of bias at imprinted genes in the placenta.

First, we generated a robust list of candidate imprinted genes by using 3SEQ to sequence RNA from individual mouse placentas. Our sample preparation methods greatly reduced sequencing noise, as demonstrated by discordance between SNPs within the same transcript. We suggest the possibility that 3SEQ reduces chemical biases by removing the use of random hexamers and creating a one-to-one correspondence between transcripts and sequenced reads. These two advantages would eliminate some biases in sample preparation and increase the quantitative nature of our results. Additionally, sequencing RNA from multiple individuals, instead of pooling samples, reduces the effect of random sequencing errors and noise on the results. In all, we developed an improved methodology for the identification of novel imprinted gene candidates. This led to a shorter gene list with a greater degree of overlap with previously characterized imprinted genes and both known placental imprinted genes and potential candidate lists from other studies. This gene list is also strikingly replicable among extra-embryonic tissues and time points.

Second, we observed strong evidence that at least three loci escape silencing of the paternal X chromosome. While previous studies have demonstrated this silencing in the female mouse placenta (Huynh and Lee, 2003; Takagi and Sasaki, 1975; West et al., 1977), they have not clarified the extent of imprinting on the X chromosome and whether any paternal loci are able to escape silencing. The presence of such loci is a strong indicator of the flexibility of placental imprinting. The loci we identified are spread out across the chromosome and not clustered, indicating that this is likely to be a gene-specific loss of silencing, rather than one affecting larger scale domains on the chromosome. These genes include *Eif2s3x*, a translation initiation factor which has been previously shown to escape X chromosome inactivation, *Alas2*, the enzyme responsible for the first step in heme biosynthesis, *Fhl1*, a forkhead-like protein which most likely binds to DNA, and *Slc38a5*, an amino-acid transporter. How these genes escape silencing on the paternal X chromosome, and what their role in placental development may be, is unknown, but these are interesting avenues for future study.

Third, we found evidence of a large-scale preference for the maternal allele among significantly biased genes in the placenta. Previously-reported discoveries of maternal bias in the placenta have recently been called into question due to the tight intercalation of maternal and fetal tissues in the placenta, which leads to the possibility of maternal tissue contamination (Hudson et al., 2011). This observed bias extends to 75 of 78 novel candidate imprinted genes as well as hundreds of highly biased sites in individual samples. While we might have expected these results to be caused by maternal contamination, we find little evidence of this. After a thorough analysis, we observed no correlation between high expression in the decidua and high levels of allelic bias at known paternally imprinted genes, candidate imprinted genes, or across all genes. Thus, it is unlikely that decidual

contamination plays a large role in skewing the results towards maternal expression, and the observed preference for the maternal allele across many genes is likely controlled by allelic bias within the fetal placenta.

Interestingly, a placental expression bias towards maternal alleles goes against the observation that placental tissues develop mostly normally with two paternal genomes and fail to grow with two maternal genomes, as well as a recent analysis showing enrichment for paternally expressed imprinted genes in the mule placenta (Wang et al., 2013). It is possible that the maternal bias we observe is necessary to regulate later developmental processes that were not captured in the parthenogenesis and androgenesis experiments. For example, these genes may regulate hormones and nutrients in order to suppress placental or fetal growth at midgestation or later. Another possibility may be a role in regulating the maternal immune response to prevent rejection of the growing fetus. Expressing more maternal cell surface markers might go some way towards camouflaging the fetal placental cells and preventing rejection of the fetus. Consistent with either hypothesis, many of the novel candidate genes are extracellular proteins. In all, the function of a large-scale maternal bias in placental tissues, which require the paternal genome to develop, is an interesting phenomenon that warrants further study.

Finally, we have demonstrated that there exists a large amount of tissue-specific, stage-specific, and individual variation in the wild-type mouse placenta. While yolk-sac samples have been proposed as a maternal-contamination-free alternative for studying allelic bias in extraembryonic lineages (Hudson et al., 2011), the relatively small degree of maternal tissue contamination in our samples coupled with the widespread transcriptional differences between yolk sac and placental samples reinforce the importance of studying the placenta. More strikingly, we observed that gene expression patterns in the placenta are largely dependent on developmental stage, with large differences between mid- and late-gestation placentas. Our variation between litters suggests that even the small putative differences in environment and slight variation in developmental stage found between genetically identical midgestation litters were reflected in placental expression pattern. This is consistent with the hypothesis that the placenta is uniquely responsive to the environment, modulating gene expression patterns continuously in response to inputs from the external environment as well as the developing embryo (Cross and Mickelson, 2006; Lambertini et al., 2008), and recent observations imprinted genes in particular are highly individually variable (Wang et al., 2013). The degree of variation between individuals underscores the plasticity of imprinting in the placenta and the importance of assaying multiple individual samples when studying the placenta.

