

 Open access • Posted Content • DOI:10.1101/382085

Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis — [Source link](#)

Qing Zhou, Taifu Wang, Lizhi Leng, Wei Zheng ...+9 more authors

Institutions: University of Copenhagen, Chinese Academy of Sciences, Central South University

Published on: 01 Aug 2018 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Dosage compensation, X-inactivation, Y chromosome, X chromosome and Embryo

Related papers:

- [Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis](#)
- [Allele-specific expression analysis reveals conserved and unique features of preimplantation development in equine ICSI embryos.](#)
- [Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells](#)
- [Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos.](#)
- [Effects of X Chromosome Number and Parental Origin on X-Linked Gene Expression in Preimplantation Mouse Embryos](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/single-cell-rna-seq-reveals-distinct-dynamic-behavior-of-sex-k7noqtblhg>

1 Single-cell RNA-seq reveals distinct dynamic behavior
2 of sex chromosomes during early human embryogenesis

3 Qing Zhou^{1,2,3 †}, Taifu Wang^{1,2 †}, Lizhi Leng^{4,5}, Wei Zheng⁶, Jinrong Huang^{1,2},
4 Fang Fang^{1,2}, Ling Yang^{1,2}, Jian Wang^{1,7}, Huanming Yang^{1,7}, Fang Chen^{1,2,3},
5 Ge Lin^{4,5,6,7}, Wen-Jing Wang^{1,2#}, Karsten Kristiansen^{1,2,3#}

6 ¹ BGI-Shenzhen, Shenzhen 518083, China,

7 ² China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China,

8 ³ Laboratory of Genomics and Molecular Biomedicine, Department of
9 Biology, University of Copenhagen, Copenhagen, Denmark,

10 ⁴ Institute of Reproductive and Stem Cell Engineering, School of Basic
11 Medical Science, Central South University, Changsha, China.

12 ⁵ Key Laboratory of Reproductive and Stem Cells Engineering, Ministry of
13 Health, Changsha, China.

14 ⁶ Reproductive & Genetic Hospital of CITIC-Xiangya, Changsha, China.

15 ⁷James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

16 ⁸ National Engineering and Research Center of Human Stem Cell, Changsha,
17 China.

18

19 [†] These two authors contributed equally to this work.

20 Correspondence and requests for materials should be addressed to Qing Zhou

21 (zhouqing1@genomics.cn), Wen-Jing Wang (wangwenjing@genomics.cn) or

22 Karsten Kristiansen (kk@bio.ku.dk)

23

24 **Abstract**

25 **Background:** Several animal and human studies have demonstrated that sex affects
26 kinetics and metabolism during early embryo development. However, the
27 mechanism governing these differences at the molecular level is unknown,
28 warranting a systematic profiling of gene expression in males and females during
29 embryogenesis.

30 **Findings:** We performed comprehensive analyses of gene expression comparing
31 male and female embryos using available single-cell RNA-sequencing data of 1607
32 individual cells from 99 human preimplantation embryos, covering development
33 stages from 4-cell to late blastocyst (E2 to E7). Consistent chromosome-wide
34 transcription of autosomes was observed, while sex chromosomes showed
35 significant differences after embryonic genome activation (EGA). Differentially
36 expressed genes (DE genes) in male and female embryos mainly involved in the cell
37 cycle, protein translation and metabolism. The Y chromosome was initially
38 activated by pioneer genes, *RPS4Y1* and *DDX3Y*, while the two X chromosomes in
39 female were widely activated after EGA. Expression of X-linked genes in female

40 significantly declined at the late blastocyst stage, especially in trophectoderm cells,
41 revealing a rapid process of dosage compensation.

42 **Conclusions:** We observed imbalanced expression from sex chromosomes in male
43 and female embryos during EGA, with dosage compensation occurring first in female
44 trophectoderm cells. Studying the effect of sex differences during human
45 embryogenesis, as well as understanding the mechanism of X chromosome
46 inactivation and its correlation with early miscarriage, will provide a basis for
47 advancing assisted reproductive technology (ART) and thereby improve the
48 treatment of infertility and possibly enhance reproductive health.

