Title

Patchy, incomplete, and heterogeneous X-inactivation in the human placenta

Authors

Tanya N. Phung[1,2], Kimberly C. Olney[1,2], Harvey J. Kliman[3], and Melissa A. Wilson[1,2*]

Affiliations

1. Center for Evolution and Medicine, Arizona State University, Tempe AZ 85282 USA
2. School of Life Sciences, Arizona State University, Tempe AZ 85282 USA
3. Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT 06520

*Corresponding author
Melissa A. Wilson
School of Life Sciences | Arizona State University | PO Box 874501 | Tempe, AZ 85287-4501
mwilsons@asu.edu

Author contributions

TNP: Project Lead, Formal Analysis, Investigation, Visualization, Writing - Original Draft
KCO: Formal Analysis, Investigation, Visualization, Writing - Original Draft
HJK: Sample access, Writing- Review and Editing
MAW: Conceptualization, Supervision, Visualization, Resources, Project Administration, Writing - Original Draft Preparation, Writing - Review and Editing, Funding Acquisition

Abstract

The placenta is formed after the first few weeks of pregnancy and is the genotype of the fetus. It acts as an immune modulator in the uterine environment to sustain a successful pregnancy. One of the X chromosomes in XX females is silenced by a process called X-inactivation. Prior research suggests that incorrect dosage on the X chromosome could lead to poor development of the placenta and ultimately result in complications in pregnancy. Previous studies of X-inactivation in the placenta were either in non-human placentas, or were limited to only a few SNPs and genes in human placentas. Thus, it is not clear whether within human placenta, X-inactivation is completely homogeneous, patchy, or mosaic. Further, X-inactivation is not complete in humans; as many as one-third of the genes on the X chromosome escape X-inactivation, but variability in genes that escape X-inactivation in the placenta has not been investigated. We sequenced RNA from 60 placenta samples from 30 full-term, uncomplicated pregnancies with female offspring. We can confidently rule out X-inactivation being completely mosaic in the human placenta. Rather, we find strong evidence that X-inactivation in the human placenta is patchy, with large potential clonal expansions of either silenced maternal or paternal X-chromosomes, with provocative suggestions of bias towards silencing the paternal X. We also find variation in the degree of silencing, where a high portion of variants (between 26.8-75.3% in any sample) are silenced incompletely. Finally, we find evidence for variability in genes that escape X-inactivation within and among placenta samples.

Significance

The placenta is formed early during development, is essential for healthy pregnancy, and is largely composed of DNA from the offspring, and thus can be XX or XY. One of the X chromosomes in XX females is silenced by a process called X-inactivation. We studied patterns of X-inactivation in two regions of the placenta across 30 placentas, and find strong evidence for large patches of either maternally or paternally silenced X-chromosomes in the human placenta. We also found that the genes that escape X-inactivation vary both across individuals, and within the placenta of a single individual, suggesting mechanisms for variability in placenta function, and implications for prenatal testing that samples only a single region of the placenta.

Introduction

In eutherian mammals, X-inactivation is a mechanism to regulate dosage of gene expression due to the differences in the number of X chromosomes: genetic females (XX) have two X chromosomes while genetic males (XY) have only one X chromosome (Lyon 1961; Wilson Sayres and Makova 2013). During X-inactivation, one of the X chromosomes in genetic females is silenced. Both paternally imprinted X-inactivation (the paternal X is preferentially silenced) and random X-inactivation have been observed. In mice, the extraembryonic lineages that ultimately give rise to the placenta and some extraembryonic membranes show paternally imprinted X-inactivation (Takagi and Sasaki 1975; K. D. Huynh and Lee 2001; Khanh D. Huynh and Lee 2005). Paternally imprinted X-inactivation has also been reported in rats (Wake, Takagi, and Sasaki 1976), cows (Xue et al. 2002), and marsupial mammals (Richardson, Czuppon, and Sharman 1971; Cooper et al. 1989; Al Nadaf et al. 2010). However, in mice, the embryonic lineages that ultimately give rise to the rest of the fetus exhibit random X-inactivation (Takagi and Sasaki 1975). Random X-inactivation has also been reported in mule and horse (Wang et al. 2012). In addition, X-inactivation in humans is thought to be random in most tissues, though the evidence for this is still a topic of debate (Ropers, Wolff, and Hitzeroth 1978; Looijenga et al. 1999; Moreira de Mello et al. 2010).

Understanding X-inactivation in the placenta can potentially lead to better understanding of pregnancy complications. The human placenta develops in the first several weeks of gastrulation (James, Carter, and Chamley 2012). The placenta, which is the genotype of the fetus, plays a critical role in pregnancy by regulating nutrition and protecting the developing fetus from the pregnant person's immune system (Gude et al. 2004). Improper placenta development can lead to complications such as preeclampsia and fetal growth restriction (Rathbun and Hildebrand 2019). Previous research has suggested an important role of X-inactivation in pregnancy complications. For example, Gong et al. (2018) found that the spermine synthase gene (SMS) that escapes X-inactivation in the placenta, but not in other tissues, may modulate fetal growth restriction, a common pregnancy complication (Gong et al. 2018). Therefore, understanding pattern of X-inactivation in this early-formed tissue is important to better understand its downstream developmental effects.

Most previous studies of X-inactivation in the placenta have been conducted in non-human species (Wang et al. 2012; Finn et al. 2014), or surveyed from very few SNPs and genes (Moreira de Mello et al. 2010; Looijenga et al. 1999). Therefore, while previous research suggested that X-inactivation is random in the human placenta, these studies cannot conclude whether X-inactivation in the human placenta is completely homogeneous where the same copy of the X chromosome is inactivated everywhere (**Figure 1A**). Or whether X-inactivation is organized into patches of cells with both maternally inactivated X patches and paternally inactivated X patches (**Figure 1B, C, and D**). Another possibility is that the inactivation in each cell is completely random, creating a mosaic pattern within the placenta (**Figure 1E**). In addition, the degree of variability in gene-specific escape from X-inactivation has not been explored within individuals or across individuals in the placenta.

To better understand patterns of X-inactivation in the human placenta, we sequenced whole-exome and whole-genome transcriptome of 30 term-placentas where the fetus is XX females. Importantly, we extracted and sequenced RNA from two sites at separate locations of the same placenta to explore heterogeneity within the placenta. We analyzed allele-specific expression to study X-inactivation. Taking all samples into consideration we observed evidence for large patches (clonal expansions) of X-inactivation in the placenta. Our observations are inconsistent with complete mosaicism in X-inactivation in the human placenta. We also observed that X-inactivation in the placenta is leaky; many variants on the X chromosome are not silenced completely, but do not show a strong signal of biallelic expression. Moreover, approximately half of genes that escape X-inactivation do not consistently escape across placentas, suggesting a high degree of variability, or ongoing evolution, in genes that escape X-inactivation. We concluded that patterns of X-inactivation in human placenta is more complex than previously thought. In addition, heterogeneity within and among placentas suggest careful consideration when interpreting results from bulk sequencing.

Methods

Samples

Working with the Yale Biobank, we identified one flash frozen DNA sample and two sites preserved in RNAlater from 30 placentas where the sex of the offspring was assigned female at birth; we confirmed the samples were genetically XX by presence of heterozygous sites across the X chromosome. We also sequenced whole exome and whole transcriptome from 12 placentas where the sex of the offspring was assigned male at birth. In 18 female placentas, we also sequenced DNA from the accompanying deciduas. The decidua is the genotype of the pregnant person. The placenta and decidua samples were collected immediately following birth from term (>= 39 weeks) uncomplicated pregnancies. Placentas were collected and sequenced at two different times, with 12 placenta samples in the first batch and 18 in the second (**Table S1**). From each placenta, from one flash frozen collection site, DNA was extracted, exome libraries were prepped and sequenced to approximately 50X coverage with 150bp paired-end sequence on the Illumina NextSeq at the Yale Genome Sequence Center. From each placenta, two separate sites were collected in RNAlater, RNA was extracted, RiboZero stranded libraries (RF) were prepared and sequenced to approximately 40 million reads per sample with 100bp paired end sequence on the Illumina NextSeq at the Yale Genome Sequence Center. We plotted the principal component for the exome data using the package *SNPRelate* in R (Zheng et al. 2012) and the multidimensional scaling plot for the RNA-seq data in **Figure S1**. We observed no separation by batch from the exome data (**Figure S1A**). We found that the whole-transcriptome data separated by batch along the first dimension (**Figure S1B**).