In all, we demonstrate that imprinting in the placenta occurs through tissue-specific patterns, with a marked preference for the maternal allele. Our list of novel candidate genes includes genes which are not expressed in later stages or yolk sac samples, as well as some which do not show bias in these other tissues. We observe tissue-specific escape of silencing on the X chromosome as well – particularly at four genes, *Alas2, Fhl1, Slc38a5*, and *Eif2s3x*, whose roles in placentation have not yet been elucidated. Most strikingly, we see a general preference for the maternal allele not only at novel candidate imprinted genes but also at thousands of SNPs in hundreds of transcripts genome-wide, specific to the mid-gestation

placenta. Overall, this paper highlights the dynamic and unique nature of the epigenetic landscape in the placenta and provides evidence that the maternal contribution in this organ is more substantial than previously appreciated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Anders S, Huber W. Differential Expression Analysis for Sequence Count Data. Genome Biology. 2010; 11:R106. [PubMed: 20979621]

Babak T, DeVeale B, Armour C, Raymond C, Cleary MA, Kooy D, Johnson JM, Lim LP. Global Survey of Genomic Imprinting by Transcriptome Sequencing. Current Biology. 2008; 18:1735–1741. [PubMed: 19026546]

Barton SC, Surani MA, Norris ML. Role of paternal and maternal genomes in mouse development. Nature. 1984; 311:374–376. [PubMed: 6482961]

Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, Lacroute P, Smith CL, Tibshirani R, Rijn M, Sidow A, West RB. 3′-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples. PLoS ONE. 2010; 5:e8768. [PubMed: 20098735]

Brockdorff N, Ashworth A, Kay G, Cooper P, Smith S, McCabe V, Norris D, Penny G, Patel D, Rastan S. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature. 1991; 351:329–331. [PubMed: 2034279]

Brown C, Ballabio A, Rupert J, Lafreniere R, Grompe M, Tonlorenzi R, Willard H. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature. 1991; 349:38–44. [PubMed: 1985261]

Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet. 2013; 45:325–329. [PubMed: 23396136]

Cross JC, Mickelson L. Nutritional Influences on Implantation and Placental Development. Nutrition Reviews. 2006; 64:S12–S18. [PubMed: 16770948]

DeVeale B, Kooy D, Babak T. Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. PLoS Genet. 2012; 8:e1002600. [PubMed: 22479196]

Devlin CM, Lahm T, Hubbard WC, Van Demark M, Wang KC, Wu X, Bielawska A, Obeid LM, Ivan M, Petrache I. Dihydroceramide-based Response to Hypoxia. Journal of Biological Chemistry. 2011; 286:38069–38078. [PubMed: 21914808]

Ferdous Z, Peterson SB, Tseng H, Anderson DK, Iozzo RV, Grande-Allen KJ. A role for decorin in controlling proliferation, adhesion, and migration of murine embryonic fibroblasts. Journal of Biomedical Materials Research Part A. 2010; 93A:419–428. [PubMed: 19569212]

Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Research. 2012; 22:2376–2384. [PubMed: 22919075]

Gregg C, Zhang J, Butler JE, Haig D, Dulac C. Sex-Specific Parent-of-Origin Allelic Expression in the Mouse Brain. Science. 2010a; 329:682–685. [PubMed: 20616234]

Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Mouse Brain. Science. 2010b; 329:643–648. [PubMed: 20616232]

Gu M, Pritlove DC, Boyd CAR, Vatish M. Placental Expression of 2,3 Bisphosphoglycerate Mutase in IGF-II Knock out Mouse: Correlation of Circulating Maternal 2,3 Bisphosphoglycerate and Fetal Growth. Placenta. 2009; 30:919–922. [PubMed: 19733906]

Harper M, Fosten M, Monk M. Preferential paternal X inactivation in extraembryonic tissues of early mouse embryos. J Embryol Exp Morphol. 1982; 67:127–135. [PubMed: 7086330]

Hu D, Cross JC. Development and function of trophoblast giant cells in the rodent placenta. Int J Dev Biol. 2010:341–354. [PubMed: 19876834]

Hudson QJ, Seidl CI, Kulinski TM, Huang R, Warczok KE, Bittner R, Bartolomei MS, Barlow DP. Extra-embryonic-specific imprinted expression is restricted to defined lineages in the post-implantation embryo. Developmental Biology. 2011; 353:420–431. [PubMed: 21354127]

Huynh KD, Lee JT. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature. 2003; 426:857–862. [PubMed: 14661031]

Iozzo RV, Moscatello DK, McQuillan DJ, Eichstetter I. Decorin Is a Biological Ligand for the Epidermal Growth Factor Receptor. Journal of Biological Chemistry. 1999; 274:4489–4492. [PubMed: 9988678]

Johnston C, Nesterova T, Formstone E, Newall A, Duthie S, Sheardown SNB. Developmentally regulated *Xist* promoter switch mediates initiation of X inactivation. Cell. 1998; 94:809–817. [PubMed: 9753327]

Lambertini L, Diplas AI, Lee M, Sperling R, Chen J, Wetmur JG. A sensitive functional assay reveals frequent loss of genomic imprinting in human placenta. Epigenetics. 2008; 3:261–269. [PubMed: 18769151]

Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, Walter J, Higgins M, Feil R, Reik W. Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet. 2004; 36:1291–1295. [PubMed: 15516931]

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

Lieberman MW, Barrios R, Carter BZ, Habib GM, Lebovitz RM, Rajagopalan S, Sepulveda AR, Shi ZZ, Wan DF. gamma-Glutamyl transpeptidase. What does the organization and expression of a multipromoter gene tell us about its functions? Am J Pathol. 1995; 147:1175–1185. [PubMed: 7485380]

Lyon MF. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature. 1961; 190:372–373. [PubMed: 13764598]

Monk M, Harper M. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. Nature. 1979; 281:311–313. [PubMed: 551278]

Nagy A, Gocza E, Diaz EM, Prideaux VR, Ivanyi E, Markkula M, Rossant J. Embryonic stem cells alone are able to support fetal development in the mouse. Development. 1990; 110:815–821. [PubMed: 2088722]

Okae H, Hiura H, Nishida Y, Funayama R, Tanaka S, Chiba H, Yaegashi N, Nakayama K, Sasaki H, Arima T. Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. Human Molecular Genetics. 2012; 21:548–558. [PubMed: 22025075]

Razin A, Webb C, Szyf M, Yisraeli J, Rosenthal A, Naveh-Many T, Sciaky-Gallili N, Cedar H. Variations in DNA methylation during mouse cell differentiation *in vivo* and *in vitro*. Proc Natl Acad Sci USA. 1984; 81:2275–2279. [PubMed: 6585800]

Reed CC, Iozzo RV. The role of decorin in collagen fibrillogenesis and skin homeostasis. Glycoconj J. 2002; 19:249–255. [PubMed: 12975602]

Rugg-Gunn PJ, Cox BJ, Ralston A, Rossant J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc Natl Acad Sci U S A. 2010; 107:10783–10790. [PubMed: 20479220]

Sakaue M, Ohta H, Kumaki Y, Oda M, Sakaide Y, Matsuoka C, Yamagiwa A, Niwa H, Wakayama T, Okano M. DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages. Current Biology. 2010; 20:1452–1457. [PubMed: 20637626]

Schonherr E, Sunderkotter C, Iozzo RV, Schaefer L. Decorin, a novel player in the insulin-like growth factor system. J Biol Chem. 2005; 280:15767–15772. [PubMed: 15701628]

Searle AG, Beechey CV. Complementation studies with mouse translocations. Cytogenetic and Genome Research. 1978; 20:282–303.

Surani MA, Barton SC. Development of Gynogenetic Eggs in the Mouse; Implications for Parthenogenetic Embryos. Science. 1983; 222:1034–1036. [PubMed: 6648518]

Surani MA, Barton SC, Norris ML. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature. 1984; 308:548–550. [PubMed: 6709062]

Takagi N, Sasaki M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature. 1975; 256:640–642. [PubMed: 1152998]

Tate SS, Meister A. gamma-Glutamyl transpeptidase: catalytic, structural and functional aspects. Mol Cell Biochem. 1981; 39:357–368. [PubMed: 6118826]

Torres-Padilla M, Parfitt D, Kouzarides T, Zernicka-Goetz M. Histone arginine methylation regulates pluripotency in the early mouse embryo. Nature. 2007; 445:214–218. [PubMed: 17215844]

Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, Feil R. Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet. 2004; 36:1296–1300. [PubMed: 15516932]

Wang X, Miller DC, Harman R, Antczak DF, Clark AG. Paternally expressed genes predominate in the placenta. Proceedings of the National Academy of Sciences. 2013

Wang X, Soloway PD, Clark AG. A Survey for Novel Imprinted Genes in the Mouse Placenta by mRNA-seq. Genetics. 2011; 189:109–122. [PubMed: 21705755]

Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG. Transcriptome-Wide Identification of Novel Imprinted Genes in Neonatal Mouse Brain. PLoS ONE. 2008; 3:e3839. [PubMed: 19052635]

West J, Frels W, Chapman V, Papaioannou V. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. Cell. 1977; 12:873–882. [PubMed: 597862]

Xiol J, Cora E, Koglgruber R, Chuma S, Subramanian S, Hosokawa M, Reuter M, Yang Z, Berninger P, Palencia A, Benes V, Penninger J, Sachidanandam R, Pillai RS. A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. Mol Cell. 2012; 47:970–979. [PubMed: 22902560]

Yalcin B, Adams D, Flint J, Keane T. Next-generation sequencing of experimental mouse strains. Mammalian Genome. 2012; 23:490–498. [PubMed: 22772437]

