Title

X chromosome inactivation in the human placenta is patchy and distinct from adult tissues

Authors and Affiliations

Tanya N. Phung[1,2], Kimberly C. Olney[1,2], Michelle Silasi[3], Lauren Perley[4], Jane O'Bryan[4], Harvey J. Kliman[4], and Melissa A. Wilson[1,2,5]

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 Maternal-Fetal Medicine, Mercy Hospital St. Louis, St. Louis, MO 63141
4. Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT 06520
5. The Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe AZ 85282 USA

Corresponding Author

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

Abstract

One of the X chromosomes in genetic females is silenced by a process called X chromosome inactivation (XCI). Variation in XCI across the placenta may contribute to observed sex differences and variability in pregnancy outcomes. However, XCI has predominantly been studied in human adult tissues. Here we sequenced and analyzed DNA and RNA from two locations from 30 full-term pregnancies. Implementing an allele specific approach to examine XCI, we report evidence that XCI in the human placenta is patchy, with large patches of either silenced maternal or paternal X chromosomes. Further, using similar measurements, we show that this is in contrast to adult tissues, which generally exhibit mosaic X-inactivation, where bulk samples exhibit both maternal and paternal X chromosome expression. Further, by comparing skewed samples in placenta and adult tissues, we identify genes that are uniquely silenced or expressed in the placenta compared to adult tissues highlighting the need for tissue-specific maps of XCI.

Introduction

X chromosome inactivation (XCI) evolved in eutherian mammals, at the same time as the development of invasive placentation, and is important for many biological processes (Natri et al. 2019). XCI is a mechanism to regulate the dosage of gene expression due to the differences in the number of X chromosomes between the sexes: genetic females (XX) have two X chromosomes while genetic males (XY) have only one X chromosome (Lyon 1961; Wilson Sayres and Makova 2013).

There are two levels of XCI. The first level is at the whole chromosome level, where either the maternal X or the paternal X is chosen for inactivation. The choice of which X is inactivated has been observed to happen in one of two ways: random XCI (the maternal and paternal X are silenced with equal probability), and imprinted XCI (the paternal X is inactivated). In mice, the extraembryonic lineages that ultimately give rise to the placenta and some extraembryonic membranes show paternally imprinted XCI (Takagi and Sasaki 1975; K. D. Huynh and Lee 2001; Khanh D. Huynh and Lee 2005). Paternally imprinted XCI 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 XCI (Takagi and Sasaki 1975). Random XCI has also been reported in mule and horse placenta (Wang et al. 2012).

The second level of XCI concerns which individual genes are subject to inactivation (silenced or not expressed) on the inactive X chromosome. Previous studies have aimed to determine what genes on the X chromosome escape XCI in a variety of tissues (see (Carrel and Brown 2017) for an overview of expression from the inactive X). Carrel and Willard (2005) studied biallelic expression in primary human fibroblast cell lines and rodent/human somatic hybrids (Carrel and Willard 2005). Cotton et al. (2015) used DNA methylation data to characterize escape status in 27 adult tissues (Cotton et al. 2015). Gene-specific silencing has also been shown to vary between tissues and individuals in mouse and human (Berletch et al. 2015; Tukiainen et al. 2017). However, previous cross-tissue studies on gene-specific escape from XCI did not include the placenta.

The human placenta is a transient tissue that is formed early in pregnancy and develops from the outer layer of the pre-implantation embryos (James, Carter, and Chamley 2012; Turco and Moffett 2019). The placenta has the genotype of the fetus and 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). It is important to study gene-specific escape in the placenta because XCI could play an important role in pregnancy complications; Gong et al. (2018) found that the spermine synthase gene (SMS) that modulates fetal growth restriction, a common pregnancy complication may escape XCI in the placenta but not in other tissues (Gong et al. 2018). Further, the degree of XCI skewing could

be associated with pregnancy loss (Sui, Chen, and Sun 2015). Therefore, understanding XCI in this early-formed tissue is critical for understanding downstream developmental effects.

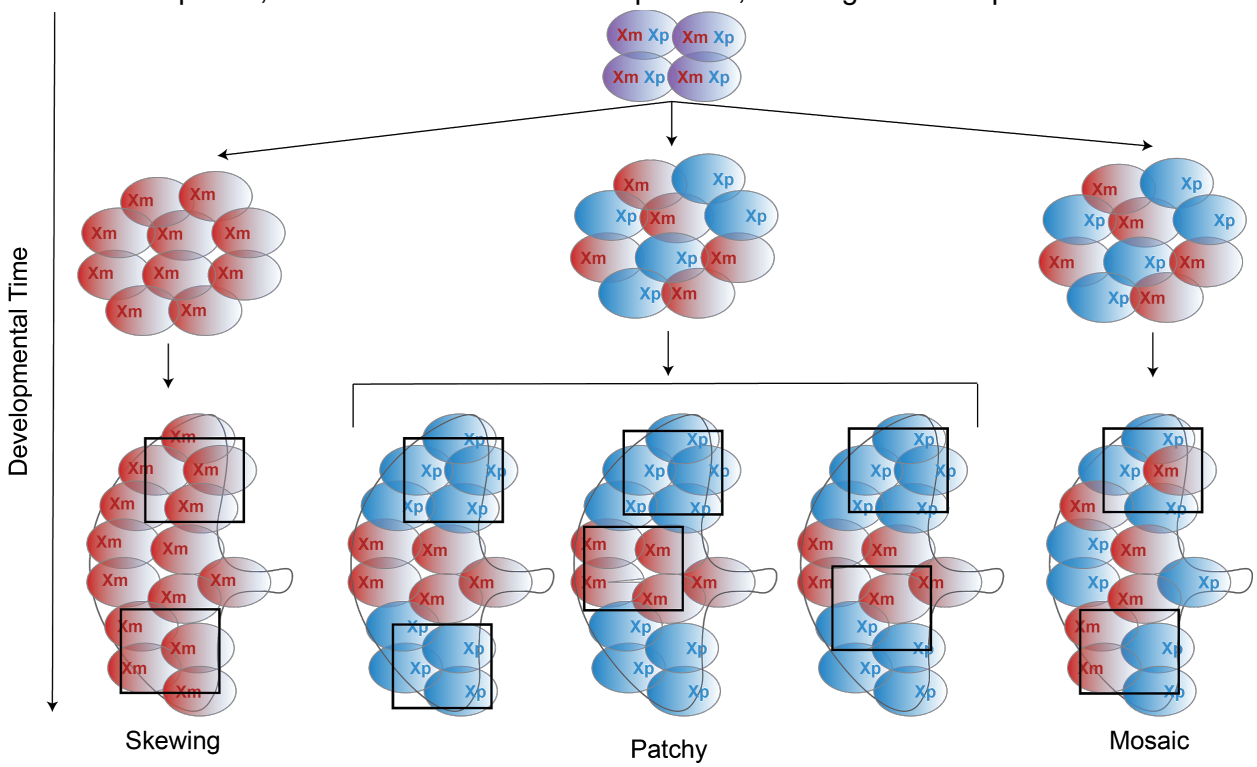
Previous research suggests that XCI across the entire X chromosome in the placenta is random and patchy, but these studies relied on a limited number of loci, and did not investigate chromosome-wide gene-specific variability (Moreira de Mello et al. 2010; Peñaherrera et al. 2012). Further, one of the challenges of previous research on XCI is that indirect methods have been used to infer inactivation. Gong et al. (2018) and Tukiainene et al. (2017) used higher expression in females as compared to males as a proxy for escape genes, rather than directly measuring inactivation status. However, Slavney et al. (2016) observed that escape genes show higher expression compared to inactivated genes in both males and females (Slavney et al. 2016).

