Single Cell Analysis Reveals Partial Reactivation of X-chromosome Instead of chromosome wide dampening in Naïve Human Pluripotent Stem Cells

Mandal S, Chandel D, Kaur H, Majumdar S, Arava M, and Gayen S*

Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore-560012, India.

*Correspondence: srimonta@iisc.ac.in

Abstract

To balance the dosage of X-linked gene expression between the sexes, eutherian female mammals inactivate one of the two X-chromosomes present in a diploid cell. Once enacted in individual cells of the early female embryo, X-inactivation is stably transmitted such that all descendant cells maintain silencing of that X-chromosome. Recently, it has been demonstrated that in human however, X-chromosome dosage compensation initiates through the dampening of X-linked gene expression from both X-chromosomes in the preimplantation embryos and which later on switches to the X-chromosome inactivation. Subsequently, Sahakyan *et al.* found that conversion of primed human pluripotent stem cells (hPSCs) to naïve state is associated with X-chromosome dampening. However, both in embryos and hPSCs, the existence of dampened X-chromosomes remains controversial. Specially, in preimplantation embryos it has been shown that there is actually inactivation of X-chromosome instead of dampening. Here, we have performed allelic analysis of X-linked genes at single cell level in hPSCs and found that there is partial reactivation of the inactive X-chromosome instead of chromosome wide dampening upon conversion from primed to naïve state. In addition, our analysis suggests that the reduced X-linked gene expression in naïve hPSCs might be due to the erasure of X-chromosome upregulation.

Keywords: X-chromosome dampening; X-chromosome inactivation; X-chromosome upregulation, Human pluripotent stem cells; Naive & primed pluripotency; Human embryos; *Xist*

Introduction

In therian mammals, presence of two X-chromosomes in female and one X and one Y chromosome in male leads to the dosage imbalance of X-chromosome between the sexes. Therefore, to balance the X-chromosome dosage between male and female, one X-chromosome becomes inactive in female cells (Lyon, 1961). On the other hand, dosage imbalance between single active X-chromosome and two copies of autosomes (AA) is compensated through the upregulation of the active-X chromosome in both male and female (Deng et al., 2011; Larsson et al., 2019; Ohno, 1967). Recently, a unique form of X-chromosome dosage compensation has been inferred in human preimplantation embryos, which happens through the dampening of X-linked gene expression from both X-chromosomes (Petropoulos et al., 2016). However, X-chromosome dampening phenomenon in human embryos remains controversial (De Mello et al., 2017; Saiba et al., 2018). De Mello *et al.* found inactivation of X-chromosome instead of dampening when they reanalyzed the same scRNA-seq dataset of Petropoulos *et al.* with more stringency. Subsequently, Sahakyan *et al.* have found that naïve human pluripotent stem cells (hPSCs) also reflect the X-chromosome dampening of embryos (Sahakyan et al., 2017a). However, similar to the human embryos, X-chromosome states in hPSCs remains elusive (Kaur et al., 2019; De Mello et al., 2017). Therefore, here we have taken single cell approach with allelic resolution to get more insight about X-chromosome states in hPSCs using the single-cell RNA-sequencing (scRNA-seq) dataset from Sahakyan *et al.*

Results

Increased *Xist* expression and reduction in X-linked gene expression upon conversion of early to late naïve state

In rodents and human, two transient pluripotent state have been described: naïve and primed (De Los Angeles et al., 2015; Zhang et al., 2016). Pluripotency first arises in inner cell mass of preimplantation blastocysts in a naïve state and then transformed to a primed state in post implantation epiblast. Conventional hPSCs derived from blastocysts represent a primed state instead of naïve state and therefore, unable to recapitulate the preimplantation X-chromosome states (Nichols and Smith, 2009; Sahakyan et al., 2017b). Therefore, to model the preimplantation X-chromosome state, Sahakyan *et al.* converted primed hPSCs to a naïve state

(Fig. 1A). The primed cell line UCLA1, they used for their study, harbored one active-X chromosome and one inactive X chromosome (X_aX_i). The transition of primed to naïve state happened through an intermediate early naïve state (Fig. 1A). It was suggested that there was reactivation of the inactive-X chromosome upon transition from primed to the early naïve state and followed by X-chromosome dampening in late naïve cells (Fig. 1A). Moreover, X-reactivation was associated with the transformation of cells from *Xist* positive to negative state, whereas X-chromosome dampening was associated with the transformation of cells from *Xist* negative to positive state. In this study, we have analyzed scRNA-seq dataset of early naïve cells and late naïve cells (Fig. 1A). First, to test if the transition of early to late naïve state is associated with the conversion of cells from *Xist* negative to positive state, we analyzed the expression of *Xist* in early and late naïve cells. We found that indeed majority of the early naïve cells had very low levels of *Xist* expression, whereas late naïve cells mostly showed significantly higher levels of *Xist* expression as it was inferred by the Sahakyan *et al* (Fig. S1; Supplementary file1). Next, we proceeded for analysis of X-linked gene expression. We sorted out top 60 out of 96 early naïve cells based on low levels of *Xist* expression and top 59 out of 96 late naïve cells having highest levels of *Xist* expression. Finally, from this list we selected out top 28 early naïve cells and 30 late naïve cells based on total gene expression (RPKM sum and mean) (Fig. 1B; Supplementary file1) for comparison of X-linked gene expression analysis. Comparison of *Xist* expression in these cells, showed significant increase in *Xist* expression in late naïve cells (Fig. 1B). From the same cells (28+30), we selected 57 X-linked genes having mean RPKM ($5 < \mu_{RPKM} < 200$) for our analysis. We found that there was a significant reduction in X-linked gene expression in late naïve cells compared to the early naïve cells (Fig. 1C; Supplementary file1). Taken together, our analysis of scRNA-Seq data showed increased *Xist* expression and reduction in X-linked gene expression during transition of early to late naïve state just like the RNA-Seq analysis of whole cell population as it was described in Sahakyan *et al.*