Yang F, Babak T, Shendure J, Disteche C. Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. 2010; 20(5):614–622. [PubMed: 20363980]
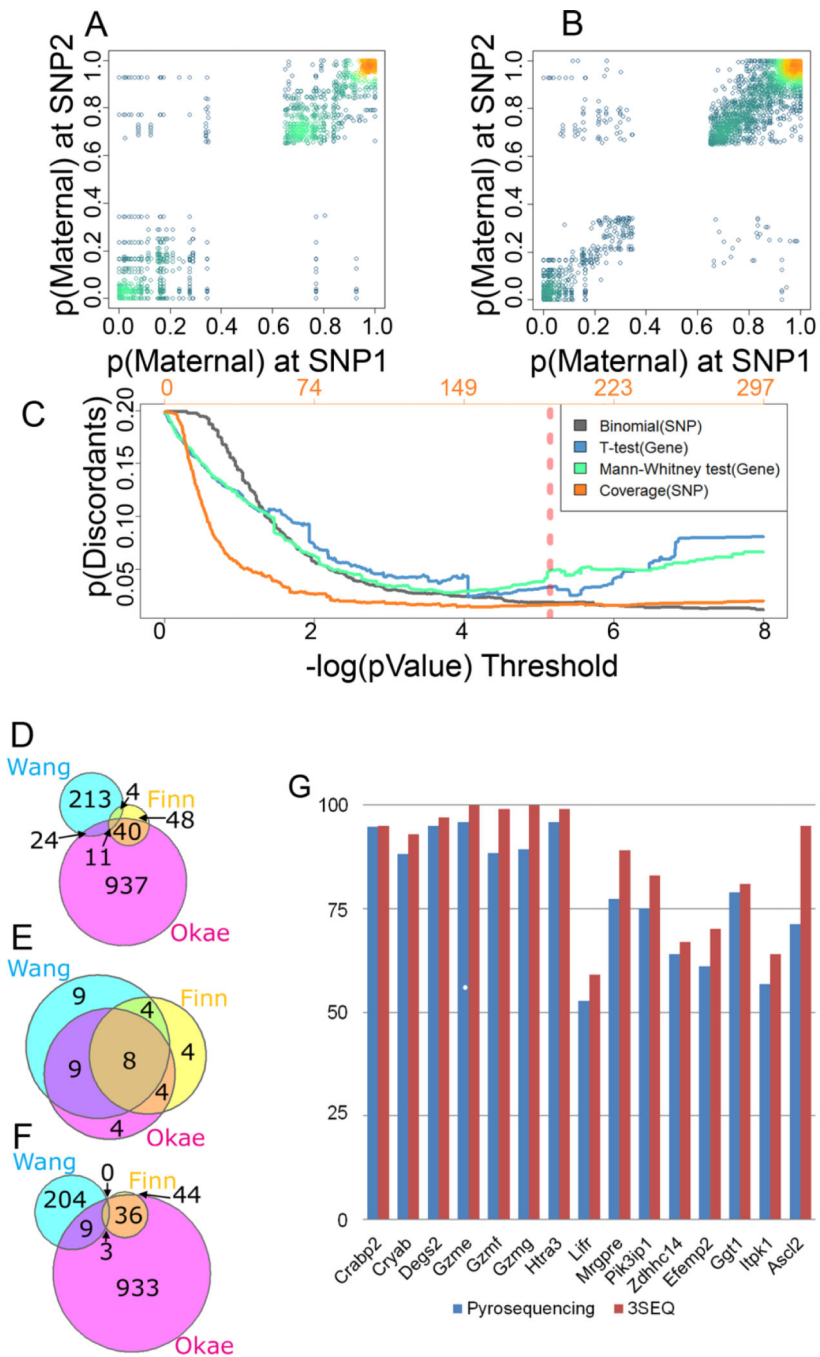
Figure 1. A Robust Analysis Yields a Replicable Gene List

A) Discordance between pairs of highly significantly biased (p<0.0001, binomial) SNPs within the same transcript. Density of points by color: orange is high density and blue low. B) Discordance between SNPs in highly significantly biased (p<7E-07, t-test or Wilcoxon mann-whitney) transcripts. Density of points by color: orange is high density and blue low. C) Trends of discordance rate using different measures of significance (by p-value on the bottom y axis) or coverage (by number of sequenced reads on the top y axis). D–F) Venn diagrams comparing the 103 genes from this analysis with other reports (Wang et al, 2011;