Exome sequence data processing

We used fastqc version 0.11.8 (Andrews 2010) and multiqc version 0.9 (Ewels et al. 2016) for visualizing quality. We trimmed adapters using bbduk as part of bbmap version 38.22 (Bushnell 2014) with the following parameters: qtrim=rl trimq=30 minlen=75 maq=20. Post trimming quality was checked using fastqc version 0.11.8 (Andrews 2010) and multiqc version 0.9 (Ewels

et al. 2016) (**Figure S2, A, B, and C)**. We used bwa-mem version 0.7.17 (Li 2013) to map to female-specific reference. Specifically, we mapped the exome samples to a sex chromosome complement informed reference genome in which the Y chromosome is hard-masked (to avoid mismapping of X-linked reads to homologous regions on the Y chromosome in the XX samples). To generate the sex chromosome complement reference genome we employed XYalign (Webster et al. 2019). XYalign created a Y-masked gencode GRCh38.p12 human reference genome for aligning XX individuals (Harrow et al. 2012). We used picard version 2.18.27 ("Picard Tools - By Broad Institute" n.d.) to mark PCR duplicates. To genotype variants, we used GATK version 4.1.0.0 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). We first used GATK's HaplotypeCaller to generate GVCF files. Second, we combined GVCF from 48 samples using GATK's CombineGVCFs (30 placenta samples and 18 decidua samples). Finally, we used GATK's GenotypeGVCFs to call variants. We obtained high-quality variants by filtering using GATK's Variant Quality Score Recalibration (VQSR) following GATK's best practice. We tabulated the number of variants before and after filtering with GATK's VQSR in **Table S2**.

RNA-seq data processing

Samples were checked for quality using fastqc version 0.11.8 (Andrews 2010) and multiqc version 0.9 (Ewels et al. 2016). Adapters were removed and sequences were trimmed both left and right for phred quality of 30, minimum length of 75 base pairs, and an average read quality of 20 or greater using bbduck version 38.22 (Bushnell 2014). Post trimming quality was checked using fastqc version 0.11.8 (Andrews 2010) and multiqc version 0.9 (Ewels et al. 2016) (**Figure S2, D, E, and F)**. Trimmed RNAseq reads were then aligned to a sex chromosome complement informed reference genome with the Y-masked (Webster et al. 2019; Olney et al. 2019) gencode GRCh38.p12 reference genome (Harrow et al. 2012). Total reads mapped and duplicate reads were visually checked using BAMtools stats (Barnett et al. 2011) (**Table S3)**.

Allele specific expression

We used GATK's ASEReadCounter (version 3.8.0) to obtain allele counts (McKenna et al. 2010). We set the minimum mapping quality to be 20, and minimum base quality to be 2. We removed duplicated reads with the flag --drf DuplicatedRead (Castel et al. 2015). We explored three thresholds for the parameter minimum depth: 10, 20, and 50. The number of heterozygous and expressed variants for each minimum depth threshold are shown in **Table S4**. As expected, as we increased the minimum depth threshold, the number of heterozygous and expressed variants decreases (**Table S4**). Since the patterns are similar between different threshold (**Figure S3**), we presented the results when using minimum depth of 20 throughout the main text.

Defining threshold for biased-allele expression

For each heterozygous and expressed variant, ASEReadCounter tabulates the number of reads (the count) for the reference allele and the alternate allele. The total read count is equal to the sum of the read count for the reference allele and the read count for the alternate allele. If the read count of the reference allele is greater than the read count of the alternate allele, we define the reference allele to be the biased allele, and vice versa. To determine the threshold for

biased-allele expression, we examined the allele balance on the X chromosome and an autosome, specifically chromosome 8 (**Figure S4**). Allele balance is defined as the ratio between the read count of the biased allele over the total read count. As expected, we observed that there are more variants with skewed ratio on the X chromosome as compared to chromosome 8. Some variants with skewed ratio are expected on the autosomes because of imprinted genes on the autosomes (Ishida and Moore 2013). We observed that the majority of the ratios on the X chromosome are greater than 0.8 while the majority of the ratio on chromosome 8 is less than 0.8 (**Figure S4**). Therefore, we chose 0.8 as a threshold where a ratio greater than 0.8 or less than 0.2 is taken as evidence for skewness in allele expression while a ratio between 0.2 and 0.8 is evidence of biallelic expression.

Phasing the X chromosome

To study whether patterns of X-inactivation in the placenta are completely homogeneous, patchy, or mosaic, we employed a phasing strategy where we defined that all of the variants skewed in the same direction are on the same haplotype (see **Supplementary Notes 1**). We defined "haplotype 1" to contain the biased allele where the allele balance (i.e the ratio between the read count of the biased allele and the total read count) is greater than 0.5. In cases where the allele balance is exactly 0.5, we picked either the reference allele or the alternate allele randomly with equal probability.

Since we have two sites, A and B, for each individual. We phase the X chromosome for each individual using biallelic and expressed variants for either site A or site B. For each collection site, we tabulated the number of heterozygous and expressed variants where the allele balance is greater than 0.8. Then, site A or B is chosen as the reference phasing if it is the one with the most skewed pattern of X-inactivation, defined as having more heterozygous and expressed variants where the allele balance is greater than 0.8 (**Supplementary Note 1**).

Determining whether the maternal or paternal X chromosome is silenced

In 18 of the placenta samples, we also sequenced whole exome of the decidua. Because the genotype of the decidua is the genotype of the pregnant person (Mori et al. 2016), we can leverage this information to determine which X chromosome, maternal or paternal, is silenced. For each heterozygous and expressed variants of the placenta, we can assign whether the biased allele (allele balance is greater or equal to 0.8) is paternal or maternal in origin if this variant is homozygous in the decidua. For example, if the genotype of the placenta is A/T and the genotype of the decidua is T/T, we assigned the T allele to be the maternal allele. If most of the biased alleles are the maternal allele, we designated the expressed haplotype to be the maternal haplotype, and vice versa.

Determining patterns of X chromosome inactivation per gene

To study this, we focused on genes for which there is at least one heterozygous variant that is expressed. For each of these genes, in each sample, we calculated the allele balance. In genes where there are more than one heterozygous and expressed variants, we obtained the allele balance summing across total read counts from all variants. For the calculation of X-inactivation status by gene we excluded the five samples that showed biallelic expression (**Figure 2:**

OBG0066, OBG0111, OBG0051, OBG0039, and OBG0156), resulting in a total of 55 samples. To confidently classify genes into “escape”, “silence”, or “variable escape”, we only consider a gene if it is expressed in a minimum of 10 (out of 55 samples). To leverage information across 55 samples, we classified a gene as confidently ”escape” if the gene shows biallelic expression (allele balance between 0.2 and 0.8) in at least 80% of the expressed samples. We classified a gene as confidently “silenced” if at most 20% of the expressed samples show biallelic expression (allele balance between 0.2 and 0.8). We classified a gene as exhibiting “variable escape” if more than 20% but less than 80% of the expressed samples show biallelic expression (allele balance between 0.2 and 0.8).