Reduction in X-linked gene expression in late naïve cells is independent of *Xist*

X-chromosome dampening in preimplantation embryos is associated with the expression of *Xist* from both X-chromosomes. In fact, X-dampening initiates concomitantly with the initiation of *Xist* expression and therefore it is thought that *Xist* might have important role in dampening process. Therefore, we wanted to examine what fraction of cells show *Xist* expression from both X-chromosomes in late naïve cells. In fact, Sahakyan *et al.* tested allelic expression of *Xist* at single cell level using RNA-Seq data and found ~28% of cells having *Xist*

expression from both X-chromosomes. Consistent with the Sahakyan's analysis we also found ~32 % cells (8 of 25) having the *Xist* expression from both X-chromosomes (Fig. 1D; Supplementary file 2). Next, we tested if the reduction in X-linked gene expression in late naïve cells was restricted to the *Xist* biallelic cell population only as in embryos X-dampening is mostly associated with the biallelic *Xist* expression. To test this, we compared the global X-linked gene expression of *Xist* biallelic cells with the *Xist* monoallelic cells. However, we did not find any significant difference in X-linked gene expression between *Xist* biallelic and monoallelic cells (Fig. 1E; Supplementary file 2). Interestingly, we observed that there is no significant difference also when we compared with the *Xist* negative cells even (Fig. 1E; Supplementary file 2). Altogether, we concluded that there is no significant correlation between the reduction in X-linked gene expression and *Xist* expression in late naïve cells.

No evidence of X-chromosome inactivation or dampening

We then looked further about the mechanisms behind the reduction in X-linked gene expression in late naïve hPSCs. At first, we considered it could be due to the either X-chromosome inactivation or dampening. To examine if it is due to the X-inactivation, we analyzed allelic expression pattern of multiple X-linked genes distributed across the X-chromosome based on single nucleotide polymorphisms (SNPs) in both early and late naïve cells (n=12 SNPs covering 10 genes). Interestingly, we found allelic gene expression pattern between the cells varied dramatically both in early and late naïve cells (Fig. 2A; Fig. S2; Supplementary file 2). Importantly, there was incomplete reactivation of X-linked genes in early naïve state as there many cells showed monoallelic expression along with biallelically expressed X-linked genes (Fig. 2A; Fig. S2). Almost similar pattern of allelic expression of X-linked genes was observed in late naïve cells. In addition, we found that there is no difference in allelic expression pattern between *Xist* monoallelic and biallelic cells (Fig. 2A). To validate the accuracy of our allele specific expression analysis, we also analyzed allelic expression of genes of an autosome (Chr17). As expected, we found mostly biallelic expression of SNPs, thus validating our allelic expression analysis (Fig. 2A). Next, we calculated percent of SNPs showing biallelic and monoallelic expression in each cells of early and late naïve cells (Fig. 2B). We found that there was huge variation in percent of biallelic vs monoallelic expression between cell to cell and there was overall little increase in fraction of SNPs showing biallelic expression in late naïve cells compared to the early naïve cells. Overall, we found that both early and late naïve cells have partially reactivated X-chromosome. Moreover, frequency of reactivation is increased in late naïve cells compared to the early naïve cells, however reactivation remained incomplete.

We concluded that late naïve cells do not harbor inactive X-chromosome rather they have partially reactivated X-chromosome. If the late naïve cells underwent X-inactivation, then we should have expected monoallelic expression of most of the X-linked genes, which we did not observe. Next, we looked if X-dampening is leading the reduction in X-linked gene expression in late naïve cells. In fact, from the allelic analysis of X-linked gene expression it is clear that late naïve cells harbor partially reactivated X-chromosome and if they harbored dampened X-chromosomes, we should have expected biallelic expression of most of the genes, which we did not observe. In addition, we compared median expression of biallelically expressed genes of early cells with that of the late naïve cells. If there are X-dampening, we should have seen significant decrease in median expression. However, we did not find any significant differences (Fig. 2C). Therefore, we concluded that there is neither X-inactivation or dampening upon conversion of early to late naïve cells. Next, to find out the mechanism behind the reduction in X-linked gene expression in late naïve cells we checked X-chromosome ploidy of the cells. We thought, it may be possible that upon conversion of early to late naïve state, cells are losing one X-chromosome. Though, from the allelic expression pattern in late naïve cells, it is clear that cells harbored two X-chromosomes as evident by the biallelic expression of some X-linked genes (Fig. 2A). However, it is also possible that they may lose some part of the X-chromosome in late naïve cells and which is leading to the reduction in X-linked gene expression. To test this, we analyzed gene expression ratio of X-linked genes across the X-chromosome and we did not find significant differences between early vs. late naïve cells (Fig. 2D; Supplementary file 3). Therefore, we confirmed that loss of X-chromosome is not the cause of reduction in X-linked gene expression in late naïve cells, some other mechanism might be existing for this.