49 **Key words:** single-cell RNA-seq, embryogenesis, sex differences, dosage
50 compensation

51

52 **Background**

53 From the moment of fertilization in mammals, the sex of the preimplantation
54 embryo is determined by the spermatozoon carrying either an X or Y chromosome [1,
55 2]. In recent years, several mammalian and human studies have aimed to
56 molecularly and functionally characterize male and female embryos during *in vitro*
57 development [3, 4]. There are three main aspects of IVF embryos that differ between
58 males and females: 1) patterns of development, including morphology and gene
59 transcription; 2) kinetics and timing of development, including growth rates; 3)
60 mortality during intrauterine development [5]. At the 2-cell stage the percentage of

61 successful culturing differs between male and female mouse embryos [6]. Mouse
62 embryos that carry a Y chromosome develop more quickly *in vitro* than XX embryos
63 [7]. For bovine embryos, addition of the embryonic colony stimulating-factor 2 (CSF2)
64 to the culture medium increases the survival of female embryos at the morula stages,
65 but not male embryos [8]. Moreover, several animal studies have demonstrated that
66 sex affects metabolism during early embryonic development [1, 9, 10]. Thus, much
67 evidence regarding sex differences comes from animal models. For human embryos
68 derived via assisted reproductive technology (ART), it has been reported that
69 embryonic mortality before blastocyst formation is male-biased, as abnormalities
70 occur more frequently in male embryos [11]. In other studies, male IVF embryos are
71 reported to display an increased number of cells and higher metabolic activity than
72 female embryos and develop at a significantly faster rate [12, 13]. While these
73 observations clearly point to developmental differences between male and female
74 embryos, the molecular mechanisms governing these differences prior to the
75 expression of the sex-determine gene *SRY* remain to be established.

76 At the stage before implantation, sex-specific differences in gene expression
77 become apparent. These have been demonstrated initially in genes that derive from
78 sex chromosomes (at 8-cell stage in mouse [14]) , and later in the autosomes
79 (blastocyst in mouse [15]). In bovine embryos, expression of key enzymes involved in
80 establishing genome methylation, as well as histone methylation, is upregulated in
81 male blastocyst compared to their female counterparts [16]. For humans, although

82 Y-chromosome-driven effects have been detected in pluripotent stem cells in a
83 transcriptional study [17], a systematic profiling of gene expression comparing male
84 and female embryos during early development is needed.

85 Another elusive event in early embryogenesis is X chromosome dosage
86 compensation. In mouse, a period of double X chromosome activation occurs
87 between the 4-cell and 16-cell stages [18, 19]. The burst of transcription from both X
88 chromosomes results in a proteome exhibiting distinct differences between the
89 sexes. Failure to accomplish dosage compensation normally results in early
90 miscarriage and embryonic lethality [20, 21]. In human, it has been reported that X
91 chromosome inactivation occurs in all three lineages of *in vitro* blastocyst embryos
92 on day 7, and that the expression of both X chromosomes is reduced before the
93 random silencing of an entire X chromosome [22]. However, detailed information of
94 the precise temporal activation and inactivation of the X chromosome during early
95 development is still lacking.

96 The recent development of single-cell sequencing technology has allowed
97 characterizing of individual embryonic cells at multiple levels [23-28] providing
98 comprehensive transcriptional atlases [19, 22, 29, 30]. Here we aimed to determine
99 whether male and female embryos differ in relation to gene expression levels during
100 early development. By analyzing available transcriptome data, we revealed a
101 dynamic pattern of expression for the sex chromosomes and the process of X
102 chromosome dosage compensation in female embryos.

103 **Data description**

104 To examine whether sex differences already affect transcriptional patterns
105 during early human embryogenesis, we collected available sequencing data on the
106 transcriptome of 1607 individual cells from 99 human preimplantation embryos
107 ranging from the 4-cell stage to late blastocyst [22, 30] (Figure 1A, E2-E7). A total of 3
108 to 17 embryos and 12 to 466 cells were analyzed per stage (Figure 1B). Single-cell
109 RNA-seq data, processed data and raw reads, on human early embryonic
110 development were downloaded from two publicly available datasets: GSE36552 (78
111 cells from 4-cell to late blastocyst)[30]; ArrayExpress: E-MTAB-3929 (1529 cells from
112 E3 to E7 embryos)[22]. The DNA methylation data of human embryos was from
113 GSE49828 (covering developmental stages from 4-cell stage to post-implantation)
114 [25].