Okae et al, 2012). D) All genes. E) Previously known imprinted genes. F) Novel candidate imprinted genes. G) Pyrosequencing results confirming bias of several candidate genes; proportion maternal reads is shown, those derived from pyrosequencing in blue and those derived from 3SEQ in red.
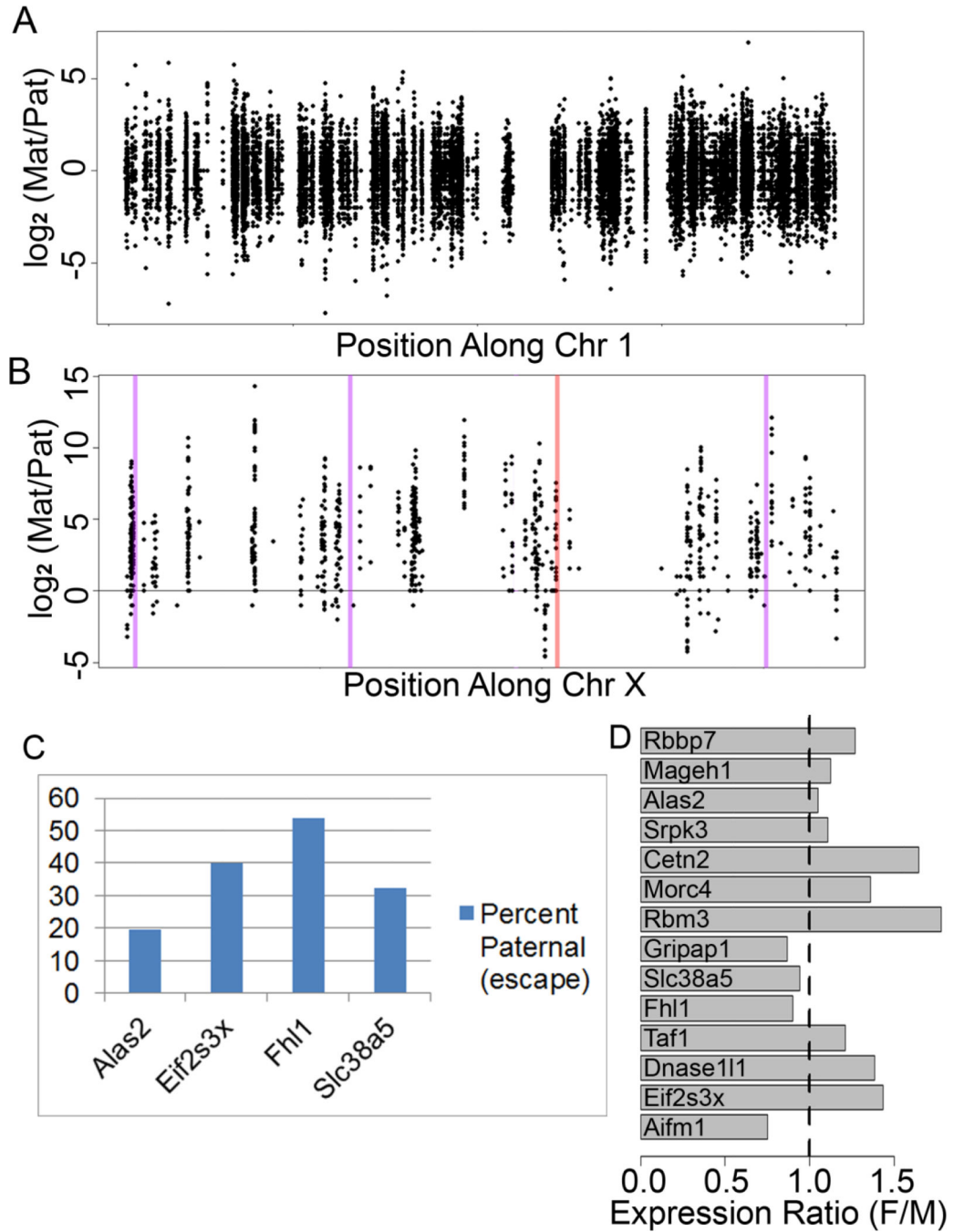
Figure 2. Genes that Escape X Chromosome Inactivation