Erasure of X-upregulation might be the cause of reduced X-linked gene expression in late naïve cells

We investigated if the erasure of X-chromosome upregulation might be leading the reduction in X-linked gene expression in late naïve cells. Sahakyan *et al.* in their analysis of X-linked gene expression did not consider the possibility of having upregulated X-chromosome in primed cells. Recent years, existence of X-chromosome upregulation has been increasingly documented in mammals (Chen and Zhang, 2016; Lin et al., 2011; Sangrithi et al., 2017; Wang et al., 2016). Moreover, De Mello *et al.* recently have shown that indeed primed hPSCs harbor upregulated active-X chromosome. Based on this, we first wanted to confirm if UCLA1 primed cells truly harbor upregulated active-X chromosome. To test this, we calculated X to autosomal (X:A) expression ratio. For this, we used whole cell population RNA-Seq dataset as the

162 scRNA-Seq dataset was not available for UCLA1 primed cells. If primed cells harbor one
163 upregulated active-X chromosome and one inactive-X chromosome then we should expect X:A
164 ratio more than 0.5 and near to 1. Indeed, we found X:A ratio ~0.8 in primed cells, which
165 strongly suggests that UCLA1 primed cells harbor upregulated active-X chromosome (Fig. 2E;
166 Supplementary file 3). Interestingly, we found significant increase in X:A ratio upon transition
167 of primed to early naïve state (Fig. 2E). As there was X-reactivation upon transition of primed
168 to early naïve cells, it is obvious that X:A ratio will increase provided there was not complete
169 erasure of X-upregulation as well. On the other hand, we observed decrease in X:A ratio from
170 early to late naïve state, which indicated that it might be possible that the X-upregulation is
171 getting erased upon transition of early to late naïve cells and thereby leading to the overall X-
172 linked gene expression. Albeit, the decrease in X:A ratio from early to late naïve state is not
173 significant as expected. We think increase in biallelic X-liked gene expression from early to
174 late naïve state is masking this to some extent. To make sure that changes in X:A ratio is not
175 contributed by the autosomal gene expression we compared autosomal gene expression in
176 primed early naïve and late naïve cells, which showed no significant differences (Fig. 2F).
177 Altogether, we think that X-dampening is not the causing component for the reduction in X-
178 linked gene expression in late naïve cells and erasure of X-upregulation might be leading to
179 the decrease of X-linked gene expression (Fig. 2G).

180

181 **Discussion**

182 The initiation of X-chromosome dosage compensation in human remains elusive till date.
183 Although it has been proposed recently that X-chromosome dosage compensation is initiated
184 in female human preimplantation embryos through the dampening of X-linked gene expression
185 from both X-chromosomes, other studies have provided evidence of inactivation of one of the
186 X-chromosome. The major reason behind why till date this issue remains unresolved is the lack
187 of availability of surplus number of human embryos for the experiment. Therefore, hPSCs
188 derived from the human embryos serve as the alternative system to resolve this issue. However,
189 conventional hPSCs represent the primed state instead of the naïve state (Davidson et al., 2015;
190 Sahakyan et al., 2017b). Recently, Sahakyan *et al*. converted primed hPSCs to naïve state to
191 model the X-chromosome states of preimplantation embryos. They suggested that naïve hPSCs
192 have dampened X-chromosomes like preimplantation blastocysts, which strengthened the
193 dampening observation of human embryos a step further. In this study, using scRNA-seq

194 dataset, we have demonstrated that upon conversion of primed hPSCs to naïve state, there is
195 partial reactivation of the inactive-X chromosome instead of the X-chromosome dampening.

196 Sahakyan *et al.* demonstrated that conversion of primed to naïve state happened through one
197 intermediate early naïve state; i.e. Primed to early naïve and then to late naïve (Fig1A). They
198 studied X-chromosome dynamics during the conversion of primed to naïve state using bulk
199 RNA-sequencing using cell population. Here, we have analyzed the single cell RNA-Seq
200 dataset generated from the early naïve state and late naïve state. Single cell RNA-Seq provides
201 better clarity on cellular state and gene expression to distinguish the heterogeneity among the
202 cells in a population. We found that early naïve cells were mostly *Xist* negative and becomes
203 *Xist* positive in late naïve state along with the reduction in X-linked gene expression, which is
204 consistent with the Sahakyan *et al.* Moreover, we found in late naïve state 32% cells showed
205 *Xist* expression from both X-chromosomes, similar to Sahakyan *et al.* In contrary, in
206 preimplantation embryos ~85% cells express *Xist* from both X-chromosomes (Okamoto et al.,
207 2011). It is believed that *Xist* might have important role in X-dampening. However, we found
208 that reduction in X-linked gene expression was *Xist* independent as *Xist* biallelic, monoallelic
209 and negative cells showed almost similar level of gene expression in late naïve cells (Fig. 1E).
210 Next, our allelic analysis of X-linked gene expression revealed that early and late naïve cells
211 harbored partially reactivated X-chromosome (Fig. 2A), which is dissimilar to the observation
212 of Sahakyan *et al.* where they found almost all genes they analyzed showed biallelic
213 expression. Importantly, we found huge variation in allelic pattern of gene expression in
214 different cells, such as same gene in some cells is monoallelic, whereas in others biallelic.
215 Therefore, in that scenario, whole cell population analysis will always show biallelic
216 expression of the genes, indeed which was the case in Sahakyan's study. In addition, we
217 observed that the frequency of biallelic is increasing in late naïve cells compared to the early
218 naïve cells, which indicated that the cells are still undergoing through the reactivation process.
219 Besides this, we did not find significant difference of the median expression of biallelically
220 expressed genes between early and late naïve cells, which indicated that there is no X-
221 dampening. Altogether, we concluded that the conversion of primed to naïve state is associated
222 with the incomplete reactivation of X-chromosome instead of X-inactivation or X-dampening.
223 Importantly, some observations by Sahakyan *et al.* also indicated that there is an incomplete
224 erasure of epigenetic memory on the inactive X in the naïve state. They found, upon
225 differentiation, naïve hPSCs undergo non-random X-inactivation dissimilar to the normal
226 development, and the same X-chromosome that was inactive in the primed hPSCs got selected