Coordinates used of PARs and XIST
As defined for GRCh38.p13 (“Human Genome Overview - Genome Reference Consortium” n.d.), we used the following coordinates for the pseudoautosomal region 1 (PAR1), PAR2, and XIST.
PAR1: position 10,001 to position 2,781,479 on the X chromosome
PAR2: position 155,701,383 to position 156,030,895 on the X chromosome
XIST: position 73,820,892 to position 73,851,867 on the X chromosome

Results

Evidence for completely homogeneous or patchy patterns of X-inactivation in the human placenta
We found evidence for either large patchiness, or completely homogenous X-inactivation in the human placenta. In 25/30 placentas both sites in the placenta exhibited extremely skewed X-inactivation, either with both extraction sites showing the same inactivated X chromosome (13 placentas; **Figure 2A**) or each site showing a single, but different, X chromosome inactivated (12 placentas; **Figure 2A**). In the remaining five placentas we observed one extraction site showing skewed X-inactivation, and the other showing both X chromosomes being expressed (**Figure 2A**). In this subset, we postulated that one of our samples was collected on the boundary of two different patches of X inactivated cells (**Figure 1D**). We never observed both sites in the same placenta exhibiting biallelic X-chromosome expression (**Figure 2A**). To validate that the patterns we are observing on the X chromosome are indeed X-inactivation, we repeated the same analyses on chromosome 8, where all samples showed biallelic expression (**Figure 2B**).

We observed that the inactivated X is mostly the paternal X chromosome. In 18/30 placentas, we also sequenced the whole exome of the decidua, allowing us to determine whether the maternal or paternal X chromosome is inactivated (see Methods). In 8/18 placenta samples where both extraction sites show the same inactivated X chromosome, it is the paternal X that is silenced (**Table S5**). In 3/18 placenta samples where one extraction site show skewed X-inactivation while the other extraction site show both X chromosomes being expressed, in two of these cases, the paternal X is silenced (**Table S5**). In 7/18 placenta samples where both extraction sites show the opposite X chromosome being activated, we also observed that at one extraction site, the maternal X is silenced and at the other extraction site, the paternal X is

silenced in 6/7 samples (**Table S5**). Overall, while we are underpowered to determine whether the paternal X or the maternal X that is silenced, we consistently observed the paternal X being silenced.

Evidence for a range of complete silencing, incomplete silencing, and escape

Despite evidence for extreme skewing or patchiness in the human placenta, we still observed incomplete silencing. To further gain insights into how X-inactivation varies across the X chromosome, we plotted the allele balance across the X chromosome (see Methods for how allele balance is measured). **Figure 3** shows a representation of a placenta from each category that was observed in **Figure 2**. Across the X chromosome, in the most highly skewed samples, we observed most variants show an allele balance greater than 0.8. This is in contrast to chromosome 8 of the autosomes, where we observed that most of the sites have an allele balance between 0.2 and 0.8 (**Figure S5**). However, across all samples, even those with the most highly skewed patterns of X-inactivation, we do not observe completely mono-allelic expression for all variants (**Figure 3; Figure S3, Table S6**). The percentage of variants showing completely mono-allelic expression (allele balance is 1) ranges from 11.8% to 65.5% (**Table S6**). The percentage of variants showing leaky silencing (0.8 ≤ allele balance < 1) ranges from 26.8% to 75.3% (**Table S6**). We replicated this analysis in 12 placentas from XY males, and showed that (if we falsely assumed males are biallelic in the nonPAR region), in these XY male samples we did observe almost exclusively mono-allelic expression in the nonPAR region of the X (**Figure S6**), suggesting that we are truly observing weak escape from the inactive X chromosome across many sites in XX individuals.

In addition to leaky silencing, we also observed individual instances of variants showing evidence for more extreme escape from X-inactivation. Here we defined a variant with biallelic expression if its allele balance is greater than 0.2 and less than 0.8 (0.2 < allele balance < 0.8). In all placenta samples, as expected, we observed biallelic expression of most variants in the pseudoautosomal regions at both extraction sites (grey boxes in **Figure 3; Figure S3**). We also observed individual variants showing biallelic expression in the nonPAR regions of the X chromosome, but the number of biallelic sites in the nonPAR regions varied. We found anywhere between 0 to 40% of the variants showing biallelic expression in the nonPARs region of the X chromosome (**Table S6).**

Heterogeneity in genes that escape X chromosome inactivation within and across individuals

We observed a large degree of heterogeneity in the genes that showed biallelic expression across all placenta samples. We identified 257 genes with at least one heterozygous and expressed variant in each gene. 19 genes (~7%) escape XCI across all samples and 114 genes (~44%) are silenced across all samples (**Figure 4**). 124 genes (~49%) show a variable degree of escaping, ranging from 3% to 97%, indicating that a gene that escape XCI in one sample does not necessarily escape XCI in other samples (**Figure 4**). Therefore, there is a large degree of heterogeneity in genes that escape XCI in the placenta.

We also observed tremendous heterogeneity in the genes that showed biallelic expression both between, and perhaps surprisingly, within, placentas. Because a gene that shows evidence for escaping XCI does not necessarily escape XCI across all samples (**Figure 4**), we wanted to confidently identify genes that are escaping inactivation. To do this, we employed a stringent criteria where we categorized genes across the X chromosome into three categories based on comparisons across placentas: 1) Escape, genes that show biallelic expression in greater than 80% of the samples; 2) Silenced, genes that show biallelic expression in fewer than 20% of placenta samples; and, 3) Variable escape, genes that showed biallelic expression in between 20-80% of samples. Since we could not identify genes that escape and genes that are inactivated using the five placentas that show biallelic expression, we excluded these samples (**Figure 2**). Additionally, we only considered a gene if there are at least ten samples with expression data for that gene. Using this definition, we confidently identified 18 “escape” genes, 32 “variable escape” genes, and 67 “silenced” genes (**Figure 5; Figure S7**). Even with these stringent criteria, we still observe heterogeneity in genes that escape XCI and in genes that are silenced (**Figure 5**).

We observed a high degree of heterogeneity in escape within the same placenta. For variants that are shared between site A and site B of the same placenta, we tabulated the proportion where the allele balance is showing the same pattern between the two sites, either both are monoallelic or both are biallelic. Even in samples where the two sites are expressing the same haplotype, the proportion of variants that show consistent patterns between the two sites ranges from 82% to 99% (**Table S7**). The proportion of variants that show consistent patterns between the two sites is even less for samples where site A is monoallelic but site B is biallelic or where site A and site B are monoallelic for opposite haplotypes (**Table S7**). Further, we observed instances where the two sites of the same placenta showed an opposite pattern for the same gene (**Figure 5**). For example, in placenta OBG0175, the gene PUDP is silenced at site A but escapes inactivation at site B (**Figure 5A**). Overall, we observed 15 instances of escape genes and 35 instances of silenced genes showing an opposite pattern between two extraction site of the same placenta. These results suggest a large degree of heterogeneity within the same placenta.

Discussion

We analyzed allele specific expression of 30 XX female placentas and found that the human placenta does not exhibit mosaic X-inactivation; rather, it exhibits large patches of maternal or paternal X chromosome expression or only one X chromosome being expressed. We can rule out the human placenta being completely mosaic because we never observed both sites of the same placenta showing both X chromosomes being expressed (**Figure 2, Figure 3, Figure S3**). We observed that in placentas where both extraction sites show the same inactivated X chromosome, it is always the paternal X that is silenced (**Table S5**), suggesting that the human placenta could be paternally imprinted. Paternally imprinted X-inactivation has been reported in some tissues in other placental mammals such as mouse, rat, and cow (Takagi and Sasaki

1975; K. D. Huynh and Lee 2001; Khanh D. Huynh and Lee 2005; Wake, Takagi, and Sasaki 1976; Xue et al. 2002). In addition, since X-inactivation in marsupial mammals is paternally imprinted, these results suggest that X-inactivation in marsupial mammals and placental mammals could have a common origin (Heard and Disteche 2006). The second possibility is that the human placenta is patchy in all individuals, but that we don't have the resolution to observe this in our sample. In 13/30 placentas where both extraction sites show inactivation for the same X chromosome, it could be that we were sampling both extraction sites from patches with the same inactivated X chromosome (**Figure 1B**), but that other patches existed.