Here we utilized next-generation sequencing to characterize genome-wide patterns of XCI in the human placenta, to compare patterns of XCI between the placenta and other tissues in humans, and to analyze genes that escape XCI that are unique to the placenta. We examined allele specific expression by sequencing whole-exome and whole-transcriptome (from two locations) of 30 full-term placentas from uncomplicated pregnancies. We further utilized data from the Genotype-Tissue Expression (GTEx) projects to examine XCI patterns in adult tissues (GTEx Consortium 2020). Using our experimental approach, we have the capacity to observe: 1) extreme skewing in the human placenta, where all placentas show the same X inactivated in both sites; 2) patchiness, where the placenta exhibits patches of daughter cells either maternal or paternal X expression only; or, 3) mosaicism, where clusters of cells in a sample exhibit both maternal and paternal X chromosome expression (**Figure 1**). Here we limit analysis of gene-specific inactivation to placental and adult tissues with skewed XCI to directly measure allele-specific expression. We observed evidence for skewing (large patches of exclusively maternal or paternal X expression) in the human placenta that is not present in adult tissues. We also compared genes that are inactivated, genes that escape from inactivation, and genes that show variable escape between the placenta and adult tissues. While a majority of genes are concordant for inactivation status between the placenta and adult tissues (71%), we observed a subset of genes that show opposite patterns between the placenta and adult tissues.

Our results provide additional evidence that XCI in the human placenta is random and patchy. Our study further shows that patterns of X-inactivation differ between embryonic tissue and adult tissues, consistent with different developmental structures of these tissues. Further, we identified genes with unique XCI status in the placenta. We hypothesize that these genes could play an important role in the placenta and pregnancy.

Figure 1. Schematic possible patterns of X chromosome inactivation.

Prior to XCI, both the maternal and the paternal X chromosomes are present (purple cells). The blue cell represents the maternal X is inactivated, and the red cell represents the paternal X is inactivated. If XCI occurs very early in development and propagates to all daughter cells, we would observe extreme skewing where all placentas show the same X inactivated. In this case, the same X chromosome is inactivated in both sites sampled from this study. If XCI occurs at an intermediate stage, XCI is random but is present with large patches of daughter cells with only the maternal X or paternal X. In this case, three possible scenarios can happen: 1. Both of the sites sampled in this study happen to come from two patches with the same X inactivated, 2. Each site sampled in this study comes from two patches with different X inactivation, and 3. One of the sites sampled in this study comes from the boundary of two patches. If XCI occurs very late in development, there are either small or no patches, creating a mosaic pattern of XCI.



Results

Whole chromosome level of X chromosome inactivation differs between the placenta and adult tissues

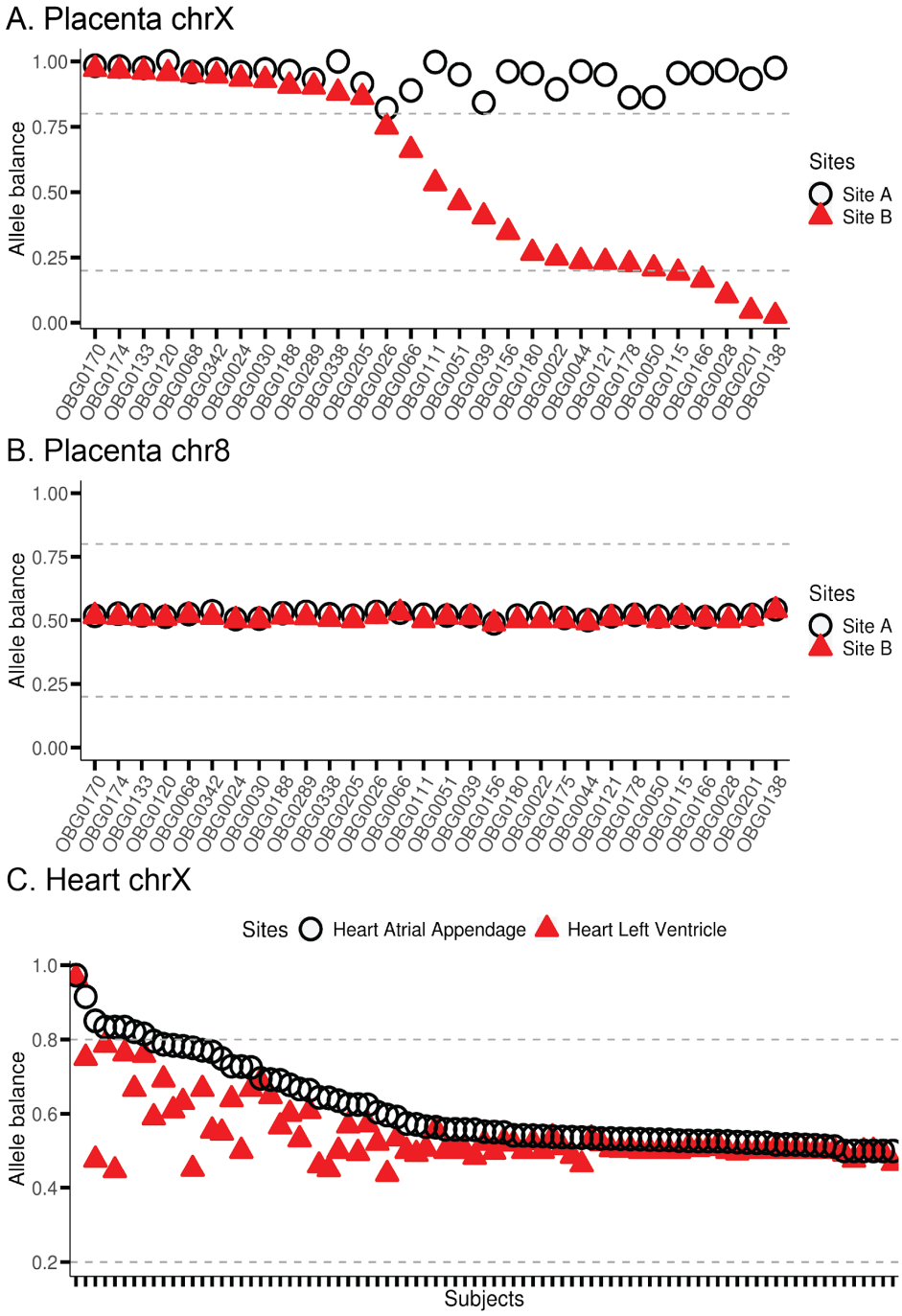
We observed evidence for patchiness in XCI in the placenta. To examine patterns of XCI in the human placenta, we sequenced the whole transcriptome from two separate extraction sites of the same placenta. We then employed a phasing strategy on the X chromosome using whole exome and whole transcriptome sequence data. Briefly, at heterozygous sites that exhibit allelic imbalance, we reasoned that the biased alleles (the alleles that are expressed in a higher proportion) are all on the same X chromosome (see **Methods**). In 17/29 placentas, both sites in the placenta exhibited extremely skewed X-inactivation, either with both extraction sites showing the same inactivated X chromosome (12 placentas; **Figure 2A**; **Figure S7A**) or each site showing the opposite X chromosome inactivated (5 placentas; **Figure 2A**; **Figure S7B**). In the remaining twelve placentas we observed one extraction site showing skewed XCI, and the other showing variable proportions of both X chromosomes being expressed (**Figure 2A**; **Figure S7C**). In this subset, we postulated that one of our samples was collected on the boundary of two different patches of X inactivated cells. To validate that the patterns we observed on the X chromosome are indeed XCI, we repeated the same analyses on chromosome 8, where all samples showed biallelic expression (**Figure 2B**). Our results suggest that the human placenta is organized into large patches with either the maternal or paternal X chromosome being inactivated.