227 for inactivation. Even, the naïve cells showed an accumulation of H3K27me3 repressive marks
228 on one of the X-chromosomes, that has not been seen in preimplantation blastocyst (Okamoto
229 et al., 2011). We think presence of H3k27me3 is the result of incomplete erasure of inactive-X
230 marks, which support our observation of partially reactivated X-chromosome. Taking these
231 facts together, we strongly suggest that the naïve cells is still in the process of removal of
232 inactive-X epigenetic marks. In contrary, Sahakyan *et al.* found that RNA-fluorescent in situ
233 hybridization (FISH) analysis of three X-linked genes (*HUWE1, ATRX, and THOC2*) showed
234 biallelic expression in naïve cells. We think this small number of genes assayed is not enough
235 to conclude the X-chromosome wide gene expression status.

236 On the other hand, we found presence of upregulated X-chromosome in primed UCLA1.
237 Moreover, our analysis indicates that there might be erasure of X-upregulation upon conversion
238 of primed to naïve state. In fact, Theunissen *et al.* found a reduction in X-linked gene
239 expression in male naïve hPSCs compared to that in primed hPSCs, which may be due to the
240 erasure of X-chromosome upregulation (Theunissen et al., 2016). Our data also indicate that
241 the erasure of upregulation is might be incomplete in early naïve state and gets completed in
242 late naïve state (Fig. 2E; 2G). In that scenario, we believe that erasure of X-upregulation is
243 leading to the reduction in X-linked gene expression upon transition of early to late naïve state.
244 Therefore, in future, thorough analysis of X-chromosome dynamics in light of erasure of X-
245 upregulation is important.

246 In conclusion, our study provides compelling evidence that X-chromosome states in naïve
247 hPSCs do not reflects the X-chromosome states of human preimplantation embryos. We
248 suggest that the conversion of primed cells to the naïve state leads to the partial reactivation of
249 the inactive X-chromosome instead of chromosome wide dampening. Importantly, our data
250 indicates that erasure of X-upregulation might leading the reduction in X-linked gene
251 expression in naïve hPSCs. Finally, we think more improvement is necessary to establish a
252 suitable culture condition to establish naïve hPSCs recapitulating the X-chromosome states of
253 preimplantation embryos.

254

255

256

257

258 **Methods**

259 **Key resources table**

Softwares/Tools	Source/ Reference	Identifiers (RRID)
FASTQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	SCR_014583
STAR	https://github.com/alexdobin/STAR	SCR_015899
SAMtools	https://github.com/samtools/samtools	SCR_002105
HTSeq-count	https://htseq.readthedocs.io/en/release_0.9.1/	SCR_005514
Picard	http://broadinstitute.github.io/picard/	SCR_006525
GATK	https://software.broadinstitute.org/gatk/	SCR_001876
R	https://www.r-project.org/	SCR_001905
CGH Explorer	http://www.softgenetics.com/CGHExplorer.html	SCR_010920

260

261 **Experimental model and subject details** NA

262

263 **Method details**

264 **Data acquisition:** RNA-Seq datasets used for the analysis were acquired from Gene
265 Expression Omnibus (GEO) under the accession number GSE872309 (Sahakyan et al., 2017a).
266 We analyzed 192 single cell RNA-Seq data consisting of 96 early naive (Exp2) and 96 late
267 naive (Exp1) cells. We also analyzed whole cell population RNA-Seq data from UCLA1
268 primed, UCLA1 Clone4 *Xist* negative, UCLA1 Clone12 *Xist* positive, UCLA1 Clone9 *Xist*
269 positive, each having two replicates.

270 **Reads mapping and counting:** Reads were mapped to the human genome (hg38) using STAR
271 with default settings (Dobin et al., 2013). Mapped reads were then processed using SAM tools
272 (Li et al., 2009) and the number of reads mapping to each gene was counted using HTSeq-
273 count (Anders et al., 2015).

