



OPEN

Concurrent X chromosome inactivation and upregulation during non-human primate preimplantation development revealed by single-cell RNA-sequencing

Ana Luíza Cidral^{1,3}, Joana C. Moreira de Mello^{2,3}, Joost Gribnau² & Lygia V. Pereira¹✉

In mammals, dosage compensation of X-linked gene expression between males and females is achieved by inactivation of a single X chromosome in females, while upregulation of the single active X in males and females leads to X:autosome dosage balance. Studies in human embryos revealed that random X chromosome inactivation starts at the preimplantation stage and is not complete by day 12 of development. Alternatively, others proposed that dosage compensation in human preimplantation embryos is achieved by dampening expression from the two X chromosomes in females. Here, we characterize X-linked dosage compensation in another primate, the marmoset (*Callithrix jacchus*). Analyzing scRNA-seq data from preimplantation embryos, we detected upregulation of *XIST* at the morula stage, where female embryos presented a significantly higher expression of *XIST* than males. Moreover, we show an increase of X-linked monoallelically expressed genes in female embryos between the morula and late blastocyst stages, indicative of XCI. Nevertheless, dosage compensation was not achieved by the late blastocyst stage. Finally, we show that X:autosome dosage compensation is achieved at the 8-cell stage, and demonstrate that X chromosome dampening in females does not take place in the marmoset. Our work contributes to the elucidation of primate X-linked dosage compensation.

In mammals, dosage compensation between the single X chromosome in males and the two X in females is achieved by the process of X chromosome inactivation (XCI) which leads to the transcriptional silencing of one X in females¹. In eutherian mammals the process is initiated during preimplantation development with the expression of the long non-coding RNA *XIST* from the future inactive X (Xi)^{2,3}. Until recently, most studies of XCI initiation were performed in mice due to the ease in performing experiments in this animal model and the capability of mouse embryonic stem cells (mESCs) to undergo XCI when differentiated in vitro. Thus, in mice, XCI is initiated at the 4-cell stage with *XIST* being expressed exclusively from the paternal allele, leading to imprinted XCI (reviewed in⁴). At the blastocyst stage, cells from the epiblast (EPI) reactivate the inactive paternal X and, upon implantation, undergo a second round of XCI where each cell randomly inactivates either the paternal or the maternal X.

In contrast, studies of XCI initiation in humans have been hampered by difficulties in performing experiments in human embryos and lack of an adequate in vitro model system, since human embryonic stem cells (hESCs) are mostly in a post-XCI state⁵. Nevertheless, several differences between the process in mice and humans have been observed, including initiation of *XIST* expression at the 8-cell stage⁶, *XIST* expression from both X chromosomes in female or from the single X chromosome in male embryos^{7,8} and lack of imprinted XCI⁹.

¹National Laboratory for Embryonic Stem Cells (LaNCE), Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, São Paulo, SP 05508-090, Brazil. ²Department of Developmental Biology, Onco Institute, Erasmus MC University Medical Center, 3015GE Rotterdam, The Netherlands. ³These authors contributed equally: Ana Luíza Cidral and Joana C. Moreira de Mello. ✉email: lpereira@usp.br

Embryo	Number of cells	Number of cells after filtering	Cell type
Zygote_1	1	1	–
Zygote_2	1	1	–
Zygote_3	1	–	–
4cell_1	4	4	Blastomere
8cell_1	7	5	Blastomere
8cell_2	8	5	Blastomere
CompactedMorula_1 ^a	22	15	Morula
CompactedMorula_2	10	9	Morula
CompactedMorula_3 ^a	21	21	Morula
EarlyICM_1	9	7	Inner cell mass
EarlyICM_2 ^a	10	8	Inner cell mass
EarlyICM_3 ^a	23	8	Inner cell mass
LateICM_2	39	26	Primitive endoderm, epiblast
LateICM_4 ^a	40	14	Primitive endoderm, epiblast

Table 1. Summary of embryos and cells analyzed. ^aFemale embryos.

Intriguingly, using single-cell RNA sequencing (scRNA-seq) data from human preimplantation embryos, Petropoulos et al.⁸ identified X-linked dosage compensation in female blastocysts in the absence of XCI. The authors proposed a dampening process of X-linked gene expression from both active X chromosomes as an alternative mechanism of dosage compensation in human embryos⁸. Using a more stringent analysis combining the aforementioned and additional scRNA-seq data, Moreira de Mello et al. excluded the hypothesis of X-chromosome dampening and showed evidence that initiation of random XCI takes place between the morula and early blastocyst stage in humans¹⁰. More recently, Zhou et al. showed that XCI initiates around the time of implantation and is not complete by day 12 of human development¹¹. The conflicting results regarding *XIST* expression and timing of initiation of XCI pose a challenge in the efforts to uncover the mechanisms of dosage compensation in humans.

In this study, we aimed to investigate features of the XCI in a species that is evolutionarily closer to humans than mice. The marmoset (*Callithrix jacchus*) is a new world monkey widely used in biomedical research, whose evolutionary distance from humans is around 40 million years. We analyzed scRNA-seq data from 14 marmoset embryos, from zygote to the late blastocyst stage¹². Using allele-specific expression analysis we show that there is no evidence of X dampening during this primate's early development and that random XCI is also initiated at the preimplantation stage, concomitant with upregulation of the active X in male and female embryos.