A,B) Plots showing bias at SNPs (as log ratio, y axis) versus position on a chromosome (x axis) for all female samples. A) Chromosome 1. B) Chromosome X. Vertical bars show locations of potential genes of interest; Red bar: *Xist*. Purple bars: *Alas2, Fhl1*, and *Slc38a5*. C) Bar plot showing percent paternal expression detected via pyrosequencing. D) Bar plot shows ratios between mean normalized expression in female and male samples at genes which escape X chromosome imprinting.
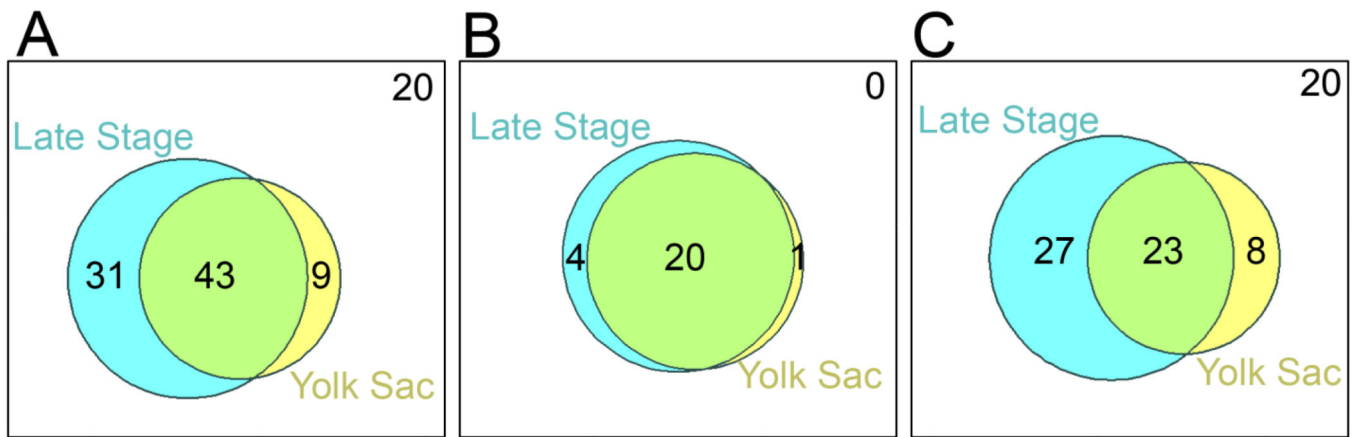
Figure 3. Candidate Imprinted Genes Replicate in Different Tissues and Time Points
A) All 103 imprinted candidate genes. B) Only previously known imprinted genes. C) Only novel candidate genes.
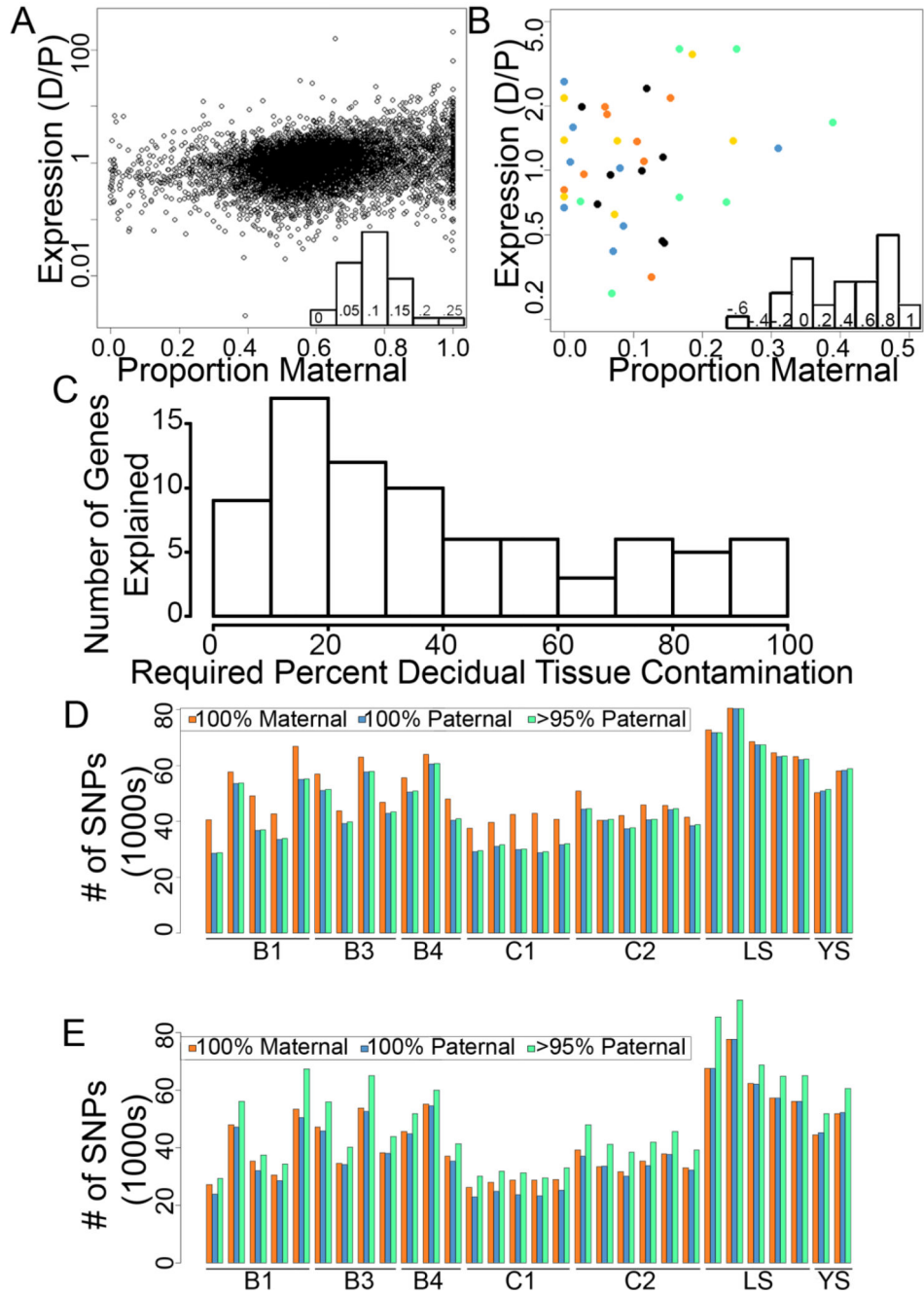
Figure 4. The Maternal Genome is Preferred in Placental Samples
A) Maternal allele percentage (x axis) as a function of relative decidua to placenta expression ratio (y axis) in one sample. Distribution of Pearson's correlation coefficients in the bottom right corner. Plots for all other samples are in figure S3. B) Maternal allele percentage (x axis) as a function of relative decidua to placenta expression ratio (y axis) at known paternally expressed genes. Five representative samples are shown (color coded). Each dot represents one gene in one sample. Distribution of Person's correlation coefficients for all samples is in the bottom right corner. C) Histogram distribution of decidual tissue