274 **Expression analyses:** For both single cell data and whole cell population dataset, the counted
275 reads from HTSeq were used to calculate RPKM (Reads Per Kilobase per Million) using
276 rpkmforgenes (Ramsköld et al., 2009). For single cell RNA-Seq analysis, we checked
277 for *Xist* expression in all 192 single cells. For early naive cells we selected 60 cells out of 96

showing very low *Xist* expression (0 to 0.5 rpkm) and for late naive cells, 59 out of 96 showing very high *Xist* expression (3 to 39 rpkm). For quality control of the selected (59+60) cells, we considered RPKM sum (> 500) and mean (> 0.2) and we reduced to 28 early naive cells and 30 late naive cells. Outlier cells were removed from the further analysis. To check the expression of X-linked genes, we considered only those genes which were having $5 < \mu_{RPKM} < 200$ (n=57 genes) and these genes were only considered for further downstream analyses.

Variant calling and SNP annotation: Aligned reads from the 58 (28 early naive + 30 late naive) single cells were marked for duplicate reads using Picard tools v2.18.11(https://broadinstitute.github.io/picard/). To retrieve allelic read counts for SNPs, GATK (v3.8) (Van der Auwera et al., 2013; Depristo et al., 2011) "HaplotypeCaller" was used. Only validated SNPs specific to UCLA1 cell line (GSM2420529) were taken into consideration for downstream analysis. SNPs were then annotated using dbSNP Build 152 (GRCh38.p12). We considered a SNP as informative, which showed biallelic expression in at least one cell.

Allele expression analyses: The allelic expression of two different alleles was calculated by directly counting the allele-specific reads covering a SNP position mapped to the reference or the alternative allele and then dividing it by the total number of reads covering that position. Low coverage SNPs (≤ 5 reads) were excluded from the analysis, and SNPs that occurred at least in ten different cells were only retained. Biallelic expression was considered when an allele expression ratio ranged from 0.1 to 0.9 and monoallelic expression otherwise. Also, we only considered allelic ratios of SNPs (n=12, covering 10 X-liked genes) that were having RPKM ≥ 4 or else deemed as not determined.

Ploidy: To analyze the X-chromosomal ploidy, we considered 73 genes having RPKM mean ≥ 3 distributed across the X-chromosome for 56 cells (28 early naive cells + 28 late naive cells). We calculated gene expression ratio through dividing the individual gene's RPKM by its median value across all the 56 cells. To visualize the profile of X-chromosome in all these cells, the normalised gene expression values (gene expression ratio) were used to build a moving average plot (neighbourhood size, k=9) using methods described by Mayshar et al. (Mayshar et al., 2010).

X chromosome to autosomes expression ratio: For calculation of X to autosomal gene expression ratio (X:A) we used whole cell population RNA-Seq dataset. For each cell line, normalized gene expressions for all the expressed autosomal and X-linked genes (RPKM>0) excluding escapees and genes in the pseudo autosomal regions of X-chromosome were

310 considered for X:A ratio. X:A ratios were calculated by dividing the median RPKM values of
311 the X chromosome by those of the autosomes. Also, for comparison of autosomal gene
312 expression analysis in these cell lines, we considered the expressed genes with RPKM>0.

313 **Statistical tests and Box-plots:** All statistical tests and boxplots were performed using R
314 version 3.5.1(R Development Core Team, 2016).

315

316 **Figure legends**

317 **Figure 1: Transition of early naïve to late naïve state is associated with increased *Xist***
318 **expression and reduction in X-linked gene expression.** (A) Schematic representation of
319 different stages of conversion of primed hPSCs to naïve state as described in Sahakyan *et al*.
320 We performed analysis of single cell RNA-Seq dataset acquired from early and late naïve state.
321 (B) Comparison of *Xist* expression (RPKM) between early naïve (n=28 cells) and late naïve
322 cells (n=30 cells). $p<0.00001$ (Mann Whitney U test) (C) Comparison of X-linked gene
323 expression (n=57) between early and late naïve state. $p<0.00001$ (Mann Whitney U test) (D)
324 Allelic expression of *Xist* in late naïve cells (n=25 cells). (E) Comparison of X-linked gene
325 expression among Xist biallelic, Xist monoallelic and Xist negative cells. Non-significant at
326 $p<0.05$ (Kruskal-Wallis test)

327 **Figure 2: Reduction in X-linked gene expression in naïve hPSCs not because of X-**
328 **dampening and might be due to the erasure of X-upregulation.** (A) Allelic expression
329 analysis of X-linked genes in early and late naïve cells at single cell levels. Right, validation
330 of our allelic gene expression analysis by analyzing allelic expression of an autosome (Chr17).
331 (B) Histogram showing the percent of SNPs showing monoallelic and biallelic in each cells of
332 early and late naïve state. (C) Comparison of median of expression level of biallelically
333 expressed genes between early and late naïve cells. Non-significant at $p<0.05$ (Kruskal-Wallis
334 test) (D) Ploidy analysis of the cells of early and late naïve state through the analysis of gene
335 expression ratio across the X-chromosome. (E) Analysis of ratio of X-linked gene expression
336 and autosomal gene expression in primed, early naïve and late naïve cells. $P<0.05$ (Student's
337 t-test) (F) Comparison of autosomal gene expression among primed, early naïve and late naïve
338 cells. Non-significant at $p<0.05$ (Kruskal-Wallis test) (G) Proposed model representing the X-
339 chromosome states during the conversion of primed hPSCs to the naïve state.