115 **Analyses**

116 **Transcriptional profiling reveals differences in expression patterns** 117 **between sex chromosomes during early embryogenesis**

118 We firstly generated the comprehensive transcriptional map of early embryos
119 during these stages. In the result of dimensionality reduction by t-distributed
120 stochastic neighbor embedding (t-SNE), we noticed that the primary segregating
121 factor is the time point of development, not the sex, as samples are clearly classified
122 in agreement with embryonic day (Figure 1C). We then investigated the expression
123 at the chromosome level comparing male and female embryos. Following EGA,

124 differences in gene expression become apparent. At 8-cell stage we observed
125 significant differences in transcription of genes on the sex chromosomes (Figure 1D,
126 $p < 0.00001$, Mann-Whitney-Wilcoxon test), and sex-dependent differences of
127 expression of genes on autosomes become detectable during development,
128 especially when embryos develop to late blastocysts (Figure 1E, Figure S1, $p < 0.00001$,
129 Mann-Whitney-Wilcoxon test).

130 Next, we performed differential expression analysis comparing female and male
131 cells within embryonic stages. The identified DE genes locate on both sex
132 chromosomes and autosomes for all stages (Table S1, S2). In agreement with
133 previously reported results [22] we found that there is a significant enrichment of DE
134 genes on sex chromosomes (Figure S2, $p < 0.001$, fisher's exact test). Functional
135 annotation based on Gene Ontology (GO) revealed a stage-specific function for these
136 DE genes (Figure 1G). At the 8-cell stage, these genes are mainly involved in cell cycle
137 control, cell division and chromosomal segregation, and later in the morula stage, DE
138 genes play a role in the chromatin organization. During formation of the blastocyst,
139 sex-dependent differences in gene expression are related to chromatin assembly,
140 translation elongation, and metabolism, and in the late blastocyst stage, DE genes
141 comprise genes involved in regulation of lipid transport and neuron differentiation.
142 All these results indicate that expression differences between males and females are
143 manifest already during early developmental stages of embryogenesis, and thus,
144 regulate various biological processes of development.

145 **Initial transcription of the Y chromosome is initiated by few pioneer**
146 **genes**

147 To further investigate the temporal expression patterns of genes on the sex
148 chromosomes, we profiled the expression of Y-linked genes. In total, we detected
149 27 Y-linked genes exhibiting distinct expression patterns during the early embryonic
150 stages (Figure 2A). For example, transcript from the *PCDH11Y* gene could be
151 detected in E2 embryos before EGA, while down-regulation and transcriptional
152 silencing are observed during later development. Notably, the sex-determining *SRY*
153 gene is as expected not activated at these early stages (Figure S3), reflecting that
154 male gonadal differentiation in the developing embryo first occurs
155 post-implantation [31]. The majority of the Y-linked genes exhibits low levels of
156 expression, but we noticed that two pioneer genes, *RPS4Y1* and *DDX3Y*, exhibit high
157 expression immediately after EGA. These two genes are widely expressed in all
158 male cells, and the sex-specific differences are maintained and even accentuated in
159 the following stages (Figure 2B). Furthermore, we could cluster embryos at the
160 8-cell stage into two separate groups according to sex solely based on the
161 expression of the *RPS4Y1* gene (Figure 2C). Our analysis reveal that only pioneer
162 genes are highly transcribed during the initial activation of the Y chromosome, and
163 the consistent and high expression in all male cells indicate that the *RPS4Y1* gene
164 could serve as a potential sex-specific marker in human early embryos.