We then found that patterns fXCI in the placenta are different from adult tissues (**Figure 2, Figure 3**). We first compared the pattern of whole chromosome XCI in the two placenta samples we collected with whole chromosome XCI in two regions collected from adult human hearts (left ventricle and atrial appendages) sampled by the GTEx consortium (GTEx Consortium 2020) (**Figure 2A, 2C**). We identified 85 individuals with RNAseq data for both of these locations in the heart and employed the same phasing approach as in the placenta. In the heart tissue, both regions exhibit biallelic expression (both maternal and paternal X expression) in 91% of individuals (**Figure 2C**). This is in stark contrast to the placenta, where none of individuals exhibit biallelic expression at both regions (**Figure 2A**) .

We further found the same biallelic pattern of XCI across all 43 adult tissue samples, in contrast to the placenta (**Figure 3**). Because other adult tissues include a single sample (GTEx Consortium 2020), to assess chromosome wide XCI we compared median allele balance computed from the biased allele for each sample. While most placenta samples (90%) exhibit skewed allele balance (*i.e.*, allele balance greater than 0.8), most samples from adult tissues (89%) exhibit biallelic expression (*i.e.*, allele balance between 0.5 and 0.8) (**Figure 3**). Similar to the placenta, however, also we observed biallelic expression across all adult tissues on chromosome 8, as expected (**Figure S8**). Together, these results suggest that patterns of whole chromosome XCI differ between the placenta and adult tissues.

167 Figure 2. X chromosome inactivation is patchy in the human placenta and
168 mosaic in the human heart

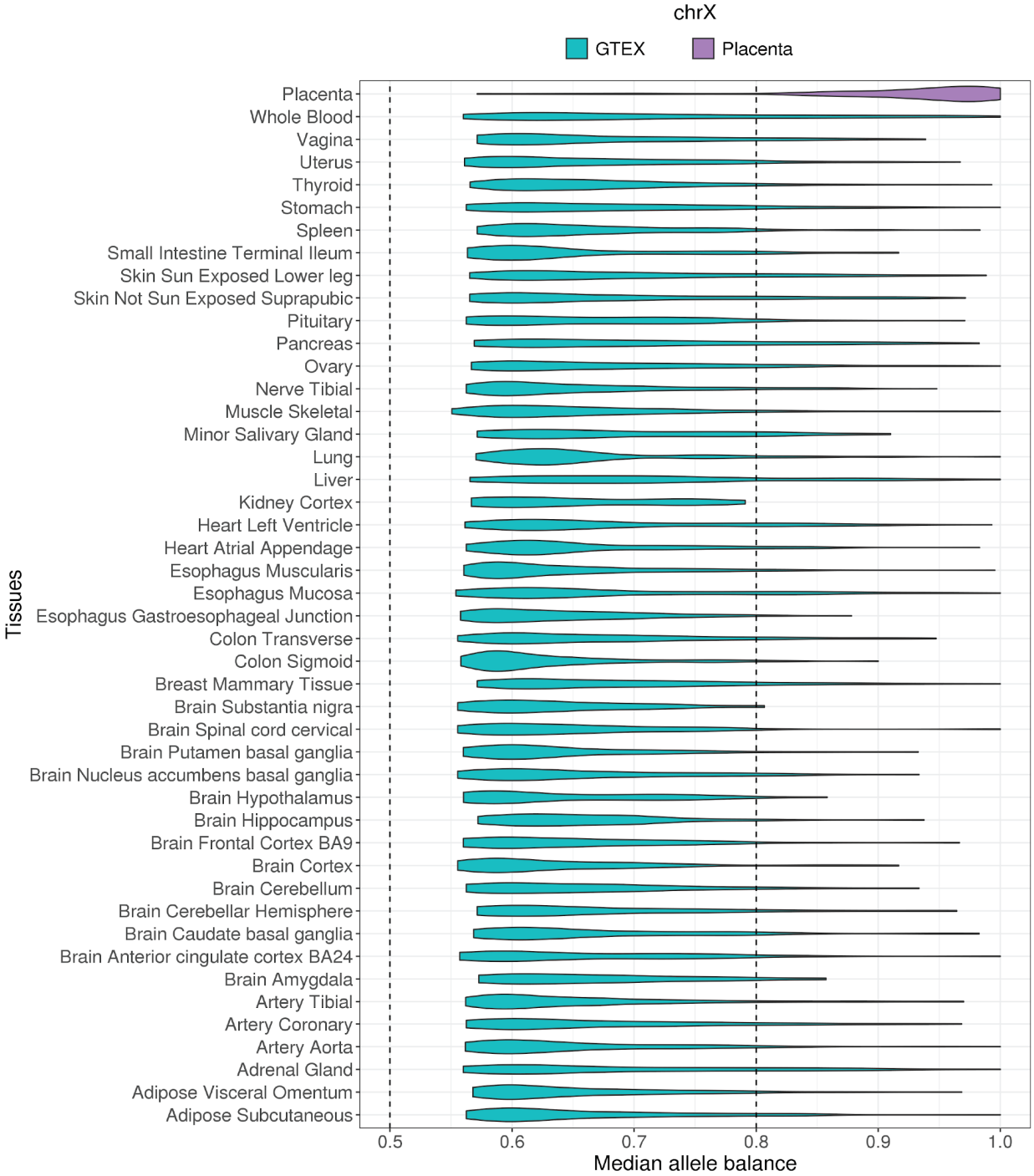
169 Phased allele balance aggregated across variants for each sample is shown for site A (open
170 black circle) and site B (filled red triangle). Each point is the median value across variants.
171 Median was used to minimize the effect of outliers. The dotted gray horizontal lines denote
172 allele balance of 0.2 and 0.8. (A) Placenta X chromosome. We removed variants that fall within
173 the pseudoautosomal regions (PARs) prior to computing medians. (B) Placenta chromosome 8.
174 (C) Heart X chromosome.