340

341 **Author's Contribution**

342 SG, SM (Susmita Mandal), DC, and HK conceptualized the study. SG supervised the study.
343 Bioinformatic analyses was done by SM (Susmita Mandal) and DC. SG, SM (Susmita
344 Mandal), DC, MA, and SM wrote, edited and proofread the manuscript. Final manuscript was
345 edited and approved by all the authors.
346

347 **Acknowledgments**

348 We thank Pavithra RV and Karunyaa M for their help in art work and Amitesh Panda for taking
349 part during the discussion. The study is supported by IISC-MHRD Start up grant,
350 Ramalingaswamy-DBT fellowship to SG, and IISC-DBT partnership program.

351

352 **References:**

353 Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-
354 throughput sequencing data. Bioinformatics *31*, 166–169.

355 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-
356 Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data
357 to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr.
358 Protoc. Bioinforma.

359 Chen, X., and Zhang, J. (2016). The X to autosome expression ratio in haploid and diploid
360 human embryonic stem cells. Mol. Biol. Evol. *33*, 3104–3107.

361 Davidson, K.C., Mason, E.A., and Pera, M.F. (2015). The pluripotent state in mouse and
362 human. Development *142*, 3090–3099.

363 Deng, X., Hiatt, J.B., Nguyen, D.K., Ercan, S., Sturgill, D., Hillier, L.W., Schlesinger, F.,
364 Davis, C.A., Reinke, V.J., Gingeras, T.R., et al. (2011). Evidence for compensatory
365 upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and
366 Drosophila melanogaster. Nat. Genet. *43*, 1179–1185.

367 Depristo, M.A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J.R., Hartl, C., Philippakis,
368 A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation
369 discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–

370 501.

371 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
372 M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics
373 *29*, 15–21.

374 Kaur, H., Rv, P., and Gayen, S. (2019). Dampened X-chromosomes in human pluripotent
375 stem cells: dampening or erasure of X-upregulation ? Chromosoma *1–3*.

376 Larsson, A.J.M., Coucoravas, C., Sandberg, R., and Reinius, B. (2019). Transcriptional
377 kinetics of X-chromosome upregulation. 1–6.

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

381 Lin, H., Halsall, J.A., Antczak, P., O'Neill, L.P., Falciani, F., and Turner, B.M. (2011).
382 Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with
383 Ohno's hypothesis. Nat. Genet. *43*, 1169–1170.

384 De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H.,
385 Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of
386 pluripotency. Nature *525*, 469–478.

387 Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (mus musculus L.).
388 Nature *190*, 372–373.

389 Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K.,
390 Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal
391 aberrations in human induced pluripotent stem cells. Cell Stem Cell *7*, 521–531.

392 De Mello, J.C.M., Fernandes, G.R., Vibranovski, M.D., and Pereira, L. V. (2017). Early X
393 chromosome inactivation during human preimplantation development revealed by single-cell
394 RNA-sequencing. Sci. Rep. *7*, 1–12.

395 Nichols, J., and Smith, A. (2009). Perspective Naive and Primed Pluripotent States. Stem Cell
396 *4*, 487–492.

397 Ohno, S. (1967). Sex chromosomes and sex-linked genes.

398 Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P.,

399 Wolf, J.P., Renard, J.P., Duranthon, V., et al. (2011). Eutherian mammals use diverse
400 strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

401 Petropoulos, S., Deng, Q., Panula, S.P., Codeluppi, S., Reyes, A.P., Linnarsson, S., Sandberg,
402 R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X-chromosome dynamics
403 in human preimplantation embryos. Cell.

404 R Development Core Team (2016). R: A language and environment for statistical computing.
405 R Found. Stat. Comput.

406 Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of
407 ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput.
408 Biol. *5*, e1000598.

409 Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E.,
410 Langerman, J., Clark, A.T., Jaenisch, R., et al. (2017a). Human Naive Pluripotent Stem Cells
411 Model X Chromosome Dampening and X Inactivation. Cell Stem Cell *20*, 87–101.

412 Sahakyan, A., Plath, K., and Rougeulle, C. (2017b). Regulation of X-chromosome dosage
413 compensation in human: Mechanisms and model systems. Philos. Trans. R. Soc. B Biol. Sci.
414 *372*.

415 Saiba, R., Arava, M., and Gayen, S. (2018). Dosage compensation in human pre-implantation
416 embryos: X-chromosome inactivation or dampening? EMBO Rep. *19*, e46294.

417 Sangrithi, M.N., Royo, H., Mahadevaiah, S.K., Ojarikre, O., Bhaw, L., Sesay, A., Peters,
418 A.H.F.M., Stadler, M., and Turner, J.M.A. (2017). Non-Canonical and Sexually Dimorphic X
419 Dosage Compensation States in the Mouse and Human Germline. Dev. Cell *40*, 289–301.e3.

420 Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J.,
421 Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular Criteria for Defining the
422 Naive Human Pluripotent State. Cell Stem Cell *19*, 502–515.