165 **Both copies of X chromosome in female are widely activated during**

166 **EGA**

167 For the X chromosome, we examined the dynamic changes for all expressed
168 genes during the genomic activation process. Still, these X-linked genes, distributed
169 along the entire chromosome, exhibit higher expression in the female than in male
170 embryos after EGA (Figure 2D). Considering the extra copy of the X chromosome in
171 female, we analyzed the allele-specific expression of X-linked genes to determine
172 whether the higher level of expression reflects the activation of both copies. We
173 were able to analyze the allelic expression for each common single nucleotide
174 variant (SNV) present in the dbSNP database within each cell. For example, female
175 embryos at the 8-cell stage show bi-allelic expression of the *HNRNPH2* gene as both
176 a T and a G allele could be identified from the RNA-seq data, whereas the transcript
177 in male embryos harbors only a T allele after EGA (Figure 2E). This is also the case for
178 *DDX3X*, a gene escaping from X-inactivation, with approximate 50% percentage of
179 reads representing expression of the alternative allele in each female cell after EGA
180 (Figure 2F). All the above results demonstrate that all X chromosomes, both in male
181 and female, exhibit wide transcriptional activity during the process of EGA. The
182 transcription of the two copies in females result in an unbalanced dosage between
183 male and female embryos in these early stages.

184 **Dosage compensation of the X chromosome in female embryos first**
185 **occurs in the trophectoderm**

186 For adult females, there is a random X chromosome inactivation (XCI) to
187 equalize the expression of X-linked genes with males [32, 33]. As we observed an
188 unbalanced dosage of X chromosome expression comparing males and females in
189 the early embryonic stages, we further focused on the process of dosage
190 compensation in females. To analyze the process of dosage compensation in females,
191 we employed tSNE analysis using only expression of X-linked genes. Despite of the
192 primary classification of development stages, we noticed a sex-specific segregation
193 within each stage (Figure 3 A), except for a slight overlap for E7 embryos as they
194 exhibit an overall 70%–85% compensation of X chromosome at that time [22].

195 Since cells are designated as trophoctoderm (TE), primitive endoderm (PE), and
196 epiblast (EPI) at the blastocyst stage, we evaluated the X chromosome expression
197 dynamics in the different lineages at this stage. From the tSNE results, we found that
198 overlap between male and female samples could only be detected in TE cells at the
199 E7 stage (Figure 3B). Contrasting the stable expression pattern in male cells (Figure
200 2A), expression of the X chromosome-linked genes in female tends to be
201 down-regulated with time during the formation of blastocyst, especially in TE cells
202 (Figure 3C, Mann-Whitney-Wilcoxon test). As expected, autosomes show
203 comparable expression in all cells. Expression of X-linked genes declines
204 significantly in TE of E7 embryos, revealing a rapid process of dosage compensation
205 of the X chromosome.

206 Besides expression profiling, the DNA methylation landscape of specific marker
207 regions on the X chromosome can also reflect the status of gene activation or
208 inactivation [34]. To further support our finding that the first dosage compensation
209 occurs in TE cells, we investigated the DNA methylation level of four reported
210 markers: *AR*, *ZDHHC15*, *SLITRK4* and *PCSK1N* [35]. In total, we collected methylation
211 data for early embryos, including 4-cell, 8-cell, morula, PE (or ICM) and TE from late
212 blastocysts, and post-implantation embryos [25]. As expected, the regions near to
213 *PCSK1N* are hemimethylated in the post-implantation embryos, as one of the X
214 chromosome has completed the inactivation and become methylated at this stage
215 (Figure 3D). Interestingly, we also discovered a low methylation level for these DNA
216 sites in TE cells, comparing with the non-methylated landscape in PE cells (or ICM).
217 Although we could not obtain clear methylation profiles of the other three loci
218 (Figure S4), the specific pattern of *PCSK1N* indicates that methylation as well as
219 inactivation of the X chromosome first occur in female TE cells.