Results

Embryo sexing. The dataset originally contained the transcriptome of 196 cells obtained from 14 marmoset preimplantation embryos, from the zygote to the late blastocyst stage. After filtering out cells with low-quality RNAseq data, we retained 124 cells representing all the above-mentioned stages (Table 1).

Embryos at the same stage were compared in terms of the number of Y-linked expressed genes and their respective expression levels. This strategy is able to sex embryos after complete embryonic genome activation, at the 8-cell stage^{12–14}. We identified two morulas, two early- and one late-blastocyst as female; and two 8-cell embryos, one morula, one early- and one late-blastocyst as male (Supplementary Fig. S1 and Table 1).

Allele-specific expression. In order to analyze allele-specific expression, we identified heterozygous SNP positions in each embryo. For a SNP position to be considered, it had to be covered by at least 20 reads; and if a gene contained variants indicating mono and biallelic expression in the same cell, that gene was discarded from the analysis. In total, 16,311 heterozygous positions and 6985 informative genes were detected, of which 181 were located on the X chromosome.

Twenty three X-linked genes biallelically expressed were detected in male embryos (Supplementary Fig. S2). This is expected in cells from stages prior to embryonic genome activation (EGA) as maternal mRNAs are still present^{12–14}. In later staged male embryos, X-linked biallelic expression may be detected due to misalignment of reads from genes located in the pseudoautosomal regions (PAR) or from genes with homologues in the Y and the X chromosomes¹⁵. In order to avoid interference from these genes, and since the marmoset PAR is not well described, we removed from the analysis all genes homologous to those in the human PAR and to those known to escape XCI (*escapees*), and all X-linked genes identified as biallelically expressed in at least one cell from a male embryo after the 8-cell stage (Supplementary Figs. S2 and S3). All subsequent analyses were performed using 140 informative X-linked genes resulting from this filtering step.

***XIST* expression.** Induction of *XIST* expression is a hallmark of XCI initiation. In the marmoset embryos, *XIST* expression was first detected at the 8-cell stage in males (Fig. 1). *XIST* expression in female compacted morulas was significantly higher than in males at the same stage (P-value = 6.7×10^{-6}). Surprisingly, at the early blastocyst stage, cells from the ICM from both male and female embryos expressed high levels of *XIST*.

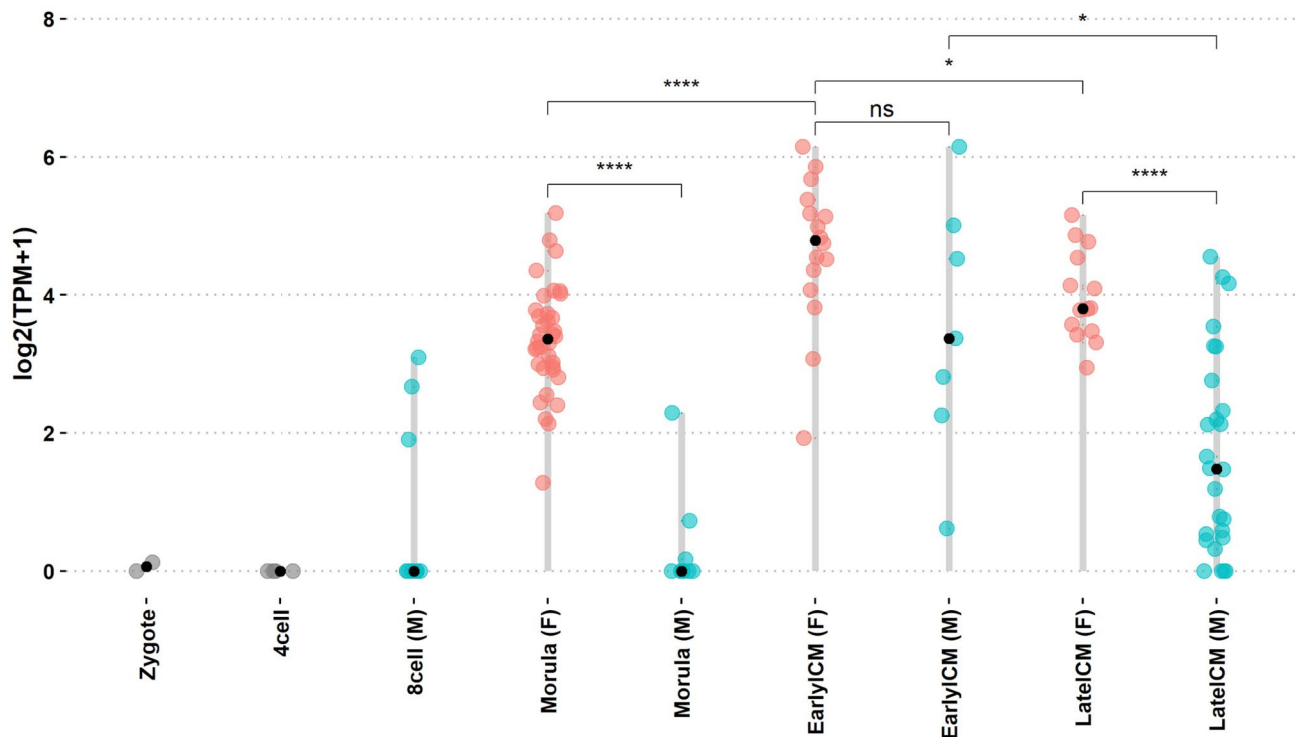


Figure 1. *XIST* expression in single cells during marmoset embryonic development. Each dot represents a single cell. Cells are grouped by embryonic stage and sex. Pink, female embryos (F); blue, male embryos (M). Distributions of *XIST* expression levels were compared using an unpaired Wilcoxon test. (*) P-value ≤ 0.05 ; (****) P-value ≤ 0.0001 ; (ns) not significant (P-value ≥ 0.05).