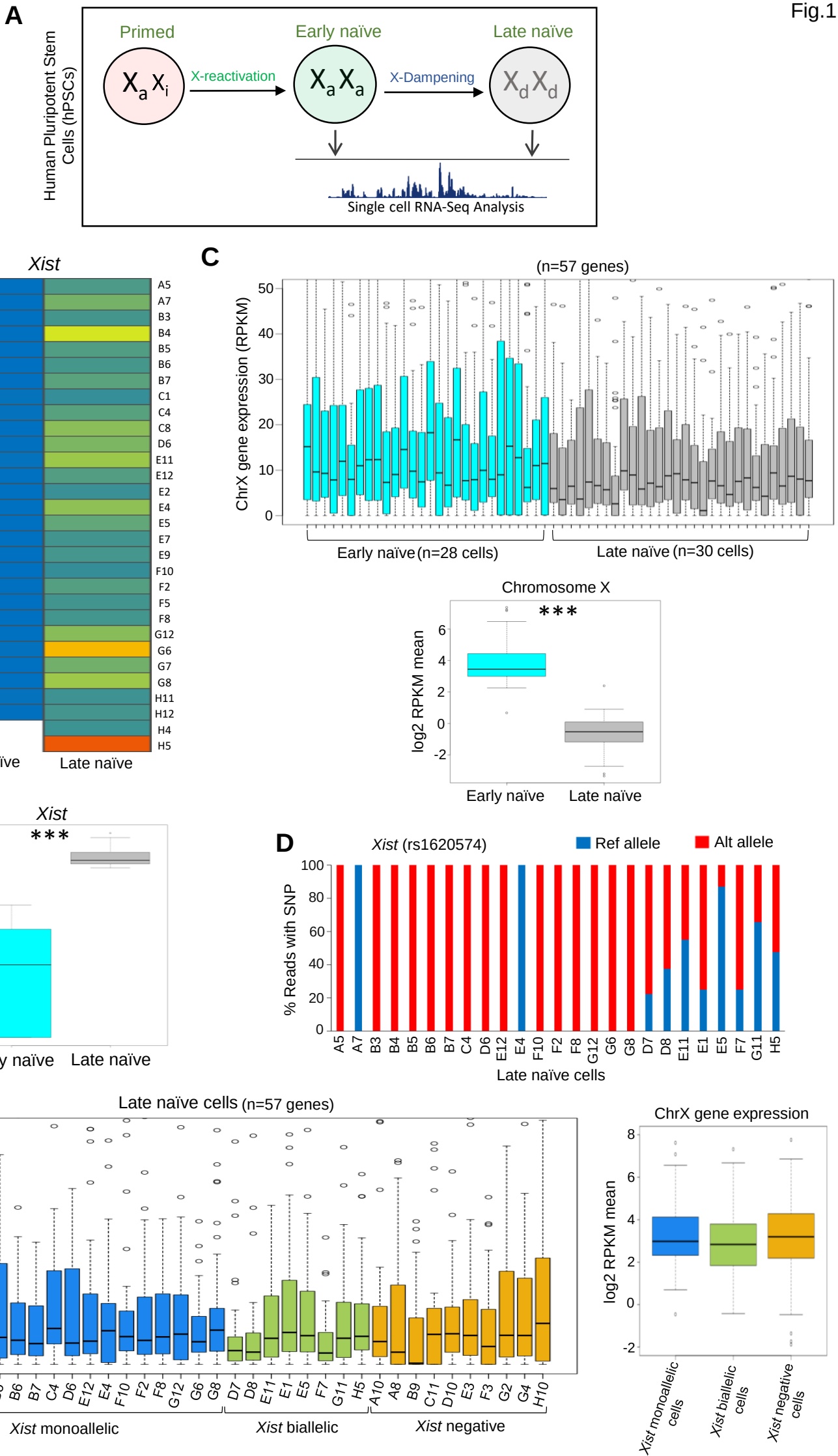
423 Wang, F., Shin, J., Shea, J.M., Yu, J., Bošković, A., Byron, M., Zhu, X., Shalek, A.K.,
424 Regev, A., Lawrence, J.B., et al. (2016). Regulation of X-linked gene expression during early
425 mouse development by Rlim. Elife *5*, 1–19.

426 Zhang, H., Gayen, S., Xiong, J., Zhou, B., Shanmugam, A.K., Sun, Y., Karatas, H., Liu, L.,
427 Rao, R.C., Wang, S., et al. (2016). MLL1 Inhibition Reprograms Epiblast Stem Cells to

428 Naive Pluripotency. Cell Stem Cell *18*, 481–494.

429

430


A
Human Pluripotent Stem Cells (hPSCs)
Primed
Early naïve
Late naïve
X-reactivation
X-Dampening
Single cell RNA-Seq Analysis
B
Xist
RPKM
Early naïve
Late naïve
C
(n=57 genes)
ChrX gene expression (RPKM)
Early naïve (n=28 cells)
Late naïve (n=30 cells)
Chromosome X
log2 RPKM mean
***
Early naïve
Late naïve
Xist
log2 RPKM
***
Early naïve
Late naïve
D
Xist (rs1620574)
Ref allele
Alt allele
% Reads with SNP
Late naïve cells
E
Late naïve cells (n=57 genes)
ChrX gene expression (RPKM)
Xist monoallelic
Xist biallelic
Xist negative
ChrX gene expression
log2 RPKM mean
Xist monoallelic cells
Xist biallelic cells
Xist negative cells