220 Discussion

221 Our study provides to the best of our knowledge the first information on
222 expression differences between male and female IVF embryos during early
223 development. The inclusion of a large number of embryos provides evidence that the
224 transcriptional differences are prominent on sex chromosomes. Our analysis
225 demonstrates distinct transcriptional patterns of genes on the sex chromosomes
226 during early embryogenesis, initial activation of pioneer genes on the Y chromosome

227 and activity of a broad region on the X chromosome. Thus, *RPS4Y1* exhibits high
228 expression at the time of EGA and shows a sex-specific expression pattern. In
229 humans, *RPS4Y1* is one of the variants encoding the ribosomal protein S4 (RPS4), and
230 its paralogous gene *RPS4X* is the first gene on long arm of the X chromosome known
231 to escape from X inactivation [36]. The amino acid differences between the
232 proteins encoded by these two genes result in the generation of two distinct, but
233 functionally equivalent, forms of ribosomes [37]. In contrast to the silencing of the
234 homologous genes in mouse [38], it has been assumed that normal human
235 development requires at least two *RPS4* genes per cell; two *RPS4X* in female cells
236 and one *RPS4X* and one *RPS4Y* in male cells. It has been reported that
237 haploinsufficiency of the ribosomal protein S4 genes may play a role in Turner
238 syndrome [39]. The high transcription of *RPS4Y1* help to balance the dosage between
239 sex [40] at the early stage (Figure S3) as there are two active X chromosomes in
240 female cells after EGA. Thus, expression of *RPS4Y1* may be used as a potential
241 marker to distinguish embryo sex at these stages, earlier than the expression of
242 other sex-determining genes.

243 The other pioneer gene on the Y chromosome, *DDX3Y*, belongs to the RNA
244 helicase family. The protein encoded by this gene shares high similarity to *DDX3X*, on
245 the X chromosome, while their functions differ [41]. As a result, activation of this
246 gene in the early stage may lead to a male-specific function, such as neuronal
247 differentiation [42] and translation initiation. In addition, certain part of the

248 ribosome family genes involved in translation elongation and metabolism shows
249 significant differences of expression (Figure S5). As a result, these genes may well
250 regulate protein synthesis and cell signaling pathways. As a sex-ratio bias in relation
251 to of embryonic mortality and growth rates during early development has been
252 reported, all these results may suggest a potential correlation between sex-specific
253 gene expression and the particular behavior of early embryos.

254 Our study revealed that most X-linked genes become transcriptional active
255 concomitant with completion of EGA in all embryos. In addition, both copies of the
256 X-chromosomes in female are activated. It has been generally assumed that the
257 germline-inactivated X might be passed onto the offspring as in two-cell mouse
258 embryos, where repetitive elements on the paternal X are suppressed [43, 44].
259 However, *de novo* inactivation of the paternal X chromosome in mouse embryos has
260 been reported [45, 46], with a re-inactivation taking place after the 4-cell stage [19].
261 For humans, we know that beyond completion of EGA at E4, female cells possess two
262 active X chromosomes [22]. From the comparison of transcription and the allelic
263 expression analysis, our research, for the first time, demonstrates that the two copies
264 of X chromosomes in female are widely activated immediately after genome
265 activation from the 4-cell to the 8-cell stage at E3.

266 The extensive datasets we investigated in the present study indicate that dosage
267 compensation of the X chromosome first occurred in TE cells. In mice, the imprinted
268 inactivation of the paternal X (Xp) chromosome occurs beyond the 4-cell stage [19].

269 Inactivity of the Xp is maintained in the TE but is reversed randomly in the ICM of the
270 blastocyst [47, 48]. Key genes, including *Atrx*, which are involved in chromatin
271 remodeling and heterochromatin formation and play a central role in the
272 X-chromosome inactivation process, have been found to be expressed in TE cells, but
273 not in other cell types (EPI). For humans, it has been reported that X chromosome
274 inactivation occurs in all three lineages at E7, and the expression of both X
275 chromosomes is reduced before the random silencing of an entire X chromosome
276 [22]. Our finding of first inactivation in TE cells raises the question as to whether
277 lineage-specific factors, similar to the situation in mouse, can regulate the process of
278 inactivation. Besides, fast inactivation of the X chromosome in TE cells, especially in
279 polar cells (Figure S6) where the first interaction between embryos and uterus occurs
280 during implantation, may result in a balanced dosage between the embryo and the
281 maternal endometrium. Thus, it may be beneficial in relation to implantation as
282 skewed X-chromosome inactivation is associated with recurrent miscarriage [48, 49].
283 While understanding whether this initial inactivation is paternal imprinted or
284 proceeds randomly is also an important question for the future.