Our observation that X-inactivation is patchy is in agreement with previous analyses of individual genes. In a subset of nine adjacent XX female full term placenta samples, where methylation in the androgen receptor gene was used as the assay, three placentas exhibited predominantly maternal X-inactivation, one exhibited paternal X-inactivation, and the remaining 5 exhibited both maternal and paternal alleles (Looijenga et al. 1999). Curiously, we never observed only maternal X being inactivated. In a different study, with at least 5 informative SNPs used to infer X-inactivation status in a single site collected from 22 placentas, the authors similarly observe some placentas have maternal X-inactivation, some have paternal X-inactivation, and others show biallelic expression across the X-linked SNPs (Moreira de Mello et al. 2010). From this, the authors proposed that the human placenta has large patches of maternal or paternal X-inactivation, but they were not able to confirm. With improved sampling and sequencing strategies, we can confirm that patterns of X-inactivation is patchy, at least in some individuals. Our results show that patterns of X inactivation in human placenta is more complex than previously thought: X-inactivation in human placenta could vary among individuals where it is completely homogeneous in some individuals but patchy in other individuals. Therefore, additional assessment of intra-placenta variation is warranted to better understand the patterns of X-inactivation across the human placenta.

By analyzing whole transcriptome data, we extend previous work on X-inactivation in the placenta to find a wide range of the degree of silencing in the human placenta. *A priori* expectation would be that the allele balance of most variants should be 1. However, the original observation of heterogeneity in escape from X-inactivation in the placenta observed that the extent of inactivation could vary between samples (Carrel and Willard 2005). Similarly, we observed instances of variants that are silenced incompletely, where the allele balance is greater than 0.8 but less than 1 (**Figure 3, Figure S3, Table S6**). Our results indicate that in addition to variants that escape X-inactivation entirely, some variants are not completely silenced. It is not known what the effect of partial silencing is on the X chromosome. We suggest the process of X-inactivation is complex, including variants that are silenced completely, variants that are partially silenced, and variants that are not silenced.

We observed a large degree of heterogeneity in genes that escape X-inactivation. We found that 49% of the genes on the X chromosome exhibit variable escape in inactivation across samples (**Figure 4**). However, when we applied a stringent threshold to classify a gene, we can confidently category 117 genes where 32/117 (27%) show variable escape (**Figure 5B**). Our finding is similar to Carrel and Willard (2005), who surveyed 94 genes on the X chromosome

across 40 human fibroblast cell lines; they found a smaller proportion of variable escape genes (20%) (Carrel and Willard 2005). Similarly, Cotton et al. (2013) who surveyed 508 genes in clonal or near-clonal lymphoblastoid cell lines and fibroblast cell lines found 146 genes (29%) with variable escape (Cotton et al. 2013). 3/32 genes (PLS3, HCFC1, and TIMP1) that we identified as variable escape were also found in Carrel and Willard (Carrel and Willard 2005) or Cotton et al. (Cotton et al. 2013). 11/32 variable escape genes in this study showed opposite patterns between Carrel and Willard (Carrel and Willard 2005) and Cotton et al. (Cotton et al. 2013) (**Table S9**). The difference in estimates of heterogeneity in X-inactivation could be due to differences in tissue of origin, differences between primary tissue and cell lines, or in the sets of genes assayed. Interestingly, TIMP1, which was confirmed to be showing variable escape patterns (Anderson and Brown 1999), also showed a signature of variable escape in our data (**Figure 5B**).

We found that 44% of the genes we could assay are silenced across all samples and only 7% of the genes escape (or are never subject to) X-inactivation across all samples (**Figure 4**). To confidently identify genes that escape X-inactivation, we used a stringent threshold to identify 18 genes that escape in at least 10 samples (**Figure 5A**). As expected, many (8/18) escaping genes are found in the pseudoautosomal regions. 1/18 escaping genes have not been identified in previous studies (Carrel and Willard 2005; Cotton et al. 2013): PUDP is a protein encoded gene but the protein has no annotated biological function (GeneCards Human Gene Database n.d.). Notably, genes that escape inactivation have been implicated as exhibiting stronger purifying selection, and being expressed across more tissues than genes that are subject to silencing (Slavney et al. 2016). **Table S8** summarizes what we know about these escape genes in terms of protein function and relevant pathways. Many of these genes, when dysregulated, are involved in human disease, but the role of inactivation in those processes has not been thoroughly studied. We identified 67 genes that are silenced in at least 10 samples (**Figure 5C**). 13 of these 67 genes that we defined as silencers are in disagreement with the results from Carrel and Willard or Cotton et al. (**Table S9**) (Carrel and Willard 2005; Cotton et al. 2013). The discrepancies between our results and Cotten et al. or Carrel and Willard could be due to differences in tissues. It is plausible that the proportion of genes that are constitutively silenced or escape X-inactivation varies across tissues. In addition, to confidently identify genes that escape, we suggest including multiple individuals because of the high degree of heterogeneity.

Our findings that the human placenta is either completely homogenous for the inactivated X or patchy or both suggest that patterns of X-inactivation in the human placenta is complex and could potentially differ across individuals. Further, we detected heterogeneity within the same placenta. We found 15 instances of escape genes and 35 instances of silenced genes that show an opposite pattern between two extraction sites from the same placenta (**Figure 5, A and B**). Since the DNA of the placenta is the DNA of the fetus, the placenta is the tissue where samples are taken for prenatal screening/diagnostic tests during pregnancy to understand conditions relating to the fetus's genetics and chromosome abnormalities (Carlson and Vora 2017). Because of the high level of variability in patterns of X-inactivation within and among placentas, heterogeneity should be taken into consideration in interpreting genetic data from the placenta.

A challenge of all future studies will be the limited heterogeneity in humans; because we are relying on allele specific expression to study genes that escape X-inactivation, we are limited in the number of genes that have at least one heterozygous and expressed variant. While previous studies used female/male ratio in gene expression as a proxy to identify genes that escape X-inactivation, we found that this metric does not accurately predict genes that escape X-inactivation in the placenta (**Figure S8, S9, Supplementary Note 2**). Therefore, we suggest that allele specific expression is still the most optimal way to study X-inactivation, at least in the human placenta. Future studies could sequence more samples to increase the power to genotype more variants and therefore increase the total number of genes that can be studied.

In this study, we found that pattern of X-inactivation in the human placenta is complex. In addition to patchiness in which X is inactivated in each region of the placenta, we found a large range in the extent of silencing, both within an individual sample, and heterogeneity in which genes are silenced across samples. We identified 7% of genes on the X chromosome that escape X-inactivation while most genes on the X chromosome show a heterogeneous pattern of silencing, ranging from leaky silencing to biallelic expression in a pattern that varied from individual to individual.

Acknowledgements

We acknowledge Research Computing at Arizona State University for providing high-performance computing and storage resources that have contributed to the research results reported within this paper. URL: http://www.researchcomputing.asu.edu.
We are very thankful for Lauren Perley at the Yale Biobank for help in identifying samples. We would also like to thank Heini Natri and Wilson lab members for helpful feedback on the manuscript. This work was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) grant R35GM124827 to MAW. We are thankful for support from the Reproductive and Placental Research Unit, Yale University School of Medicine to HJK. KCO was additionally supported as an ARCS Spetzler Scholar.