Although *XIST* expression persisted in the ICM of male late blastocyst, it was significantly downregulated (P-value = 4.1×10^{-5}).

Allele-specific X-linked gene expression. X chromosome inactivation leads to monoallelic expression of most X-linked genes in females. Thus, in order to further investigate the process during marmoset preimplantation embryo development, we calculated the proportion of monoallelically expressed genes in each embryonic stage (Fig. 2a). As expected, male embryos showed an increase of X-linked monoallelic expression between the 8-cell and late blastocyst stages, reflecting the degradation of maternal mRNA (Fig. 2a and Supplementary Fig. S4). In female embryos, we found a significant increase of monoallelic X-linked expression between the morula and blastocyst stages (Pearson's $r = 0.56$; P-value = 9.8×10^{-7} , Fig. 2a; and Supplementary Fig. S4), indicative of an ongoing XCI process. The same analysis performed in chromosome 10 did not reveal any significant increase of monoallelically expressed genes in males (Pearson's $r = 0.16$; P-value = 0.27) and a small correlation in females (Pearson's $r = 0.26$; P-value = 0.038) (Supplementary Fig. S5).

The location of all X-linked informative SNPs and their allelic-expression patterns in female embryos are shown in Fig. 2b and Supplementary Fig. S4. It is possible to see a cluster of genes close to the *XIST* locus with monoallelic expression in most of the cells at the early and late blastocyst stages.

Interestingly, we detected monoallelic expression of *XIST* from the alternative allele in one cell of the female early blastocyst. In another cell of the same embryo, the reference allele was monoallelically expressed, although with coverage of 14 reads, below our established threshold of 20 covering reads (Supplementary Fig. S6). Nevertheless, the detection of different *XIST* alleles being monoallelically expressed in different cells of the same embryo indicates that random XCI is taking place. In the same early blastocyst embryo where *XIST* showed monoallelic expression, we also detected three other genes located in a 1 Mb window from *XIST* being monoallelically expressed: *RNF12* and two uncharacterized loci (LOC100408393 and LOC100404148). In mice, *Rnf12* is an activator of XCI, participating in the upregulation of *XIST*^{18–20}. During the *Xist* accumulation process, *Rnf12* is expressed exclusively from the future active X, and is downregulated very quickly after XCI is initiated²¹. Therefore, detection of monoallelic expression of *RNF12* further suggests ongoing XCI in the marmoset early blastocyst.

X chromosome dampening. According to Petropoulos et al., dampening leads to dosage compensation in human preimplantation embryos by decreased expression of biallelically expressed X-linked genes⁸. In order to test the occurrence of X chromosome dampening in the marmoset, we calculated the median expression levels of biallelically expressed X-linked genes in female embryos during development (Fig. 3; Supplementary Fig. S7). If dampening of the two active X chromosomes occurred, we should detect a decrease in the median expression