175

176 Figure 3. Whole chromosome level of X chromosome inactivation differs
177 between the placenta and adult tissues in humans.

178 Unphased median allele balance is plotted for the placenta in this study (purple) and for 45 adult
179 tissues in the GTEx dataset (blue). Each point of the violin plot is the median allele balance for
180 each sample. Because there are no multiple site samplings for the GTEx data, unphased allele
181 balance was computed. Here, the pseudoautosomal regions of the X chromosome are
182 removed.



183

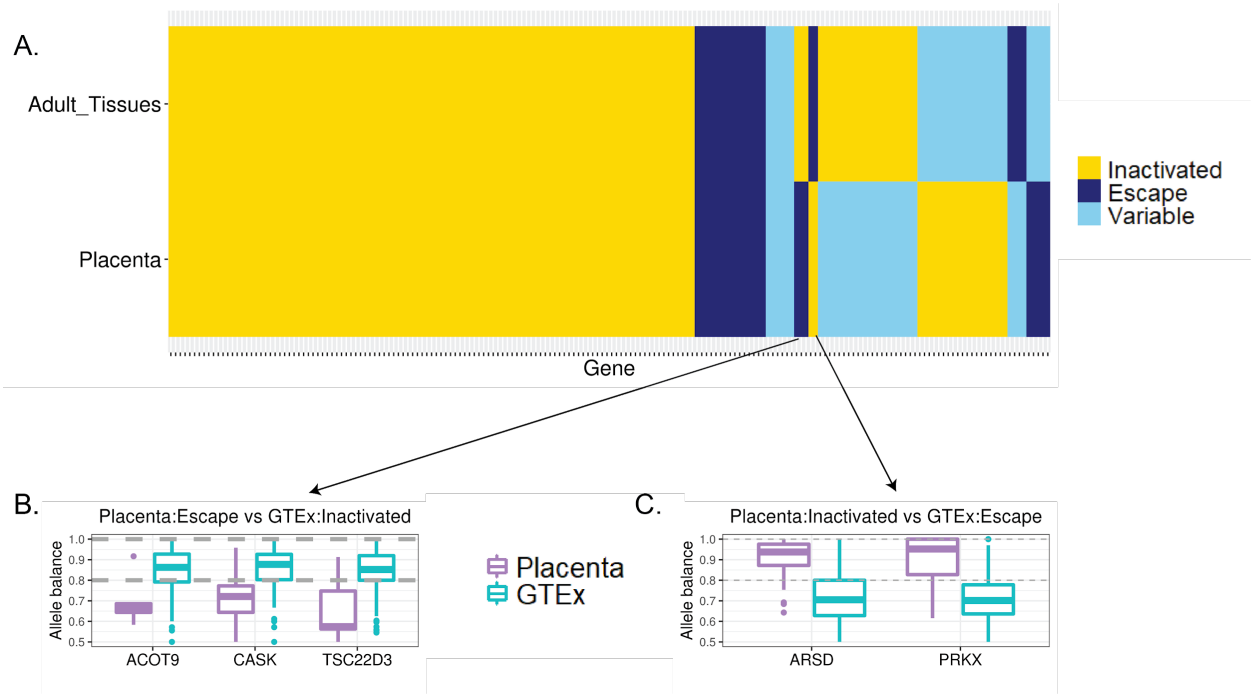
Gene-specific escape comparison between the placenta and adult tissues

We observed a large degree of variability in gene-specific escape in the placenta, both across and within individuals. We categorized genes on the X chromosome using only samples that exhibit skewed median allele balance across the X chromosome (i.e. median allele balance greater than 0.8, see **Methods**). Only 7 genes (2%) show evidence for escape across all samples with the ability to assay, while 73 genes (18%) show evidence of silencing across all samples (**Figure S9**). Rather, most genes, 316 genes (80%), show a gradient of the proportion of samples that exhibit evidence for escape. In addition, heterogeneity exists between two sites within individuals (**Figure S10)**. Given the tremendous heterogeneity in evidence for escape, we additionally investigated the ratio of female-to-male gene expression across X-linked genes. We observed that genes in all three classes (escape, variable, and silenced) exhibit a range of female-to-male gene expression ratios, with higher gene expression in females as compared to males underpowered to determine whether a gene has escaped X-inactivation (**Figure S11**).

We found similarities and differences in gene-specific escape and silencing in the human placenta when compared to the adult tissues. We used 525 skewed adult tissue samples in the GTEx dataset (out of 4958 total samples) and 52 skewed samples in the placenta dataset (out of 58 total samples) to categorize XCI inactivation status (see **Methods**). A gene is considered for this analysis if there are at least five informative samples (i.e., an informative sample is defined as having at least one heterozygous and expressed variant for that gene) in both the placenta dataset and the GTEx dataset. We found 186 such genes. Out of these 186 genes, 132 genes (71%) show the same pattern of escape or silencing between the adult tissues and the placenta dataset (**Figure 4A**). Specifically, 111 genes are inactivated, 15 genes escape, and 6 genes show variable escape between the adult tissues and the placenta. We also observed differences between the adult tissues and the placenta. We found three genes that escape in the placenta but are inactivated in adult tissues: *ACOT9*, *CASK*, *TSC22D3* (**Figure 4B**). We found two genes that are inactivated in the placenta but escape in adult tissues: *ARSD* and *PRKX* (**Figure 4C**). We examined the mapping quality and total RNA read counts of the variants in these genes to confirm that these observations are not artifacts (**Figure S12**). These results suggest that while the majority of genes on the X chromosome show consistent XCI between the placenta and adult tissues, a subset exhibits opposite patterns that could be attributed to the unique nature of the placenta tissue, which is an embryonic tissue.