285 In conclusion, we provide a comprehensive comparison of the transcriptional
286 atlas of male and female human preimplantation embryos and reveal the dynamics
287 of sex chromosomes expression and silencing during embryogenesis for the first time.
288 The precocious X inactivation and decrease in number of TE cells for IVF female
289 embryos may account for the observed preferential female mortality at early stages

290 and the sex ratio in the ART cycle [50]. Studying sex differences during human
291 embryogenesis, as well as understanding the process of X chromosome inactivation
292 and the correlation with early miscarriage, will expand the capabilities of ART and
293 possibly improve the treatment of infertility and enhance reproductive health. In
294 addition, this study of sex differences in early embryos will also provide a basis for
295 further experiments on how environmental impact during early developmental
296 stages can elicit profound and lasting effects that are different in male and female
297 offspring.

298 **Methods**

299 **Ethical approval**

300 Analyses performed at BGI comprised bioinformatics analysis of public sequencing
301 data, approved by the Institutional Review Board on Bioethics and Biosafety of BGI
302 (IRB 13067).

303 **Sequencing Data Processing**

304 For RNA-seq data, raw reads were mapped to the human genome (hg19) using
305 TopHat [51, 52] with default settings after removing the low-quality reads. Only
306 uniquely mapped reads were kept for further analysis. Chromosome level expression
307 was counted as chromosomal reads per kilobase of coding region within the
308 chromosome per million mapped reads (RPKM). The gene expression level of raw
309 reads count was calculated by HTSeq [53, 54] and RPKM values were estimated using
310 Cufflinks [51] with the annotation of RefSeq.

311 **Inference of embryonic sex and cell lineage**

312 Information on sex for each cell and embryo after EGA was classified as previously
313 described [22]. The embryos in the DNA methylation dataset were classified based
314 on the number of detected loci on the Y chromosome, using the average number in
315 oocytes and sperm as a baseline for female and male samples, respectively (Figure
316 S7). The three lineages of cells at the blastocyst stage, as well as the subpopulation
317 group of TE cells, were identified as previously reported [22].

318 **Sex differential expression analysis**

319 Differential expression analysis was performed for each stage comparing male and
320 female cells. P-values were calculated using DESeq2 [55] and a significant level
321 cut-off of adjusted $P < 0.05$ was used. A cut-off of a 2-fold change in expression was
322 used to define differentially expressed genes. We performed this analysis for each
323 dataset and stage separately, and then combined the results for further annotation.
324 The functional annotation was performed using the Database for Annotation,
325 Visualization and Integrated Discovery (DAVID) [56] Bioinformatics Resource. Gene
326 Ontology terms for each stage were plotted by the GOplot package in R and
327 summarized to a representative term.

328 **Analyses of allelic expression**

329 The alignment of raw sequencing reads to the human genome was performed by
330 BWA[57], then we employed the function of mpileup in SAMtools [58] to retrieve
331 allelic read counts in the RNA-seq data for common variants in db151[59], and

332 intergenic SNVs were excluded using ANNOVAR [60]. To obtain the total read counts
333 for each site, we run the mpileup program without base quality correction and
334 filtering.

335 **Statistical analyses**

336 Mann-Whitney-Wilcoxon analyses were performed in R. For identification of DE gene
337 identification, we used an adjusted p value <0.05 as cut-off. In the functional analysis,
338 only GO terms with $p < 0.01$ were included.

339 **Additional files**

340 **Additional file 1:** Table S1. Differentially analysis results comparing males and
341 females of each embryonic day for two datasets.

342 **Additional file 2:** Table S2. The distribution of DE genes on each chromosome.

343 **Additional file 3:** Table S3. Gene Ontology enrichment results of DE genes in each
344 stage.