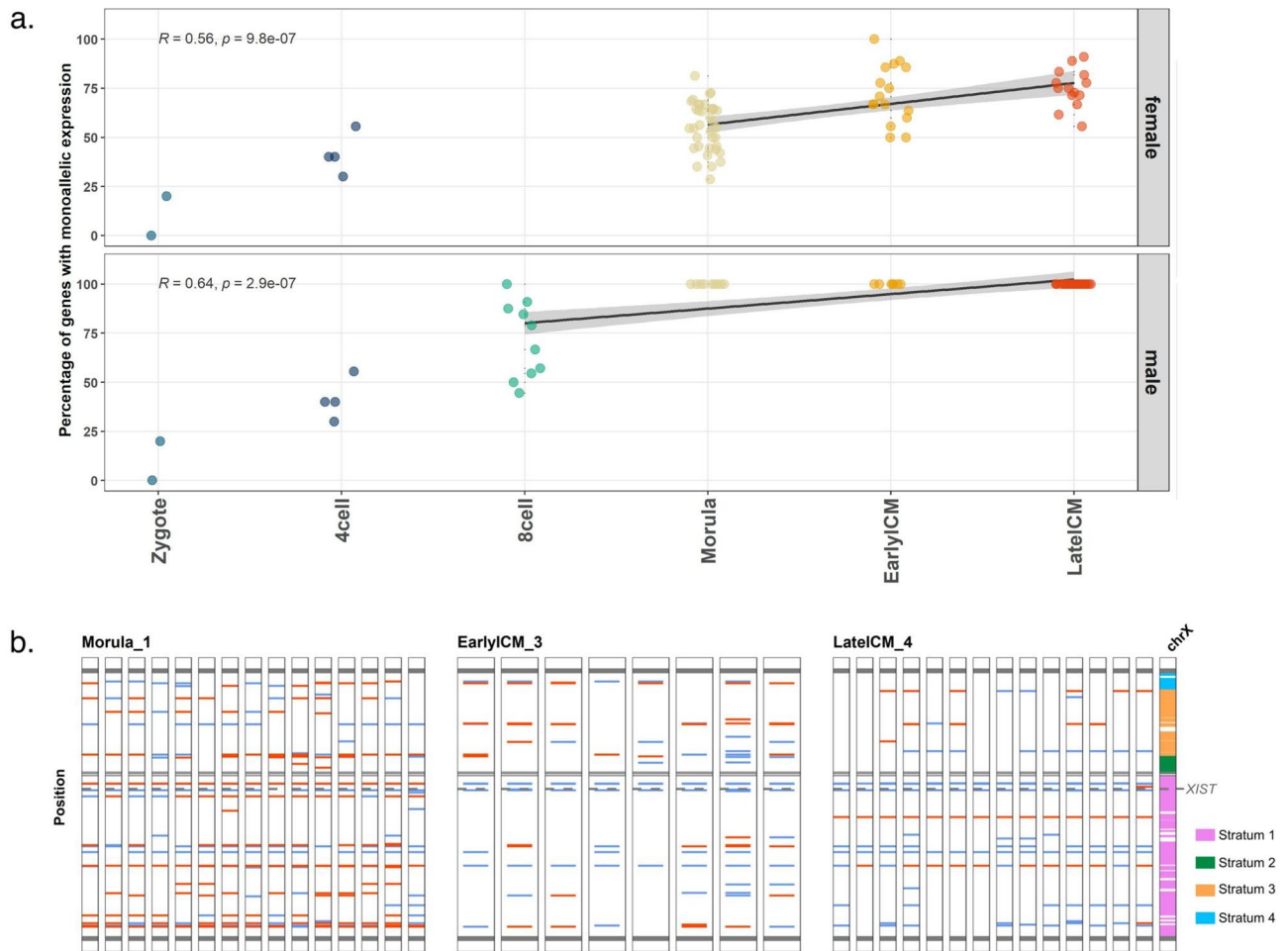


Figure 2. Allelic expression of X-linked genes. **(a)** Percentage of monoallelically expressed X-linked genes for each cell according to embryonic stage. Top panel: female embryos; bottom panel: male embryos. Each dot represents a single cell. Pearson's r values (R) are depicted in each panel. Gray area indicates a 95% confidence interval. **(b)** Expression patterns of genes along the X-chromosome of female embryos. The color of each position is assigned according to the ratio of allelic expression based on the detected SNP; blue corresponds to monoallelic expression, while red indicates biallelic expression. Each column corresponds to a single cell. The dark grey lines indicate the telomeres; the light grey lines the centromere; and the dashed line corresponds to the putative marmoset *XIST* locus. ChrX: scheme of the X chromosome showing the evolutionary strata S1, S2, S3 and S4 based on human data^{16,17}.

level of those genes during development. We found no significant differences in the expression levels of biallelically expressed X-linked genes among morula, early and late blastocyst stages (Fig. 3a), refuting the occurrence of X chromosome dampening.

Together, these data show that, while there is a growing number of monoallelically expressed genes indicative of ongoing XCI, genes that are still transcribed from both X chromosomes are not downregulated (dampened) between the morula, early and late blastocyst stages. Interestingly, we found a significant increase of the median expression levels of monoallelically expressed X-linked genes between the early and late blastocyst stages in females (Fig. 3b).

X chromosome upregulation. Upregulation of the single active X in mammalian cells was proposed as a mechanism of dosage compensation between that chromosome and the pairs of autosomes²². Indeed, X upregulation has been observed in humans, marsupials and mice^{10,23,24}. In order to further investigate the increase in X-linked gene expression during marmoset embryonic development, we looked at the expression levels of all X-linked genes in both male and female embryos. We identified an increase in total X-linked median expression levels between morula and early blastocyst stage in both males and females consistent with upregulation of the X chromosome (Fig. 3a,b; P -value = 3.6×10^{-4} and 0.031 for females and males, respectively).

In order to test if upregulation grants the X chromosome an expression level similar to the autosomes, we calculated the X:A ratio for each cell grouped by embryonic stage and sex (Fig. 4; Supplementary Fig. S8). Our results indicate that marmoset male embryos achieve dosage compensation between X chromosome and autosomes earlier than humans: at the 8-cell stage the X:A ratio is already close to 1 (Fig. 4b), whereas in humans