216 Figure 4. Gene-specific escape comparison between the placenta and
217 adult tissues

218 (A) A heatmap showing genes that are inactivated in both the placenta and adult GTEx tissues
219 (yellow), genes that escape X chromosome inactivation in both the placenta and adult GTEx
220 tissues (dark blue), and genes that exhibit variable escape across individuals in both the
221 placenta and adult GTEx tissues (light blue). Boxplots showing median allele balance across all
222 samples in the placenta (purple) and adult tissues in GTEx (blue) for the three genes that show
223 evidence for escaping XCI in the placenta but being inactivated in the adult GTEx tissues (B)
224 and for the two genes that show evidence for being inactivated in the placenta but escaping in
225 the adult GTEx tissues (C). Each point of the box plot is the median allele balance computed
226 across variants for that gene (see **Methods**).
227



228

Discussion

We observed that X chromosome inactivation in the human placenta is organized into large patches of maternal or paternal X expression. We utilized whole exome and whole transcriptome sequence data to analyze allele specific expression across female placentas and found evidence that the human placenta exhibits large patches of maternal or paternal X chromosome expression. While patchy XCI in the human placenta has been observed in humans previously, these prior studies rely on a few SNPs and a few genes in a limited number of samples (Moreira de Mello et al. 2010; Looijenga et al. 1999). For example, Looijenga et al. (1999) examined two samples from each of the nine female (46, XX) full term placenta samples, using methylation in a single gene, androgen receptor, as a read-out, they found that 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). 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. These studies typically relied on a single gene as in Looijenga et al. (1999) or did not include multiple samplings from the same placenta as in Moreira de Mello et al. (2010). With improved sampling (i.e. sampling two independent locations from each placenta, from opposing quadrants) and genome-wide sequencing, we confirm that patterns of XCI are predominantly patchy in the human placenta.

We found distinct patterns of XCI between the placenta and adult tissues in humans: while most placenta samples (~90%) exhibit skewed expression, only about 11% of samples from human adult tissues exhibit skewed expression. This is consistent with previous methylation-based analyses that observed skewed inactivation across different developmental layers of the placenta (Peñaherrera et al. 2012), suggesting that each villus tree is clonally derived from a few precursor cells, and that there is little cell migration across placental regions. This is in contrast to adult tissues, where we observe very little skewing, suggesting more cell migration during development. Though both observations are contingent on the size of the tissue sample analyzed in bulk, where samples with more cells are more likely to capture both X-chromosomes being expressed.

Although we found high concordance in genes that are inactivated and genes that escape between the placenta samples and adult tissues (**Figure 4**), we found a subset of genes with unique XCI status in the placenta. Specifically, we found five genes that show patterns that are unique to the placenta: *ACOT9*, *CASK*, and *TSC22D3* are inactivated in adult tissues, but escape XCI in the placenta (**Figure 4B**). *ARSD* and *PRKX*, both genes in the pseudoautosomal region, escape XCI in adult tissues but are inactivated consistently across regions in the placenta (**Figure 4C**). *ACOT9* has been associated with syndromic X-linked intellectual disability Turner Type (Stelzer et al. 2016) but the function of *ACOT9* in the placenta is unclear. *CASK* was shown to be subject to XCI in mice and has been implicated to play an important role in neurological disorders (Mori et al. 2019). Since the placenta and cord blood could also be

linked to neurological disorders (Mordaunt et al. 2020), it is plausible that *CASK* escapes XCI, or experiences XCI erosion, in the placenta while it is inactivated in other tissues. *TSC22D3* is an immunity related gene. Syrett al. (2019) observed that *TSC22D3* is overexpressed in females compared to males in T cells of patients with systemic lupus erythematosus (Syrett et al. 2019). Curiously, ARSD is active in adult tissues, but silenced in the placenta. Larson et al. (2017) observed that *ARSD* is inactivated in at least one ovarian tumor sample, highlighting a potential parallel between the placenta and tumor development (Larson et al. 2017). *PRKX* is also shown to be an escape gene in adult tissues here, and as previously reported (Wainer Katsir and Linial 2019), but not in the placenta. We hypothesize that these genes could be further studied for a potential role in pregnancy and pregnancy complications.

We additionally show that sex differences in gene expression are not sufficient on their own to determine whether a gene escapes inactivation or not (**Figure S12**). This highlights the importance of understanding both inactivation status, and expression differences between the sexes. Gong et al. (2018) used female-biased gene expression as a proxy for identifying potential genes that escape XCI in the placenta. 28/47 potential escape genes identified by Gong et al. (2018) were reported previously as being inactivated or unknown (Gong et al. 2018). Of these, 3/28 genes exhibit variable escape in our placenta data (*MBTPS2*, *SMS*, and *PIN4*) and only 1/28 genes exhibit evidence for escaping (*CXorf36*). This means the other 24/28 show female-biased expression but no evidence of escape in the placenta.

In conclusion, we confirmed that XCI in the human placenta is skewed, but patchy, in contrast to adult tissues, which show mosaic XCI. In addition, we identified a subset of genes showing patterns of XCI that are unique to the placenta that should be further investigated for their roles in pregnancy and placentation.

Materials and Methods

Samples collection

Working with the Yale Biobank, we collected samples from 30 placentas, where the sex of the offspring was assigned female at birth. We confirmed that the samples were genetically XX by the presence of heterozygous sites across the X chromosome. The placenta samples were collected immediately following birth from term (≥ 36 weeks and 6 days) uncomplicated pregnancies delivered by cesarean section. Placentas were collected and sequenced at two different times, with 12 placenta samples in the first batch (OBG0044, OBG0068, OBG0111, OBG0115, OBG0120, OBG0133, OBG0156, OBG0170, OBG0174, OBG0175, OBG0178, and OBG0166) and 18 in the second (OBG0022, OBG0024, OBG0026, OBG0028, OBG0030, OBG0039, OBG0050, OBG0051, OBG0066, OBG0121, OBG0138, OBG0180, OBG0188, OBG0201, OBG0205, OBG0289, OBG0338, and OBG0342).