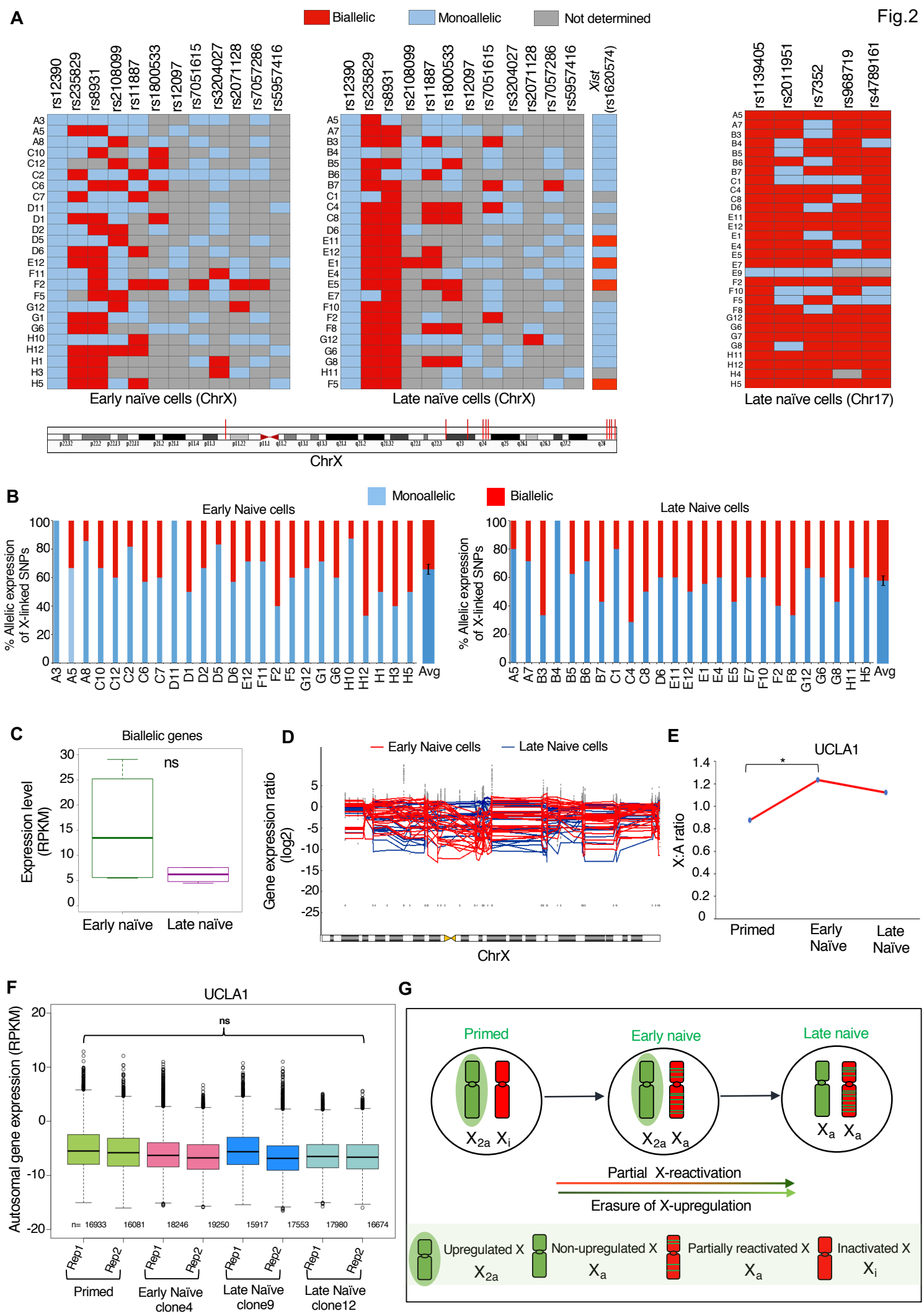


Fig.2
A
Biallelic
Monoallelic
Not determined
rs12390
rs235829
rs8931
rs2108099
rs11887
rs1800533
rs12097
rs7051615
rs3204027
rs2071128
rs7057286
rs5957416
Xist (rs1620574)
rs1139405
rs2011951
rs7352
rs968719
rs4789161
Early naïve cells (ChrX)
Late naïve cells (ChrX)
Late naïve cells (Chr17)
ChrX
B
Monoallelic
Biallelic
Early Naive cells
Late Naive cells
% Allelic expression of X-linked SNPs
Avg
C
Biallelic genes
ns
Expression level (RPKM)
Early naïve
Late naïve
D
Early Naive cells
Late Naive cells
Gene expression ratio (log2)
ChrX
E
UCLA1
X:A ratio
Primed
Early Naïve
Late Naïve
F
UCLA1
ns
Autosomal gene expression (RPKM)
n= 16933
16081
18246
19250
15917
17553
17980
16674
Rep1
Rep2
Primed
Early Naïve clone4
Late Naïve clone9
Late Naïve clone12
G
Primed
Early naive
Late naive
Partial X-reactivation
Erasure of X-upregulation
Upregulated X
Non-upregulated X
Partially reactivated X
Inactivated X