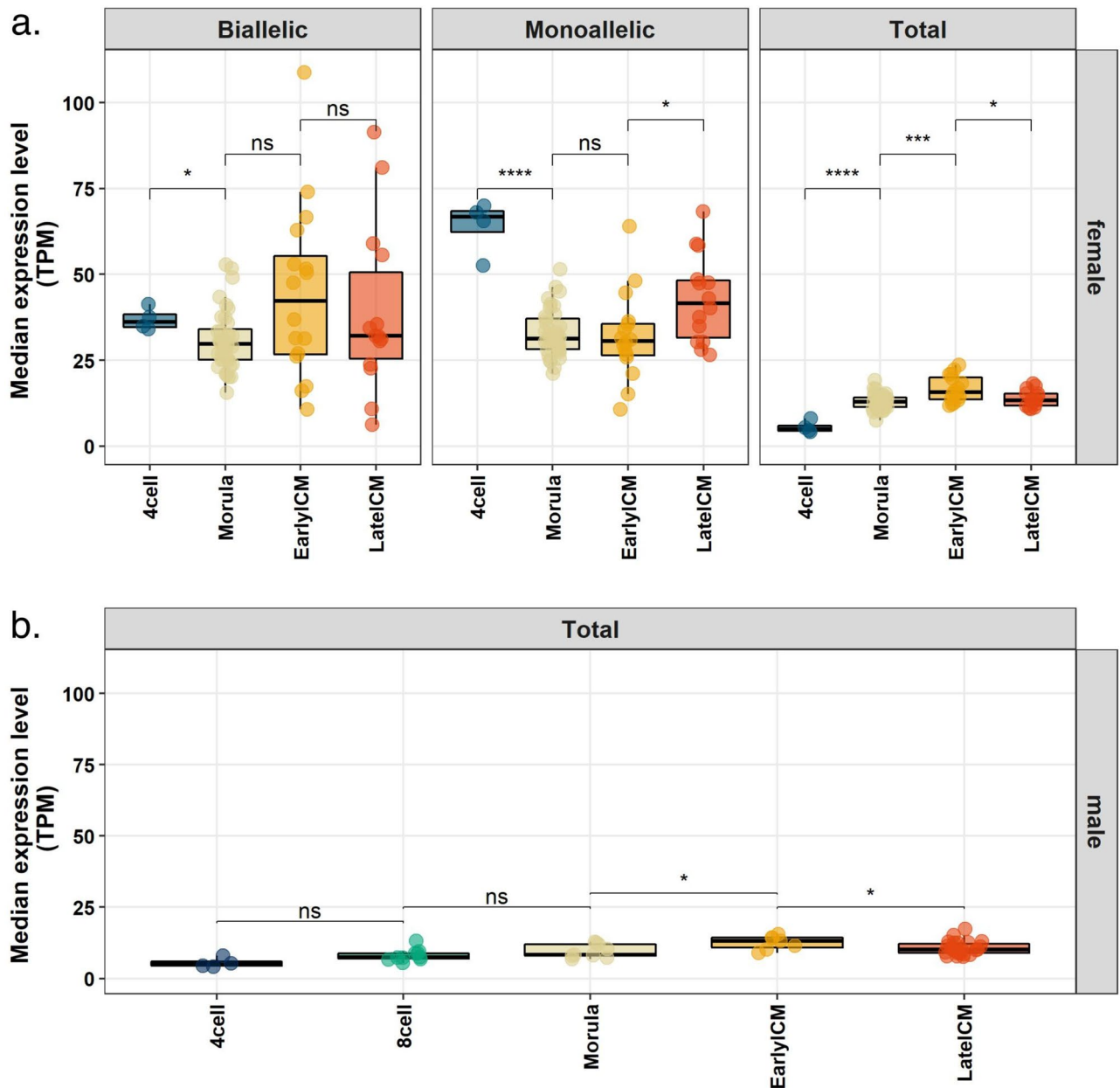


Figure 3. Expression levels of X-linked genes in marmoset preimplantation embryos. Each dot represents a single cell. Distribution of the median expression levels of X-linked genes as follows: (a) left panel: biallelically expressed; middle panel: monoallelically expressed; right panel: all X-linked expressed genes in each cell of female embryos (biallelic, monoallelic and non-informative). (b) Median expression levels of all expressed X-linked genes in each cell of male embryos. Stages were compared to each other using an unpaired Wilcoxon test. (*) P-value ≤ 0.05 ; (***) P-value ≤ 0.001 ; (****) P-value ≤ 0.0001 ; (ns) not significant (P-value ≥ 0.05).

this is not yet achieved by the blastocyst stage¹⁰. In marmoset female morulas, early and late blastocyst, the X:A ratio was greater than 1 but less than 2 (Fig. 4b), consistent with the occurrence of both X upregulation and incomplete XCI as observed in human embryos¹⁰. Moreover, X:A ratios were significantly different between male and female in all analyzed stages (Fig. 4a; morula, P-value = 1.9×10^{-5} ; early blastocyst, P-value = 5.4×10^{-5} ; late blastocyst P-value = 4.8×10^{-7}), further indicating incomplete XCI. Together with the analysis of allele-specific expression level (Fig. 3), these results suggest that X upregulation takes place and is complete earlier than XCI during marmoset preimplantation embryonic development.

Discussion

Despite remarkable advances in dissecting the mechanisms of XCI in mice, much less is known about the process during human embryonic development. Still, important differences between the process in the two species have been identified, including lack of imprinted XCI in humans⁹ and different timing of random XCI initiation^{6–8,10},