Placental tissue samples were obtained through rapid sampling (≤30 minutes post-delivery). Tissue samples were taken from the “maternal” side, midway between the chorionic and basal

312 plates, from the periphery of the lobules, avoiding maternal tissue, consistent with best practices
313 (Konwar et al. 2019) . Sampling sites were free of visible infarction, calcification, hematoma, and
314 tears (areas of frank visible pathology were avoided). Whenever possible, tissue samples were
315 obtained from distinct cotyledons of the placenta in opposing quadrants of the placenta (far from
316 one another spatially).
317
318 The sampling protocol was as follows:
319 1. Orient the placenta maternal side up (basal plate uppermost) and identify sampling
320 areas in each of the four placental quadrants.
321 2. At each site, remove the basal plate (to remove maternal tissue) ~1-2mm by trimming
322 with a pair of sterile scissors to expose villous tissue.
323 3. Cut a “grape-sized” (approximately 1-2 cm3; 5-6g) tissue lobule from each of the four
324 quadrants.
325 4. Wash the tissue thoroughly twice in phosphate buffered saline (PBS) solution and blot
326 on clean gauze.
327 5. From each “grape-sized” lobule, cut away eight smaller pieces (~1-2 mm3) using a
328 scalpel.
329 6. For each sampling quadrant, place four of the ~50mg tissue pieces in a labeled cryovial.
330 Snap freeze immediately in liquid nitrogen. Store samples at -80°C until ready for use;
331 aliquot individual tissue pieces as needed per research protocols.
332 7. For each sampling quadrant, place four of the tissue pieces in a labeled cryovial
333 containing 1 mL of RNAlater® solution (RNA stabilizing agent). Store the cryovials in the
334 4°C benchtop fridge for a minimum of 48 hours, and a maximum of 7 days, per
335 RNAlater® manufacturer protocol. After a minimum of 48 hours, use a pipette to remove
336 the RNAlater solution from the cryovials and immediately snap freeze in liquid nitrogen.
337 Store samples at –80°C until ready for use; aliquot individual tissue pieces as needed
338 per research protocols.

339 Whole exome sequencing

340 DNA was extracted from one flash frozen collection site for each individual. Exome libraries
341 were prepped, and sequenced to approximately 50X coverage with 100bp paired-end sequence
342 on the Illumina NextSeq at the Yale Genome Sequence Center.

343 Whole transcriptome sequencing

344 From each placenta, two separate sites from opposite quadrants were collected. RNA was
345 extracted and RiboZero stranded libraries (RF) were prepared and sequenced to approximately
346 40 million reads per sample with 100bp paired-end sequence on the Illumina NovaSeq at the
347 Yale Genome Sequence Center.

348 Exome sequence data processing

349 We used fastqc version 0.11.8 (Andrews 2010) for quality control and aggregated results from
350 fastqc by using multiqc version 0.9 (Ewels et al. 2016). We trimmed adapters using bbduk as

351 part of bbmap version 38.22 (Bushnell 2014). Sequences were trimmed both left and right for
352 phred quality of 30, minimum length of 75 base pairs, and average read quality of 20 or greater
353 to keep only high quality reads (Bushnell 2014). Quality was checked after trimming (**Figure**
354 **S1)**. We confirmed genetic XX females by examining the reads mapped ratio between the X
355 chromosome and chromosome 19 (chrX/chr19), between the Y chromosome and chromosome
356 19 (chrY/chr19), and between the Y chromosome and the X chromosome (chrX/chrY). We
357 observed that sample OBG0175 has a lower chrX/chr19 reads mapped ratio than other samples
358 in this study (**Figure S2**). Therefore, we removed sample OBG0175 from further analyses. We
359 used bwa-mem version 0.7.17 (Li 2013) to map to a female-specific reference genome.
360 Specifically, we mapped the exome samples to a sex chromosome complement informed
361 reference genome in which the Y chromosome is hard-masked (to avoid mismapping of X-
362 linked reads to homologous regions on the Y chromosome in the XX samples). Mapping exome
363 samples to a reference genome with the Y chromosome hard-masked has been shown to
364 increase the number of variants genotyped on the X chromosome (Webster et al. 2019). To
365 generate the sex chromosome complement reference genome we employed XYalign (Webster
366 et al. 2019). XYalign created a Y-masked gencode GRCh38.p12 human reference genome for
367 aligning XX individuals (Harrow et al. 2012). We used picard version 2.18.27 ("Picard Tools - By
368 Broad Institute" n.d.) to mark PCR duplicates. To genotype variants, we used GATK version
369 4.1.0.0 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). We first used
370 GATK's HaplotypeCaller to generate GVCF files. Second, we combined GVCF from 66 samples
371 using GATK's CombineGVCFs (30 placenta samples and an additional 36 samples from a
372 separate study in the group (unpublished) here to increase power to genotype variants). Finally,
373 we used GATK's GenotypeGVCFs to call variants. Following GATK's best practice, we obtained
374 high-quality variants by filtering using GATK's Variant Quality Score Recalibration (VQSR). We
375 tabulated the number of heterozygous variants for each sample in **Table S1**.
376
377 Because we sequenced the placenta samples in two separate batches, we plotted the principal
378 component for the exome data using the package *SNPRelate* in R (Zheng et al. 2012) and
379 observed no separation by batch from the exome data (**Figure S3**).

380 Whole transcriptome data processing

381 Samples were checked for quality using fastqc version 0.11.8 (Andrews 2010) and results were
382 aggregated across samples using multiqc version 0.9 (Ewels et al. 2016). Adapters were
383 removed and sequences were trimmed both left and right for phred quality of 30, minimum
384 length of 75 base pairs, and average read quality of 20 or greater using bbduck version 38.22
385 (Bushnell 2014). Quality was checked after trimming (**Figure S4)**. Trimmed RNA-seq female
386 placenta reads were then aligned to a sex chromosome complement informed reference
387 genome with the Y chromosome hard masked with Ns (Webster et al. 2019; Olney et al. 2019)
388 using the gencode GRCh38.p12 reference genome (Harrow et al. 2012). RNA-seq samples
389 were aligned using HISAT2 (Kim, Langmead, and Salzberg, n.d.); which has been shown to be
390 a robust aligner for high throughput RNA-seq reads. Total reads mapped and duplicate reads
391 were visually checked using BAMtools stats (Barnett et al. 2011) (**Table S2)**.

Obtaining allele counts for allele specific expression analysis

We used GATK ASEReadCounter (version 4.1.0.0) to obtain allele counts (McKenna et al. 2010) with the following thresholds: min-depth-of-non-filtered-base = 1, min-mapping-quality = 10, and min-base-quality = 10. GATK ASEReadCounter automatically filters out duplicated reads. Duplicated reads are suggested to be removed for allele specific expression analysis (Castel et al. 2015). The number of heterozygous and expressed variants for each sample are reported in **Table S3**.