Data and Code availability

Data is available at PRJNA530785: https://www.ncbi.nlm.nih.gov/sra/PRJNA530785.
Code is available on GitHub at: https://github.com/SexChrLab/PlacentaSexDiff.

Figures

Figure 1. Schematic of sampling and possible patterns of X-inactivation. In each individual's placenta, we extracted RNA-seq from two separate locations which we called site A and site B. In this schematic, yellow circles denote cells where "haplotype 1" is expressed and blue circles denote cells where "haplotype 2" is expressed. We outlined all possible scenarios that we observed in the data and what patterns of X-inactivation it suggests. (A)-(B): we observe that both sites expresse "haplotype 1" and this could indicate that the placenta is completely homogeneous (A) or that the placenta is patchy but we sampled both sites with the same haplotype (B). (C): we observe that site A expresses "haplotype 1" but site B expresses "haplotype 2". This pattern is consistent with a patchy pattern of X-inactivation. (D): we observe that site A expresses "haplotype 1" and site B expresses both haplotypes. This pattern is also consistent with a patchy pattern of XCI, but that we sampled site B at the boundary between the two patches. (E): we observe that both sites express both haplotypes, consistent with a mosaic pattern of X-inactivation, where the placenta is made of cells with random inactivation of the maternal or the paternal X chromosome.

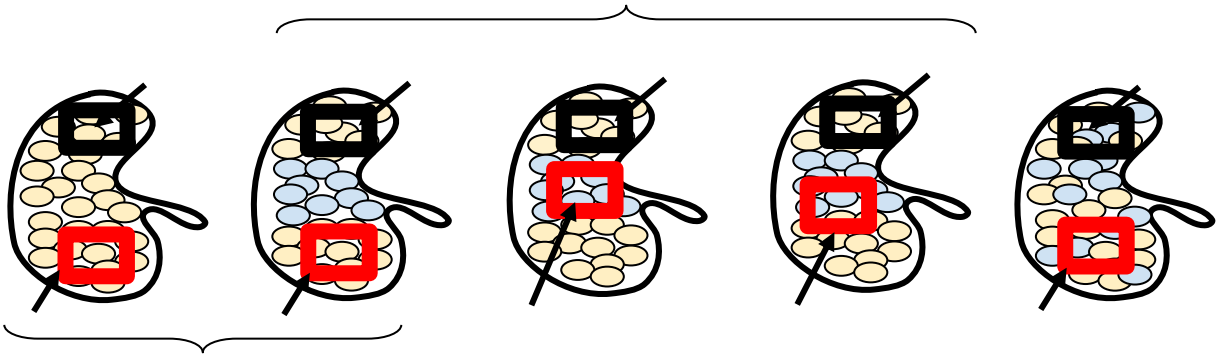


Figure 2. Maternal and paternal inactivation suggests patchy X-inactivation across the placenta. Allele balance aggregated across the X chromosome for each placenta sample is shown for site A (open black circle) and site B (filled red triangle). Each point is the median value. Median was used to minimize the effect of outliers. (A) X chromosome. We removed variants that fall within the pseudoautosomal regions (PARs) prior to computing medians. (B) Chromosome 8. The dotted gray horizontal lines denote allele balance of 0.2 and 0.8. Placenta samples are ordered from left to right: both sites expressing the same haplotype (**Figure 1A, 1B**), site A expresses "haplotype 1" but site B expresses both haplotypes (**Figure 1D**), and both sites express opposite haplotype (**Figure 1C**).

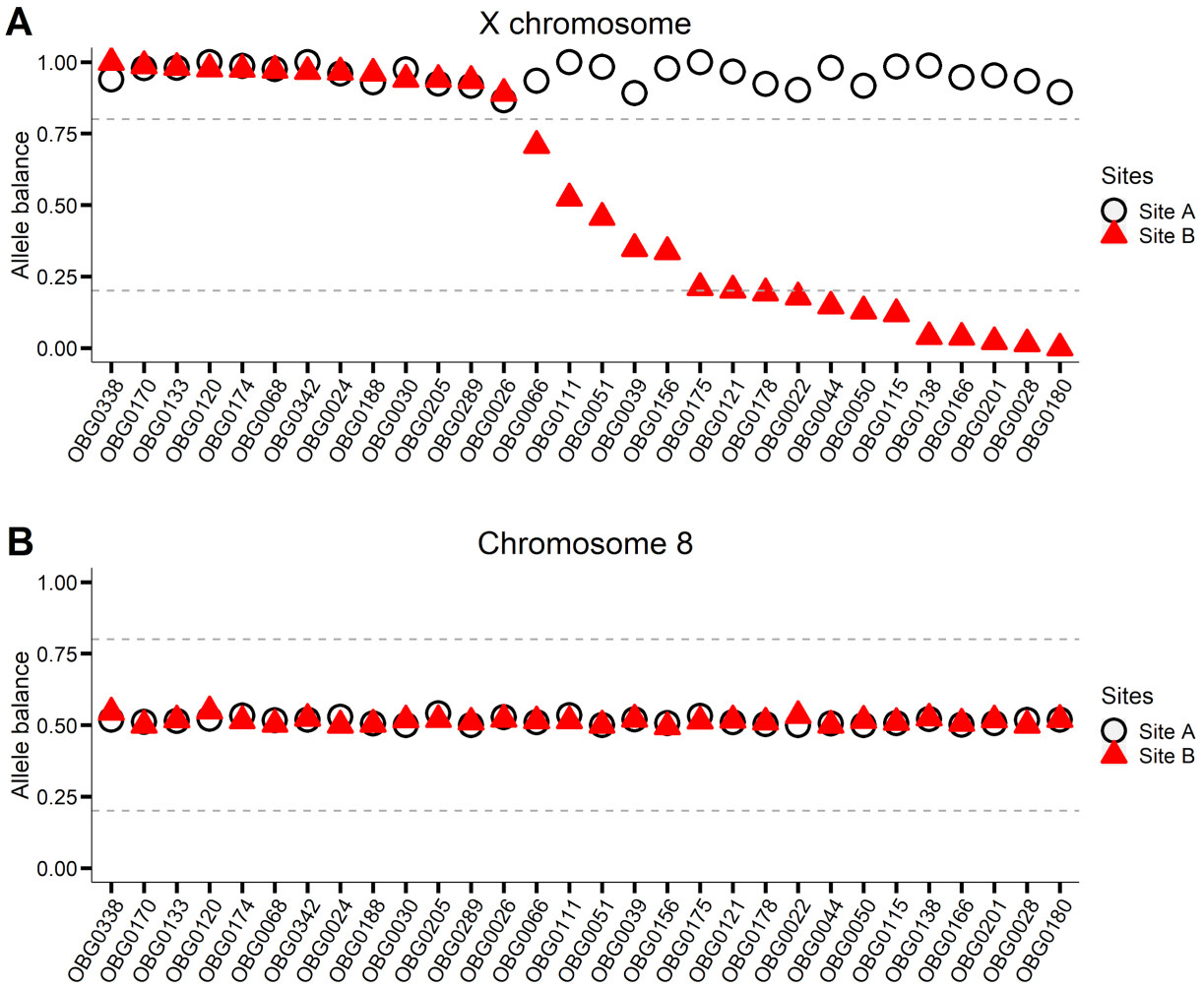
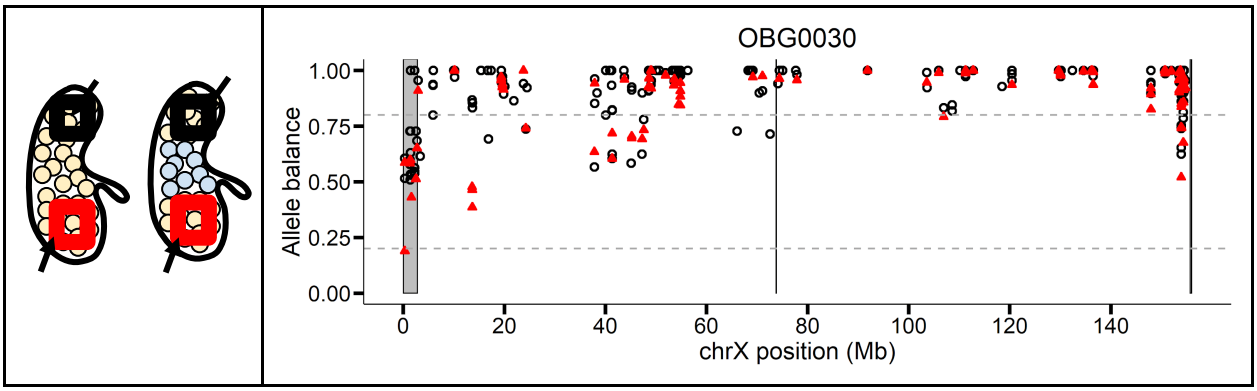
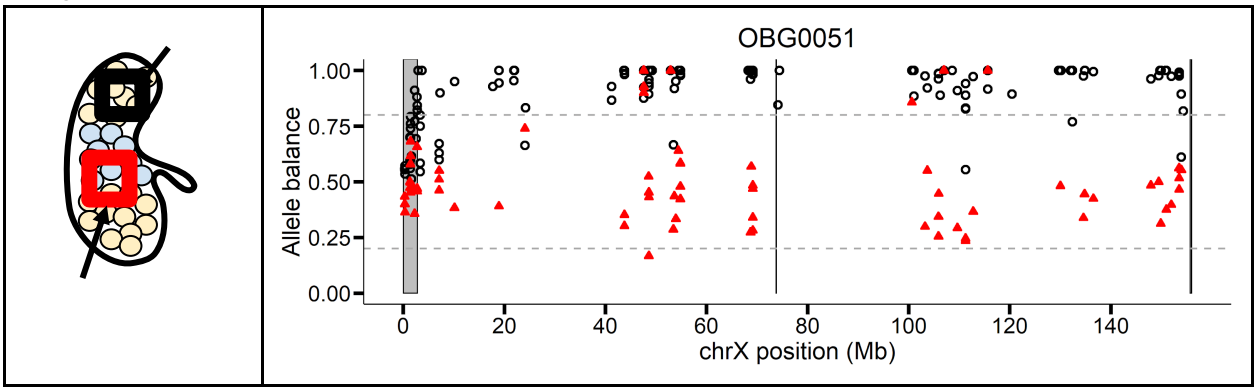