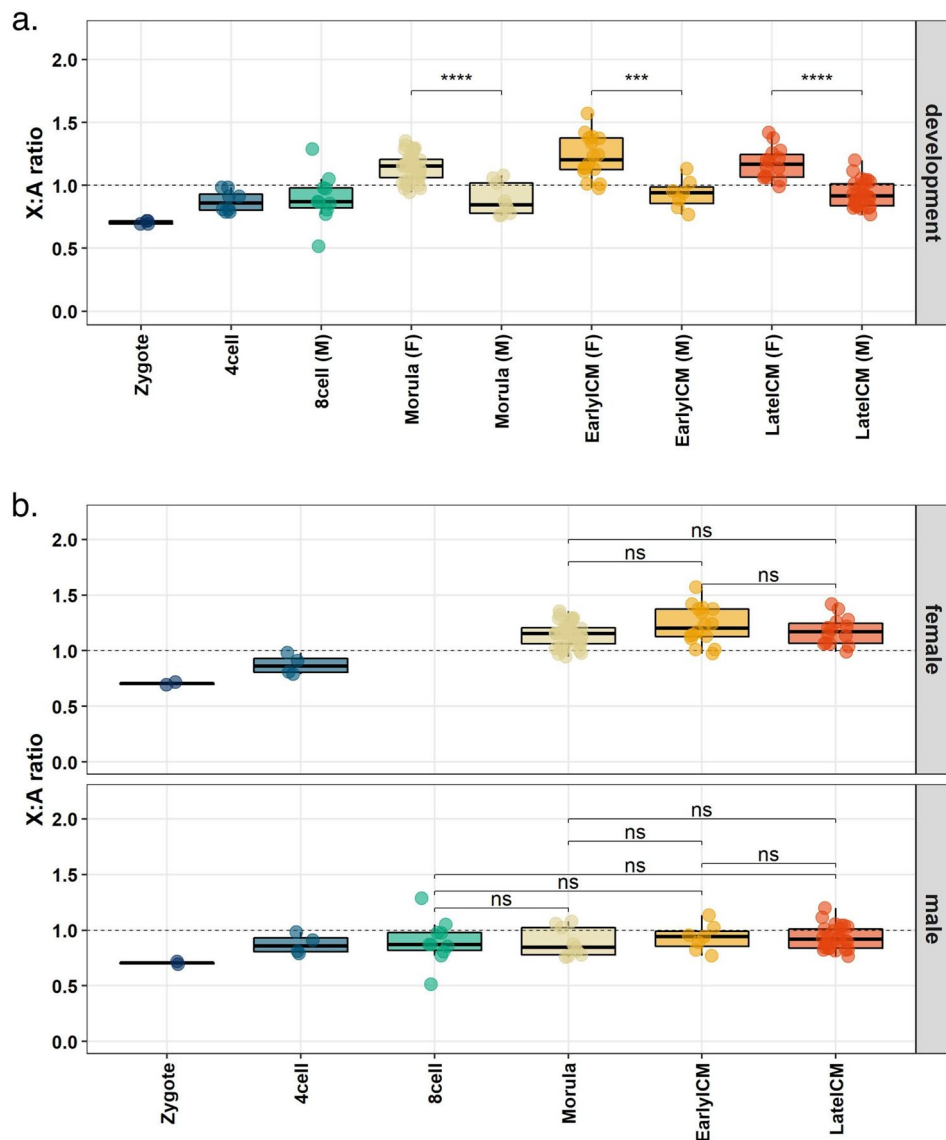


Figure 4. X-to-autosome dosage compensation. **(a)** X:A ratio distribution calculated per cell of each embryonic stage during marmoset preimplantation development. Unpaired Wilcoxon tests were used to compare males and females at the same developmental stage. (***) P-value ≤ 0.001 ; (****) P-value ≤ 0.0001 . **(b)** Same as in **(a)** with one panel for each sex. Top panel: female embryos; bottom panel: male embryos. Note that, although not sexed, zygotes and 4-cell stage embryos are duplicated in order to maintain panels uniform. Unpaired Wilcoxon tests were used to compare different stages in males and females separately. (ns) not significant (P-value ≥ 0.05).

although there is still disagreement regarding the latter. While some studies have identified initiation of random XCI in pre and perimplantation embryo stages^{5,10,11}, others argue that dosage compensation in human blastocysts is achieved by X chromosome dampening^{8,25}. Although the hypothesis of dampening has been also refuted in pigs²⁶, the study of X-linked dosage compensation in an animal model evolutionarily closer to humans may contribute in elucidating these issues. Therefore, here we sought to characterize the dynamics of X chromosome regulation during preimplantation development in another primate, the marmoset (*Callithrix jacchus*).

By analysing scRNA-seq data, we were able to perform a chromosome-wide analysis of X-linked genes in marmoset preimplantation embryos. Our results are consistent with an ongoing process of XCI in the early blastocyst of the female marmoset, characterized by an increase in *XIST* expression and in the percentage of monoallelically expressed X-linked genes, as observed in human preimplantation embryos^{10,11}.

Monoallelic expression of alternative *XIST* alleles in different cells of the same early blastocyst embryo suggests that XCI takes place in a random pattern. This is expected since the expression data were of cells from the ICM and adult marmoset females present random XCI²⁷. In mice, extraembryonic tissues present imprinted inactivation of the paternal X²⁸, whereas random XCI is present in human placenta⁹. Due to lack of scRNA-seq data from trophoctoderm cells, we were not able to exclude the existence of imprinted XCI in the marmoset.

Male marmoset early blastocysts showed upregulation of *XIST* comparable to that in female cells at the same stage, while in male late blastocysts *XIST* expression was significantly decreased. Interestingly, *XIST* expression has also been detected in rhesus macaques (*Macaca mulatta*) and human male blastocysts^{5,7,29}, suggesting that *XIST* expression in preimplantation male embryos is a feature of primates.

We further investigated X upregulation by analysis of the X:A ratio. We showed that male marmoset embryos achieve X:A dosage compensation during the preimplantation development. Accordingly, female marmoset embryos presented an X:A ratio between 1 and 2, suggestive of ongoing XCI in upregulated X chromosomes. In contrast, human embryos achieve X:A dosage compensation only after implantation¹¹.

Finally, we investigated the existence of X dampening in the marmoset embryos which should lead to decreased expression levels of biallelically expressed genes during development. Instead, we found no significant differences in expression levels of those genes between the morula and early/late blastocyst stages, refuting the occurrence of dampening.