percent required to explain observed proportion maternal reads at novel maternally imprinted candidate genes. D) Counts of highly biased SNPs (100% maternal (orange), 100% paternal (blue), 95% or more paternal (green)). E) Counts of highly biased SNPs after subtracting modeled decidual reads (100% maternal (orange), 100% paternal (blue), 95% or more paternal (green)).
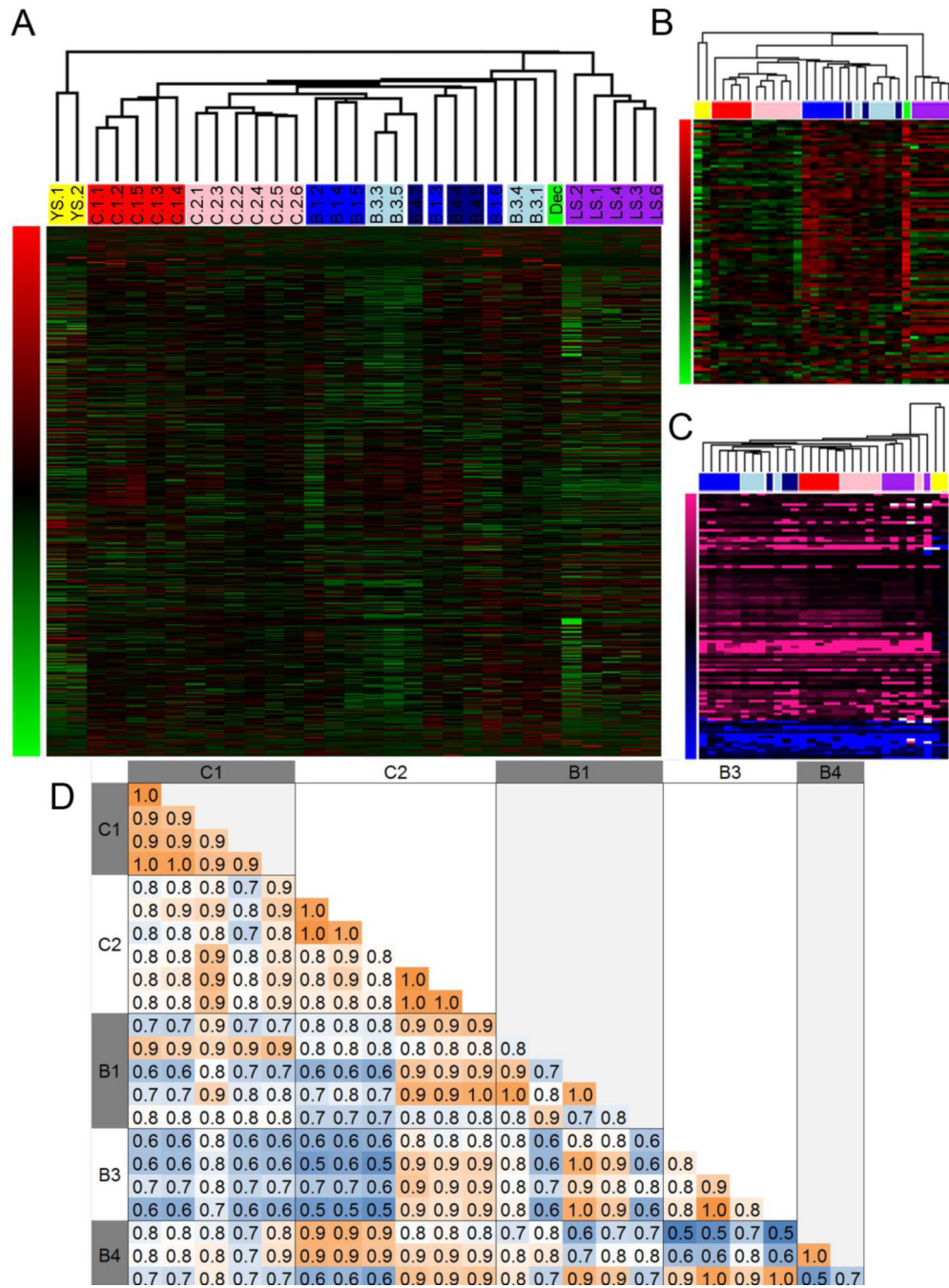
Figure 5. Transcription and Bias Vary in a Litter-Specific Manner
A) Heatmap and cluster diagram showing expression (red: high, green: low) across all genes. Samples are labeled and color coded by litter/tissue (Yellow: yolk sac. Red: Cast litter 1. Pink: Cast litter 2. Blue: B6 litter 1. Light blue: B6 litter 3. Dark blue: B6 litter 4. Purple: LS litter. Green: Decidua.) B) Heatmap and cluster diagram showing expression at biased genes. Samples are labeled and color coded by litter/tissue (Yellow: yolk sac. Red: Cast litter 1. Pink: Cast litter 2. Blue: B6 litter 1. Light blue: B6 litter 3. Dark blue: B6 litter 4. Purple: LS litter. Green: Decidua.) C) Heatmap and cluster diagram showing proportion