345 **Additional file 4:** Figure S1. **(A-C)** Plots of additional stages (E4-E6) for genome-wide
346 expression per chromosome in female (pink box) and male (lightblue box) cells.

347 Chromosomal RPKM values were calculated as chromosomal reads per kilobase of
348 transcript per million reads mapped. The chromosome with a significant difference
349 was marked with a red star if $p < 10^{-5}$ in the Mann-Whitney-Wilcoxon test.

350 **Additional file 5:** Figure S2. **(A-B)** The number of significantly differentially expressed
351 genes on each chromosome comparing males and females at 8-cell stage and late
352 blastocyst in the first dataset (Yan *et al*), stratified by autosomes (green), X

353 chromosome (red) and Y chromosome (blue); **(C-G)** Number of DE genes for each
354 chromosome from E3 to E7 in the other dataset (Petropoulos *et al*). The significant
355 enrichment of DE genes on sex chromosomes was marked with red star (Fisher's
356 exact test, $p < 0.001$).

357 **Additional file 6:** Figure S3. **(A)** Boxplots showing expression level of *SRY* from E3 to
358 E7; **(B)** Expression level of *RPS4X* and expression of *RPS4X* and *RPS4Y* in males
359 (light-blue) and females (pink) at each embryonic day(E3-E7). The significant results
360 were marked if $p < 0.001$ in the Mann-Whitney-Wilcoxon test.

361 **Additional file 7:** Figure S4. Integrative genome view (IGV) of the DNA methylation
362 patterns of three reported loci (*AR*, *ZDHHC15*, *SLITRK4*) as markers for determining X
363 chromosome inactivation or activation patterns. The height of bars shows the
364 percentage of methylation at each loci, ranging from 0% to 100%. The genomic index:
365 chrX:66,765,297-66,765,584 (*AR*); chrX:74,694,462-74,694,958 (*ZDHHC15*);
366 chrX:142,722,666-142,723,065 (*SLITRK4*).

367 **Additional file 8:** Figure S5. Heatmap showing the expression of differentially
368 expressed ribosomal genes comparing male and female embryos at E6.

369 **Additional file 9:** Figure S6. The t-SNE plot of blastocyst cells at E6 **(A)** and E7 **(B)**
370 represented by the expression of X-linked genes. The assignment of subpopulation of
371 cells is indicated as mural (blue), polar (orange) and others (red) for males (triangle)
372 and females (dot).

373 **Additional file 10:** Figure S7. Bar chart indicating the number of CpG islands detected
374 on the Y chromosome for all embryos in the DNA methylation dataset.

375 **Conflict of interest**

376 The authors declared no conflict of interest.

377 **Funding**

378 This work is supported by the Shenzhen Municipal Government of China (No.
379 JCYJ20170412152854656 and No. JCYJ20160429174400950).

380 **Author's contribution**

381 Q.Z., W.J.W and K.K. conceived and designed the study; Q.Z., T.W., J.H., L.Y., F.F., L.L.
382 and W.Z. performed the data analysis; J.W., H.Y., F.C., G.L. oversaw the study; Q.Z., T.W.
383 prepared the figures; Q.Z., T.W., W.J.W. and K.K. wrote and revised the manuscript; all
384 authors reviewed the final version of the manuscript.

385 **Acknowledgements**

386 The authors thank Chen Ye for data download and management; Huanzi Zhong for
387 constructive advice about the final manuscript.

388 **Reference**

- 389 1. Alomar M, Tasiaux H, Remacle S, George F, Paul D, Donnay I. Kinetics of fertilization and
390 development, and sex ratio of bovine embryos produced using the semen of different bulls. *Animal*
391 *Reproduction Science*. 2008;107:48-61.
- 392 2. Setti AS, Figueira RC, Braga DP, Iaconelli A, Jr., Borges E, Jr. Gender incidence of intracytoplasmic
393 morphologically selected sperm injection-derived embryos: a prospective randomized study. *Reprod*
394 *Biomed Online*. 2012;24:420-3.
- 395 3. Serdarogullari M, Findikli N, Goktas C, Sahin O, Ulug U, Yagmur E, *et al.* Comparison of
396 gender-specific human embryo development characteristics by time-lapse technology. *Reprod Biomed*
397 *Online*. 2014;29:193-9.
- 398 4. Aiken CE, Swoboda PP, Skepper JN, Johnson MH. The direct measurement of embryogenic