Figure 3. Patterns of X-inactivation across the entire X chromosome. In each plot, allele balance at each heterozygous and expressed variant is plotted as a function of the position on the X chromosome. Open black circles denote variants on extraction site A. Filled red triangles denote variants on extraction site B. Gray boxes denote the pseudoautosomal regions and XIST.

A. Both extraction sites show the same inactivated X chromosome



B. One extraction site shows skewed X-inactivation and the other shows both X chromosomes being expressed



C. Each site shows different X chromosome inactivated

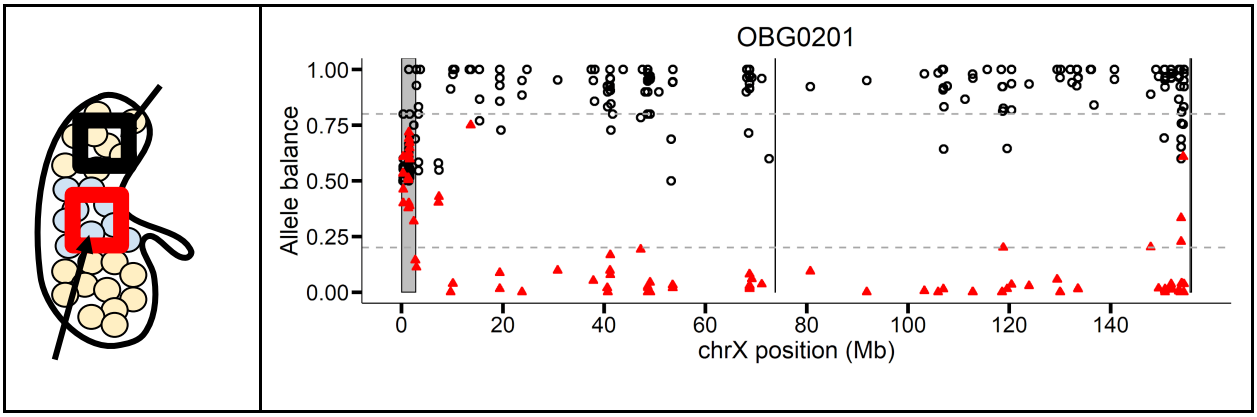


Figure 4. Heterogeneity in proportion of genes that escape X-inactivation or are silenced. Across 60 samples, there are a total of 257 genes where there is at least one heterozygous and expressed variant in each gene (X-axis). For each gene, the purple bar denotes the proportion of samples that show evidence for that gene escaping X-inactivation. The blue bar denotes the proportion of samples that show evidence for that gene being silenced. Across all genes with at least one heterozygous and expressed variant, 19 (7%) escape in all samples, 114 (44%) are silenced in all samples, and the remaining 124 (49%) exhibit heterogeneity in escape/silencing across samples.

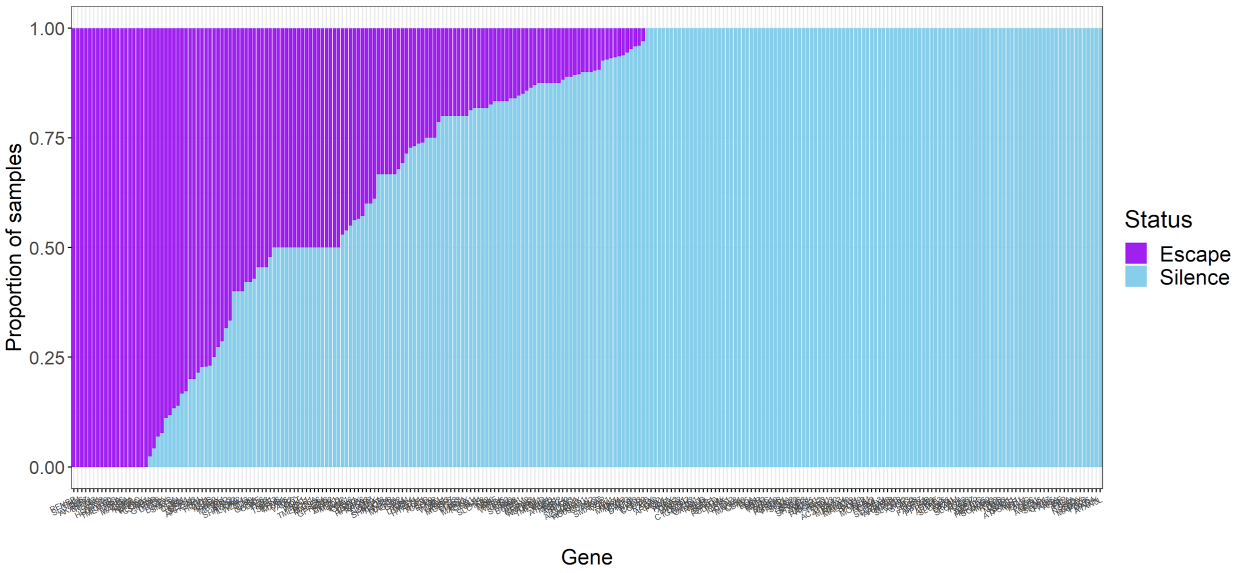


Figure 5. Heterogeneity in escape from X chromosome inactivation across and within placentas.