maternal at biased genes (pink: high, blue: low). Samples are labeled and color coded by litter/tissue (Yellow: yolk sac. Red: Cast litter 1. Pink: Cast litter 2. Blue: B6 litter 1. Light blue: B6 litter 3. Dark blue: B6 litter 4. Purple: LS litter. Green: Decidua.) D) Chart of Pearson's correlation coefficients for a pairwise comparison between each F1 individual placenta at the E11.5 timepoint. Chart titles are color coded by litter and cells are colored by pearson's R value (orange: high, blue: low).

Table 1

Top 10 Candidate Imprinted Genes

	T-TEST	WILCOX	TOT SNPS	INT SNPS	D/P	AVG PMAT YS	AVG PMAT LATE	AVG PMAT EARLY	KNOWN IMPRINTED
Apeg3	5.45E-07	9.31E-09	6	3	2.4E+00	4.7E-01	4.7E-03	7.0E-02	Y
Ascl2	1.05E-09	5.59E-09	2	2	4.0E-01	5.0E-01	9.3E-01	9.5E-01	Y
Cd81	3.29E-08	1.86E-09	5	5	3.4E+00	6.7E-01	6.8E-01	6.8E-01	Y
Cdkn1c	1.11E-04	1.86E-09	2	2	6.5E-01	9.1E-01	9.6E-01	9.7E-01	Y
Dcn	2.24E-03	1.86E-09	14	9	2.1E+01	5.4E-01	9.0E-01	9.8E-01	Y
Gab1	2.37E-11	1.86E-09	15	15	7.1E-01	2.2E-01	1.3E-01	1.8E-01	Y
Dlk1	1.22E-02	1.86E-09	7	6	1.1E+00	2.6E-02	9.0E-03	5.5E-02	Y
Fkbp6	3.69E-10	1.86E-09	4	2	8.8E-01	0.0E+00	5.2E-02	4.1E-02	N
Pik3ip1	3.85E-10	2.61E-08	5	5	5.0E+00	5.0E-01	6.2E-01	8.3E-01	N
H13	1.57E-13	1.86E-09	12	11	1.5E+00	7.7E-01	6.2E-01	6.9E-01	Y

Legend: T-test: P-value determined by t-test. Wilcox: P-value determined by Wilcoxon Mann-Whitney test. Tot SNPs: Total SNPs associated with transcript. Int SNPs: SNPs within the canonical transcribed region. D/P: Decidua to Placenta expression ratio. Avg pMat YS: Average proportion maternal in yolk sac samples. Avg pMat Late: Average proportion maternal in late-stage samples. Avg pMat Early: Average proportion maternal in midgestation samples. Known Imprinted: is the gene previously characterized as imprinted?

Table 2

Escape of X-Chromosome Imprinting

	T-TEST	WILCOX	T-TEST (STRAIN)	PATTERN	READS	SNPS	INTERNAL SNPS
Aifm1	0.99671	1.00000	0.02996	B	140.5510709	8	0
Eif2s3x	0.59271	0.55411	0.03930	C	56.35398143	4	1
Gm9112	0.33042	0.00391	0.34536	M	647.2769854	1	1
Dnase1l1	0.24162	0.01427	0.41277	M	71.90702262	1	1
Rps4x	0.16814	0.36273	0.01580	B	2950.800232	1	1
Taf1	0.14358	0.12891	0.03021	C	130.9077906	4	1
Fhl1	0.12207	0.12891	0.03844	C	157.8631743	2	2
Xist	0.11019	0.01953	0.43144	P	614.4166317	17	17
Slc38a5	0.09537	0.12891	0.00994	C	156.0470339	2	2
Gripap1	0.09148	0.02249	0.23773	M	54.32288248	7	6
Rbm3	0.08323	0.08006	0.42886	M	85.14567815	5	0
Morc4	0.07550	0.00391	0.49397	M	300.6789859	8	7
Cetn2	0.07320	0.03603	0.58172	M	103.2921476	1	1
Srpk3	0.07314	0.00391	0.52034	M	68.65064251	1	0
Alas2	0.06281	0.07422	0.02662	M	149.9125338	6	4
Mageh1	0.05944	0.00391	0.51763	M	138.4744813	5	5
Rbbp7	0.05136	0.01427	0.49963	M	85.86870404	3	1

Legend: T-test: P-value determined by t-test. Wilcox: P-value determined by Wilcoxon Mann–Whitney test. T-test (strain): P-value for strain-specific expression determined by t-test. Pattern: pattern of bias; B: B6 allele preferred, C: Cast allele preferred; M: Maternal allele preferred; P: Paternal allele preferred. Reads: Total reads, normalized by DESeq. SNPs: Total SNPs associated with transcript. Internal SNPs: SNPs within the canonical transcribed region.