Genotype-Tissue Expression (GTEx) data processing

We downloaded ASEReadCounter counts for the X chromosome and chromosome 8 from the Genotype-Tissue Expression (GTEx) project, approved for project #8834 for General Research Use in Genotype-Tissue Expression (GTEx) to MAW. We only considered tissues with more than ten samples per tissue and excluded two non-primary tissues: EBV-transformed lymphocytes and cultured fibroblasts, leaving 45 tissues to be analyzed in this study. We list the GTEx tissues, the number of samples per tissue examined in this study, and the number of skewed samples per tissue in **Table S4**.

Computing unphased median allele balance

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 the sum of the read count of the reference allele and the read count of the alternate allele. We define the biased allele to be the allele with more read counts. For each heterozygous and expressed variant, the unphased allele balance is the ratio between the read count of the biased allele and the total read count. Then, for each sample, the median allele balance is calculated by computing the median unphased allele balance across all heterozygous and expressed variants.

Defining threshold for allele specific expression

To determine the threshold for defining biased allele expression, we plotted the unphased allele balance across all variants for each individual for chromosome 8 and the X chromosome (**Figure S5**). We observed that the unphased allele balance of most variants on the X chromosome is greater than 0.8 while the unphased allele balance of most variants on chromosome 8 is less than 0.8 (**Figure S5**). Therefore, in this study, we used 0.8 as a threshold for biased allele specific expression.

Determining which X chromosome is inactivated

To determine whether the same X chromosome is inactivated at the two extraction sites for each placenta sample, we employed a phasing strategy on the X chromosome by defining that the biased alleles (the alleles with higher counts) are on the same haplotype. We restricted this analysis to contain only heterozygous and expressed variants that are shared between the two extraction sites. For consistency, we defined extraction site A to be the site with more expressed

429 variants where the unphased allele balance is greater than 0.8. We defined the activated X
430 chromosome to consist of alleles where its allele balance (the ratio between this allele's read
431 count and the total read count) is greater than 0.8. In cases where the allele balance is less than
432 0.8, we picked an allele at random with equal probability between the reference allele and the
433 alternate allele. Then, we computed the allele balance (defined as the ratio between the biased
434 allele's count and total count) for the alleles on the same X chromosome. We called this phased
435 allele balance. At each site, we computed the median phased allele balance to use as summary
436 statistics. To compute phased allele balance we followed this procedure:
437 1. Find expressed variants that are shared between site A and site B.
438 2. Using the shared expressed variants between site A and site B, tabulate the number of
439 expressed variants that exhibit biased expression (i.e., unphased allele balance is greater than
440 0.8).
441 3. Pick a site to base phasing from based on the number of expressed variants with biased
442 expression. For example, if site A has 50 expressed variants with biased expression and site B
443 has 60 such variants, phasing is based on site B.
444 4. Phasing strategy (using site B to base phasing): the expressed haplotype is generated by
445 calculate: For each expressed variant that is shared between site A and site B, pick the allele
446 with allele balance greater than 0.8 to be on the expressed haplotype. If allele balance is less
447 than 0.8, choose an allele at random with equal probability

448 Validating method based on allele specific expression to quantify XCI

449 We validated our approach to determine skewness by examining the non-pseudoautosomal
450 regions (nonPARs) of the X chromosome in males. We collected samples from 12 placentas,
451 where the sex of the offspring was assigned male at birth. Even though the X chromosome in
452 males is haploid, some variants are incorrectly genotyped as being heterozygous when diploidy
453 is assumed when genotyping with GATK (**Table S5**). Regardless, these heterozygous variants
454 exhibit mostly skewed expression, indicating that our method of combining both the whole
455 exome and whole transcriptome sequence data is robust to determine skewness (**Figure S6**).
456

457 Classifying genes into genes that are inactivated, genes that escape XCI,
458 and genes that show variable escape

459 For classifying genes, we only considered samples that show skewed expression (allele
460 balance is greater than 0.8). There are 52/58 samples in the placenta dataset and 525/4958
461 samples in the GTEx dataset exhibiting skewed expression that were used for this analysis. For
462 each gene on the X chromosome, we only considered a gene if there are at least five
463 informative samples where in each sample, there is at least one heterozygous and expressed
464 variant for that gene. If there are more than one heterozygous and expressed variant for a gene,
465 we summed up the counts for the biased allele and calculated allele balance by dividing the
466 summed counts of the biased allele by the summed counts of the total counts. We categorized a
467 gene as being inactivated if its allele balance is greater or equal to 0.8 in at least 70% of the
468 samples and the median allele balance across all samples is greater or equal to 0.8. We
469 categorized a gene as escaping from XCI if at most 30% of the samples have allele balance for

this gene equal or greater than 0.8 and the median allele balance across all samples is less than or equal to 0.75. Otherwise, the gene is annotated as showing variable escape across all individuals. See **Supplementary Note 2** for more information.

Coordinates used of PARs and XIST

As defined for GRCh38.p12 ("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 73,820,651 to position 73,852,723 on the X chromosome

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 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. This research was supported by The Yale University Reproductive Sciences Data and Specimen Biorepository, HIC#1309012696, a component of the Department of Obstetrics, Gynecology & Reproductive Sciences, Yale School of Medicine, New Haven, CT. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number F31HD101252 to KCO. KCO was additionally supported by ARCS Spetzler Scholar.

Competing interests

The authors declare no competing interests.

Data and Code Availability

The datasets generated during this study are available via controlled access: phs002240.v1.p1 NIGMS Sex Differences Placentas. The code generated during this study is available at https://github.com/SexChrLab/Placenta_XCI.

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.

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.

Berletch, Joel B., Wenxiu Ma, Fan Yang, Jay Shendure, William S. Noble, Christine M. Disteche, and Xinxian Deng. 2015. “Escape from X Inactivation Varies in Mouse Tissues.” *PLoS Genetics* 11 (3): e1005079.

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.

Carrel, Laura, and Carolyn J. Brown. 2017. “When the Lyon(ized Chromosome) Roars: Ongoing Expression from an Inactive X Chromosome.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372 (1733). https://doi.org/10.1098/rstb.2016.0355.

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., E. Magda Price, Meaghan J. Jones, Bradley P. Balaton, Michael S. Kobor, and Carolyn J. Brown. 2015. “Landscape of DNA Methylation on the X Chromosome Reflects CpG Density, Functional Chromatin State and X-Chromosome Inactivation.” *Human Molecular Genetics* 24 (6): 1528–39.

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.

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.

GTEx Consortium. 2020. “The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues.” *Science* 369 (6509): 1318–30.

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.