In conclusion, despite the limited number of female embryos, our work adds to the panorama of XCI in primates, and establishes the marmoset as a model system where, as in humans, X-linked dosage compensation between males and females is achieved by random XCI which starts during preimplantation embryo development.

Methods

Transcriptome data. The marmoset (*Callithrix jacchus*) scRNA-seq raw data is available on the ArrayExpress platform (<https://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-7078. This dataset comprises the transcriptome of 196 cells obtained from 14 preimplantation embryos (from zygote to late blastocyst stage, Table 1).

Quality filter and mapping. Adapters and low-quality reads were removed using *cutadapt*³⁰ v1.18. In addition to removing bases with a *phred* score lower than 20, sequences with poor GC content were also trimmed out. Finally, reads shorter than 30 bp were discarded.

After trimming, TopHat2³¹ v2.1.1 was used to map reads to the marmoset assembly *Callithrix jacchus*-3.2 and its respective annotation (obtained from NCBI Assembly <https://www.ncbi.nlm.nih.gov/assembly/>, NCBI *Callithrix jacchus* Annotation Release 102, GCA_000004665.1). SAMTools³² v1.9 was used to remove unmapped reads and secondary alignments as well as for sorting. The numbers of reads assigned to each gene were counted with HTSeq using mode *intersection-nonempty*³³. We used the latest marmoset gene annotation from May 2020 (*Callithrix_jacchus_cj1700_1.1*).

Data pre-processing. The marmoset preimplantation embryo dataset contained originally the transcriptome of 196 isolated cells obtained from 14 marmoset preimplantation embryos, from the zygote to the late blastocyst stage. In order to identify cells with low-quality data, a Seurat object was created considering cells with at least 200 detected genes where those genes should be expressed in at least three cells. Through visual analysis we decided to retain only cells with more than 2×10^6 and less than 5.1×10^6 reads mapped to genes (Seurat package version 3.2.2)³⁴. After filtering out cells with low-quality data, we retained 124 good quality cells still representing all of the above-mentioned stages.

Sexing. In order to determine the sex of each embryo, we combined two strategies: (1) counting the number of Y-linked expressed genes, and (2) measuring the expression levels of those genes for each individual embryo. To be considered as an expressed gene, the normalized TPM level should be greater than or equal to one. All Y-linked genes detected in embryos at the zygote and 4-cell stages were removed from the analysis, as the transcriptome at these stages is mainly composed of maternal mRNA deposited during the oogenesis. It is not expected but also not unusual to detect expression of Y-linked genes from female transcriptomes due to the similar regions between the X and Y chromosomes in mammals (mainly the PAR regions). After the filtering steps we retained five protein coding genes located on the non-recombining region of the Y chromosome. Finally, we looked for statistical differences by comparing all embryos at the same stage for both the number of Y-linked genes and their expression level with the unpaired Wilcoxon test.

Genotyping and variant calling. Genotypes for each embryo were constructed by merging the transcriptome data of all cells from the same embryo and identifying heterozygous positions with the use of VarScan³⁵ v2.3.9. Variants were also identified in individual cells with VarScan v2.3.9 and filtered with bcftools³⁶ v1.9 in order to keep only biallelic single nucleotide polymorphisms in subsequent analyses. Each cell's variants were then compared to the respective embryo genotype, a strategy that allows for the identification of differential monoallelic expression in individual cells from the same embryo. Only variants covered by 20 or more reads were considered. GATK³⁷ v4.1.0.0 VariantsToTable was used to convert the VCF output from VarScan into a readable format for data analysis.

Gene calling and allelic expression. Allelic expression was inferred from the allelic frequency of each SNP. According to Borel et al., an allelic frequency of 0–20% or 80–100% of either the reference or alternative allele corresponds to monoallelic expression, while an allelic frequency of 20–80% indicates biallelic expression³⁸.

According to the position of each variant, the corresponding gene was identified. Each gene's allele-specific expression was inferred based on the detected SNPs. When all of the gene's variants presented monoallelic expression, the gene's expression was assumed to be monoallelic. The same applies to biallelic expression from

the SNPs. If variants from the same gene were not in agreement regarding the allelic expression (mono or biallelic), that gene was discarded from the analysis.

Allelic expression pattern. The proportion of mono and biallelically expressed genes was calculated as a percentage of the total informative genes in each cell. Informative genes correspond to those genes whose expression can be inferred from their SNPs allele frequency. Total informative genes represent the sum of mono and biallelically expressed genes.

Gene expression levels. To test the existence of dampening, we extracted the expression levels only for those genes whose expression was found to be biallelic. We then calculated the median expression levels in each cell, taking into consideration the biallelic genes with expression levels greater than 0.1 FPKM (fragments per kilobase of transcript, per million). We employed the same strategy to analyze the expression levels of genes that were found to be expressed monoallelically.

For the evaluation of global expression levels from the X chromosome, we considered all genes with expression levels greater than 0.1 FPKM regardless of being informative (excluding the ones removed from the allele-specific analyses). We then found the median expression levels for each cell and grouped the results according to each embryonic stage.

X:A ratio. We calculated the median X:A ratio by using all expressed X-linked and autosomal genes (expression level greater than 0.1 FPKM). We also calculated the median X:A ratios according to the method described by Fukuda et al., and only genes with expression levels greater than 0.1 FPKM were considered (excluding the ones removed from the allele-specific analyses)³⁹. Using the 'sample' function in R, we randomly selected 100 X-linked genes and 100 autosomal genes, for which an X:A ratio was calculated for each cell. This analysis was repeated by bootstrapping with 2000 replications, allowing for the calculation of median X:A ratios (Supplementary Fig. S8).