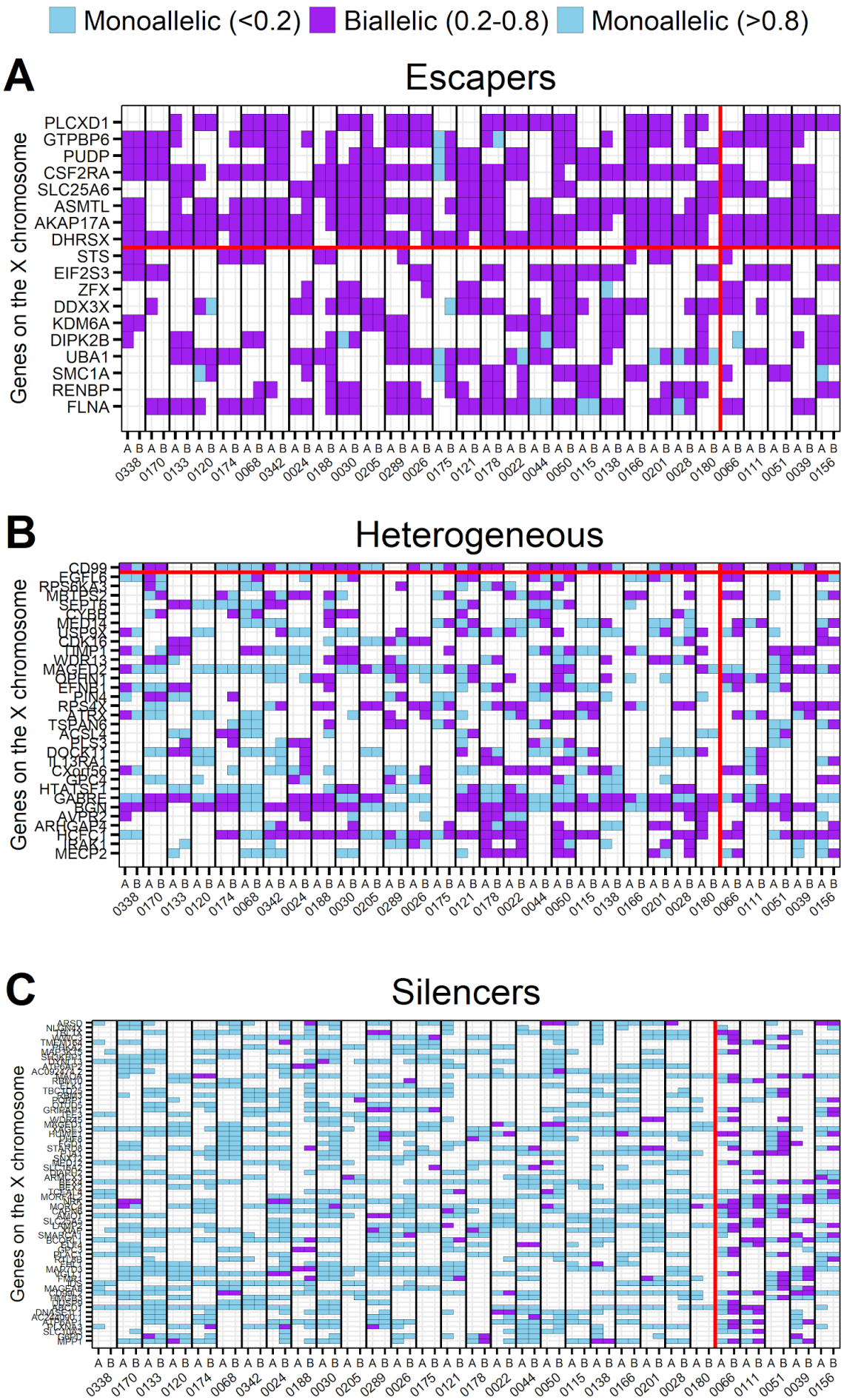
(A). We defined genes that are “escapers” are the genes where there are at least 10 samples containing at least one heterozygous and expressed variant and where more than 80% of those samples show biallelic expression (allele balance between 0.2 and 0.8).

(B). We defined genes that are “heterogeneous” are the genes where there are at least 10 samples containing at least one heterozygous and expressed variant and where between 20% and 80% of those samples show biallelic expression (allele balance between 0.2 and 0.8).

(C). We defined genes that are “silencers” are the genes where there are at least 10 samples containing at least one heterozygous and expressed variant and where less than 20% of those samples show biallelic expression (allele balance between 0.2 and 0.8).

The X axis is ordered similarly to Figure 2 with one exception: the five placentas where site A shows skewed towards haplotype 1 but site B shows biallelic expression are plotted at the end (separated with the other samples by the red vertical line). We order the Y axis by position on the X chromosome. The red horizontal line separates the pseudoautosomal regions from the non-pseudoautosomal regions. The black vertical lines separate the two sites from the same placenta for clarity.

The purple cells denote biallelic expression (allele balance between 0.2 and 0.8). The blue bars denote monoallelic expression (allele balance less than 0.2 or greater than 0.8).


Monoallelic (<0.2)
Biallelic (0.2-0.8)
Monoallelic (>0.8)
A
Escapers
Genes on the X chromosome
PLCXD1
GTPBP6
PUDP
CSF2RA
SLC25A6
ASMTL
AKAP17A
DHRSX
STS
EIF2S3
ZFX
DDX3X
KDM6A
DIPK2B
UBA1
SMC1A
RENBP
FLNA
B
Heterogeneous
Genes on the X chromosome
C
Silencers
Genes on the X chromosome
A B A B A B

References

Al Nadaf, Shafagh, Paul D. Waters, Edda Koina, Janine E. Deakin, Kristen S. Jordan, and Jennifer Am Graves. 2010. “Activity Map of the Tammar X Chromosome Shows That Marsupial X Inactivation Is Incomplete and Escape Is Stochastic.” *Genome Biology* 11 (12): R122.

Anderson, C. L., and C. J. Brown. 1999. “Polymorphic X-Chromosome Inactivation of the Human TIMP1 Gene.” *American Journal of Human Genetics* 65 (3): 699–708.

Andrews, S. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data. Available Online.”

Barnett, Derek W., Erik K. Garrison, Aaron R. Quinlan, Michael P. Strömberg, and Gabor T. Marth. 2011. “BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files.” *Bioinformatics* 27 (12): 1691–92.

Bushnell, Brian. 2014. “BBMap: A Fast, Accurate, Splice-Aware Aligner.” LBNL-7065E. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner.

Carlson, Laura M., and Neeta L. Vora. 2017. “Prenatal Diagnosis: Screening and Diagnostic Tools.” *Obstetrics and Gynecology Clinics of North America* 44 (2): 245–56.

Carrel, Laura, and Huntington F. Willard. 2005. “X-Inactivation Profile Reveals Extensive Variability in X-Linked Gene Expression in Females.” *Nature* 434 (7031): 400–404.

Castel, Stephane E., Ami Levy-Moonshine, Pejman Mohammadi, Eric Banks, and Tuuli Lappalainen. 2015. “Tools and Best Practices for Data Processing in Allelic Expression Analysis.” *Genome Biology* 16 (September): 195.

Cooper, D. W., P. G. Johnston, J. L. Vandeberg, and E. S. Robinson. 1989. “X-Chromosome Inactivation in Marsupials.” *Australian Journal of Zoology* 37 (3): 411–17.