549 “Human Genome Overview - Genome Reference Consortium.” n.d. Accessed September 5,
550 2019. https://www.ncbi.nlm.nih.gov/grc/human.
551 Huynh, K. D., and J. T. Lee. 2001. “Imprinted X Inactivation in Eutherians: A Model of Gametic
552 Execution and Zygotic Relaxation.” *Current Opinion in Cell Biology* 13 (6): 690–97.
553 Huynh, Khanh D., and Jeannie T. Lee. 2005. “X-Chromosome Inactivation: A Hypothesis
554 Linking Ontogeny and Phylogeny.” *Nature Reviews. Genetics* 6 (5): 410–18.
555 James, J. L., A. M. Carter, and L. W. Chamley. 2012. “Human Placentation from Nidation to 5
556 Weeks of Gestation. Part I: What Do We Know about Formative Placental Development
557 Following Implantation?” *Placenta* 33 (5): 327–34.
558 Kim, Daehwan, Ben Langmead, and Steven Salzberg. n.d. “HISAT: Hierarchical Indexing for
559 Spliced Alignment of Transcripts.” https://doi.org/10.1101/012591.
560 Konwar, Chaini, Giulia Del Gobbo, Victor Yuan, and Wendy P. Robinson. 2019. “Considerations
561 When Processing and Interpreting Genomics Data of the Placenta.” *Placenta* 84
562 (September): 57–62.
563 Larson, Nicholas B., Zachary C. Fogarty, Melissa C. Larson, Kimberly R. Kalli, Kate Lawrenson,
564 Simon Gayther, Brooke L. Fridley, Ellen L. Goode, and Stacey J. Winham. 2017. “An
565 Integrative Approach to Assess X-Chromosome Inactivation Using Allele-Specific
566 Expression with Applications to Epithelial Ovarian Cancer.” *Genetic Epidemiology* 41 (8):
567 898–914.
568 Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-
569 MEM.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.
570 Looijenga, L. H., A. J. Gillis, A. J. Verkerk, W. L. van Putten, and J. W. Oosterhuis. 1999.
571 “Heterogeneous X Inactivation in Trophoblastic Cells of Human Full-Term Female
572 Placentas.” *American Journal of Human Genetics* 64 (5): 1445–52.
573 Lyon, M. F. 1961. “Gene Action in the X-Chromosome of the Mouse (Mus Musculus L.).” *Nature*
574 190 (April): 372–73.
575 McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew
576 Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce
577 Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20
578 (9): 1297–1303.
579 Mordaunt, Charles E., Julia M. Jianu, Benjamin I. Laufer, Yihui Zhu, Hyeyeon Hwang, Keith W.
580 Dunaway, Kelly M. Bakulski, et al. 2020. “Cord Blood DNA Methylome in Newborns Later
581 Diagnosed with Autism Spectrum Disorder Reflects Early Dysregulation of
582 Neurodevelopmental and X-Linked Genes.” *Genome Medicine* 12 (1): 88.
583 Moreira de Mello, Joana Carvalho, Erica Sara Souza de Araújo, Raquel Stabellini, Ana Maria
584 Fraga, Jorge Estefano Santana de Souza, Denilce R. Sumita, Anamaria A. Camargo, and
585 Lygia V. Pereira. 2010. “Random X Inactivation and Extensive Mosaicism in Human
586 Placenta Revealed by Analysis of Allele-Specific Gene Expression along the X
587 Chromosome.” *PloS One* 5 (6): e10947.
588 Mori, Takuma, Enas A. Kasem, Emi Suzuki-Kouyama, Xueshan Cao, Xue Li, Taiga Kurihara,
589 Takeshi Uemura, Toru Yanagawa, and Katsuhiko Tabuchi. 2019. “Deficiency of
590 Calcium/calmodulin-Dependent Serine Protein Kinase Disrupts the Excitatory-Inhibitory
591 Balance of Synapses by down-Regulating GluN2B.” *Molecular Psychiatry* 24 (7): 1079–92.
592 Natri, Heini, Angela R. Garcia, Kenneth H. Buetow, Benjamin C. Trumble, and Melissa A.
593 Wilson. 2019. “The Pregnancy Pickle: Evolved Immune Compensation Due to Pregnancy
594 Underlies Sex Differences in Human Diseases.” *Trends in Genetics*.
595 https://doi.org/10.1016/j.tig.2019.04.008.
596 Olney, Kimberly C., Sarah M. Brotman, Valeria Valverde-Vesling, Jocelyn Andrews, and Melissa
597 A. Wilson. 2019. “‘Aligning RNA-Seq Reads to a Sex Chromosome Complement Informed
598 Reference Genome Increases Ability to Detect Sex Differences in Gene Expression.’”
599 *bioRxiv*. https://doi.org/10.1101/668376.

Peñaherrera, M. S., R. Jiang, L. Avila, R. K. C. Yuen, C. J. Brown, and W. P. Robinson. 2012. "Patterns of Placental Development Evaluated by X Chromosome Inactivation Profiling Provide a Basis to Evaluate the Origin of Epigenetic Variation." *Human Reproduction* 27 (6): 1745–53.

"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.

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.

Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, et al. 2016. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 54 (June): 1.30.1–1.30.33.

Sui, Yilun, Qianqian Chen, and Xiaoxi Sun. 2015. "Association of Skewed X Chromosome Inactivation and Idiopathic Recurrent Spontaneous Abortion: A Systematic Review and Meta-Analysis." *Reproductive Biomedicine Online* 31 (2): 140–48.

Syrett, Camille M., Bam Paneru, Donavon Sandoval-Heglund, Jianle Wang, Sarmistha Banerjee, Vishal Sindhava, Edward M. Behrens, Michael Atchison, and Montserrat C. Anguera. 2019. "Altered X-Chromosome Inactivation in T Cells May Promote Sex-Biased Autoimmune Diseases." *JCI Insight* 4 (7). https://doi.org/10.1172/jci.insight.126751.

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.

Turco, Margherita Y., and Ashley Moffett. 2019. "Development of the Human Placenta." *Development*. https://doi.org/10.1242/dev.163428.

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.

Wainer Katsir, Kerem, and Michal Linial. 2019. "Human Genes Escaping X-Inactivation Revealed by Single Cell Expression Data." *BMC Genomics* 20 (1): 201.

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,

651 Yunping Dai, Howard Levine, Lygia V. Pereira, and Xiangzhong Yang. 2002. “Aberrant
652 Patterns of X Chromosome Inactivation in Bovine Clones.” *Nature Genetics* 31 (2): 216–20.
653 Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S.
654 Weir. 2012. “A High-Performance Computing Toolset for Relatedness and Principal
655 Component Analysis of SNP Data.” *Bioinformatics* 28 (24): 3326–28.