Statistics and visualization. Data visualization analyses and statistical tests were performed using R⁴⁰ v3.6.3 and *ggplot2* package⁴¹ v3.3.2.

Code availability

Scripts used in this work can be provided on reasonable request.

Received: 23 November 2020; Accepted: 16 April 2021

Published online: 05 May 2021

References

1. Lyon, M. F. Sex chromatin and gene action in the mammalian X-chromosome. *Am. J. Hum. Genet.* **14**, 135–148 (1962).
2. Brown, C. J. et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
3. Panning, B., Dausman, J. & Jaenisch, R. X chromosome inactivation is mediated by Xist RNA stabilization. *Cell* **90**, 907–916 (1997).
4. Rebuzzini, P., Zuccotti, M. & Garagna, S. X-Chromosome inactivation during preimplantation development and in pluripotent stem cells. *Cytogenet. Genome Res.* **160**, 283–294 (2020).
5. Vallot, C. et al. XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* **20**, 102–111 (2017).
6. van den Berg, I. M. et al. X Chromosome inactivation is initiated in human preimplantation embryos. *Am. J. Hum. Genet.* **84**, 771–779 (2009).
7. Okamoto, I. et al. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
8. Petropoulos, S. et al. Single-Cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
9. de Mello, J. C. M. et al. Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS ONE* **5**, 1–8 (2010).
10. de Mello, J. C. M., Fernandes, G. R., Vrbancovski, M. D. & Pereira, L. V. Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing. *Sci. Rep.* **7**, 10794 (2017).
11. Zhou, F. et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**, 660–664 (2019).
12. Boroviak, T. et al. Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development (Cambridge)* **145**(21), dev167833. <https://doi.org/10.1242/dev.167833> (2018).
13. Chitwood, J. L., Burrue, V. R., Halstead, M. M., Meyers, S. A. & Ross, P. J. Transcriptome profiling of individual rhesus macaque oocytes and preimplantation embryos. *Biol. Reprod.* **97**, 353–364 (2017).
14. Guo, F. et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452 (2015).
15. Raudsepp, T. & Chowdhary, B. P. The eutherian pseudoautosomal region. *Cytogenet. Genome Res.* **147**, 81–94 (2016).
16. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**(5441), 964–967. <https://doi.org/10.1126/science.286.5441.964> (1999) (Erratum in: *Science* **1999 Dec 17**; **286**(5448):2273. PMID: 10542153).
17. Kelkar, A., Thakur, V., Ramaswamy, R. & Deobagkar, D. Characterisation of inactivation domains and evolutionary strata in human X chromosome through Markov segmentation. *PLoS ONE* **4**(11), e7885. <https://doi.org/10.1371/journal.pone.0007885> (2009).
18. Jonkers, I. et al. RNF12 is an X-encoded dose-dependent activator of X chromosome inactivation. *Cell* **139**, 999–1011 (2009).
19. Barakat, T. S. et al. RNF12 activates Xist and is essential for X chromosome inactivation. *PLoS Genet.* **7**, 1–12 (2011).
20. Gontan, C. et al. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**, 386–390 (2012).
21. Barakat, T. S. et al. The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Mol. Cell* **53**, 965–978 (2014).
22. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, 1967). <https://doi.org/10.1111/j.1439-0272.1973.tb00879.x>.
23. Wang, F. et al. Regulation of X-linked gene expression during early mouse development by Rlim. *Elife* **5**, 1–19 (2016).

24. Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T. & Turner, J. M. A. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature* <https://doi.org/10.1038/s41586-020-2629-6> (2020).
25. Sahakyan, A. *et al.* Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
26. Ramos-Ibeas, P. *et al.* Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. *Nat. Commun.* **10**, 1–17 (2019).
27. Machado, F. B. *et al.* 5meCpG epigenetic marks neighboring a primate-conserved core promoter short tandem repeat indicate X-chromosome inactivation. *PLoS ONE* **9**, 1–13 (2014).
28. Borenstein, M. *et al.* Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat. Struct. Mol. Biol.* **24**, 226–233 (2017).
29. Tachibana, M. *et al.* X-chromosome inactivation in monkey embryos and pluripotent stem cells. *Dev. Biol.* **40**, 1301–1315 (2012).
30. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2013).
31. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, 957–961 (2013).
32. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
33. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
34. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**(7), 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031> (2019).
35. Koboldt, D. C. *et al.* VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).
36. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
37. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
38. Borel, C. *et al.* Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* **96**, 70–80 (2015).
39. Fukuda, A., Tanino, M., Matoba, R., Umezawa, A. & Akutsu, H. Imbalance between the expression dosages of X-chromosome and autosomal genes in mammalian oocytes. *Sci. Rep.* **5**, 2–10 (2015).
40. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).
41. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, fellowship 2018/14.856-8, ALC). J.C.M.M is supported by the Oncode Institute.

Author contributions

A.L.C. design, collection and assembly of data, data analysis and figures preparation; J.C.M.M. data analysis and figure preparation; J.G. data analysis; L.V.P. conception, design and supervision of the work. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-89175-7>.

Correspondence and requests for materials should be addressed to L.V.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021