Cotton, Allison M., Bing Ge, Nicholas Light, Veronique Adoue, Tomi Pastinen, and Carolyn J. Brown. 2013. “Analysis of Expressed SNPs Identifies Variable Extents of Expression from the Human Inactive X Chromosome.” *Genome Biology* 14 (11): R122.

DePristo, Mark A., Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data.” *Nature Genetics* 43 (5): 491–98.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. “MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report.” *Bioinformatics* 32 (19): 3047–48.

Finn, Elizabeth H., Cheryl L. Smith, Jesse Rodriguez, Arend Sidow, and Julie C. Baker. 2014. “Maternal Bias and Escape from X Chromosome Imprinting in the Midgestation Mouse Placenta.” *Developmental Biology* 390 (1): 80–92.

GeneCards Human Gene Database. n.d. “PUDP Gene - GeneCards | HDHD1 Protein | HDHD1 Antibody.” Accessed September 5, 2019. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PUDP.

Gong, Sungsam, Ulla Sovio, Irving Lmh Aye, Francesca Gaccioli, Justyna Dopierala, Michelle D. Johnson, Angela M. Wood, et al. 2018. “Placental Polyamine Metabolism Differs by Fetal Sex, Fetal Growth Restriction, and Preeclampsia.” *JCI Insight* 3 (13). https://doi.org/10.1172/jci.insight.120723.

Gude, Neil M., Claire T. Roberts, Bill Kalionis, and Roger G. King. 2004. “Growth and Function of the Normal Human Placenta.” *Thrombosis Research* 114 (5-6): 397–407.

Harrow, Jennifer, Adam Frankish, Jose M. Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L. Aken, et al. 2012. “GENCODE: The Reference Human Genome Annotation for The ENCODE Project.” *Genome Research* 22 (9): 1760–74.

Heard, Edith, and Christine M. Disteche. 2006. “Dosage Compensation in Mammals: Fine-Tuning the Expression of the X Chromosome.” *Genes & Development* 20 (14): 1848–67.
“Human Genome Overview - Genome Reference Consortium.” n.d. Accessed September 5, 2019. https://www.ncbi.nlm.nih.gov/grc/human.
Huynh, K. D., and J. T. Lee. 2001. “Imprinted X Inactivation in Eutherians: A Model of Gametic Execution and Zygotic Relaxation.” *Current Opinion in Cell Biology* 13 (6): 690–97.
Huynh, Khanh D., and Jeannie T. Lee. 2005. “X-Chromosome Inactivation: A Hypothesis Linking Ontogeny and Phylogeny.” *Nature Reviews. Genetics* 6 (5): 410–18.
Ishida, Miho, and Gudrun E. Moore. 2013. “The Role of Imprinted Genes in Humans.” *Molecular Aspects of Medicine* 34 (4): 826–40.
James, J. L., A. M. Carter, and L. W. Chamley. 2012. “Human Placentation from Nidation to 5 Weeks of Gestation. Part I: What Do We Know about Formative Placental Development Following Implantation?” *Placenta* 33 (5): 327–34.
Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.
Looijenga, L. H., A. J. Gillis, A. J. Verkerk, W. L. van Putten, and J. W. Oosterhuis. 1999. “Heterogeneous X Inactivation in Trophoblastic Cells of Human Full-Term Female Placentas.” *American Journal of Human Genetics* 64 (5): 1445–52.
Lyon, M. F. 1961. “Gene Action in the X-Chromosome of the Mouse (Mus Musculus L.).” *Nature* 190 (April): 372–73.
McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303.
Moreira de Mello, Joana Carvalho, Erica Sara Souza de Araújo, Raquel Stabellini, Ana Maria Fraga, Jorge Estefano Santana de Souza, Denilce R. Sumita, Anamaria A. Camargo, and Lygia V. Pereira. 2010. “Random X Inactivation and Extensive Mosaicism in Human Placenta Revealed by Analysis of Allele-Specific Gene Expression along the X Chromosome.” *PloS One* 5 (6): e10947.
Mori, Mayumi, Agnes Bogdan, Timea Balassa, Timea Csabai, and Júlia Szekeres-Bartho. 2016. “The Decidua-the Maternal Bed Embracing the Embryo-Maintains the Pregnancy.” *Seminars in Immunopathology* 38 (6): 635–49.
Olney, Kimberly C., Sarah M. Brotman, Valeria Valverde-Vesling, Jocelyn Andrews, and Melissa A. Wilson. 2019. “‘Aligning RNA-Seq Reads to a Sex Chromosome Complement Informed Reference Genome Increases Ability to Detect Sex Differences in Gene Expression.’” *bioRxiv*. https://doi.org/10.1101/668376.
“Picard Tools - By Broad Institute.” n.d. Accessed September 5, 2019. http://broadinstitute.github.io/picard/.
Rathbun, Kimberly M., and Jason P. Hildebrand. 2019. “Placenta Abnormalities.” In *StatPearls*. Treasure Island (FL): StatPearls Publishing.
Richardson, B. J., A. B. Czuppon, and G. B. Sharman. 1971. “Inheritance of Glucose-6-Phosphate Dehydrogenase Variation in Kangaroos.” *Nature: New Biology* 230 (13): 154–55.
Ropers, H. H., G. Wolff, and H. W. Hitzeroth. 1978. “Preferential X Inactivation in Human Placenta Membranes: Is the Paternal X Inactive in Early Embryonic Development of Female Mammals?” *Human Genetics* 43 (3): 265–73.
Slavney, Andrea, Leonardo Arbiza, Andrew G. Clark, and Alon Keinan. 2016. “Strong Constraint on Human Genes Escaping X-Inactivation Is Modulated by Their Expression Level and Breadth in Both Sexes.” *Molecular Biology and Evolution* 33 (2): 384–93.
Takagi, N., and M. Sasaki. 1975. “Preferential Inactivation of the Paternally Derived X Chromosome in the Extraembryonic Membranes of the Mouse.” *Nature* 256 (5519): 640–

42.
Tukiainen, Taru, Alexandra-Chloé Villani, Angela Yen, Manuel A. Rivas, Jamie L. Marshall, Rahul Satija, Matt Aguirre, et al. 2017. “Landscape of X Chromosome Inactivation across Human Tissues.” *Nature* 550 (7675): 244–48.
Van der Auwera, Geraldine A., Mauricio O. Carneiro, Chris Hartl, Ryan Poplin, Guillermo Del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. “From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 43: 11.10.1–33.
Wake, N., N. Takagi, and M. Sasaki. 1976. “Non-Random Inactivation of X Chromosome in the Rat Yolk Sac.” *Nature* 262 (5569): 580–81.
Wang, Xu, Donald C. Miller, Andrew G. Clark, and Douglas F. Antczak. 2012. “Random X Inactivation in the Mule and Horse Placenta.” *Genome Research* 22 (10): 1855–63.
Webster, Timothy H., Madeline Couse, Bruno M. Grande, Eric Karlins, Tanya N. Phung, Phillip A. Richmond, Whitney Whitford, and Melissa A. Wilson. 2019. “Identifying, Understanding, and Correcting Technical Artifacts on the Sex Chromosomes in next-Generation Sequencing Data.” *GigaScience* 8 (7). https://doi.org/10.1093/gigascience/giz074.
Wilson Sayres, Melissa A., and Kateryna D. Makova. 2013. “Gene Survival and Death on the Human Y Chromosome.” *Molecular Biology and Evolution* 30 (4): 781–87.
Xue, Fei, X. Cindy Tian, Fuliang Du, Chikara Kubota, Maneesh Taneja, Andras Dinnyes, Yunping Dai, Howard Levine, Lygia V. Pereira, and Xiangzhong Yang. 2002. “Aberrant Patterns of X Chromosome Inactivation in Bovine Clones.” *Nature Genetics* 31 (2): 216–20.
Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. “A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data.” *Bioinformatics* 28 (24): 3326–28.