-
- 399 volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction*.
400 2004;128:527-35.
- 401 5. Legato MJ. *Principles of Gender-Specific Medicine: Gender in the Genomic Era*.
402 ELSEVIER:Academic Press. 2017:292-3.
- 403 6. Sato E, Xian M, Valdivia RP, Toyoda Y. Sex-linked differences in developmental potential of single
404 blastomeres from in vitro-fertilized 2-cell stage mouse embryos. *Horm Res*. 1995;44 Suppl 2:4-8.
- 405 7. Valdivia RP, Kunieda T, Azuma S, Toyoda Y. PCR sexing and developmental rate differences in
406 preimplantation mouse embryos fertilized and cultured in vitro. *Mol Reprod Dev*. 1993;35:121-6.
- 407 8. Hansen PJ, Dobbs KB, Denicol AC, Siqueira LG. Sex and the preimplantation embryo: implications
408 of sexual dimorphism in the preimplantation period for maternal programming of embryonic
409 development. *Cell Tissue Res*. 2016;363:237-47.
- 410 9. Holm P, Shukri NN, Vajta G, Booth P, Bendixen C, Callesen H. Developmental kinetics of the first
411 cell cycles of bovine in vitro produced embryos in relation to their in vitro viability and sex.
412 *Theriogenology*. 1998;50:1285-99.
- 413 10. Kochhar HP, Peippo J, King WA. Sex related embryo development. *Theriogenology*. 2001;55:3-14.
- 414 11. Orzack SH, Stubblefield JW, Akmaev VR, Colls P, Munne S, Scholl T, *et al*. The human sex ratio
415 from conception to birth. *Proc Natl Acad Sci U S A*. 2015;112:E2102-11.
- 416 12. Ray PF, Conaghan J, Winston RM, Handyside AH. Increased number of cells and metabolic activity
417 in male human preimplantation embryos following in vitro fertilization. *J Reprod Fertil*.
418 1995;104:165-71.
- 419 13. Alfarawati S, Fragouli E, Colls P, Stevens J, Gutiérrez-Mateo C, Schoolcraft WB, *et al*. The
420 relationship between blastocyst morphology, chromosomal abnormality, and embryo gender. *Fertility*
421 *and Sterility*. 2011;95:520-4.
- 422 14. Lowe R, Gemma C, Rakyan VK, Holland ML. Sexually dimorphic gene expression emerges with
423 embryonic genome activation and is dynamic throughout development. *BMC Genomics*. 2015;16:295.
- 424 15. Kobayashi S, Isotani A, Mise N, Yamamoto M, Fujihara Y, Kaseda K, *et al*. Comparison of Gene
425 Expression in Male and Female Mouse Blastocysts Revealed Imprinting of the X-Linked Gene at
426 Preimplantation Stages. *Current Biology*. 2006;16:166-72.
- 427 16. Bermejo-Álvarez P, Rizos D, Rath D, Lonergan P, Gutierrez-Adan A. Epigenetic differences between
428 male and female bovine blastocysts produced in vitro. *Physiological Genomics*. 2008;32:264.
- 429 17. Ronen D, Benvenisty N. Sex-dependent gene expression in human pluripotent stem cells. *Cell Rep*.
430 2014;8:923-32.
- 431 18. Gardner DK, Larman MG, Thouas GA. Sex-related physiology of the preimplantation embryo. *Mol*
432 *Hum Reprod*. 2010;16:539-47.
- 433 19. Deng Q, Ramskold D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random
434 monoallelic gene expression in mammalian cells. *Science*. 2014;343:193-6.
- 435 20. Sullivan AE, Lewis T, Stephenson M, Odem R, Schreiber J, Ober C, *et al*. Pregnancy outcome in
436 recurrent miscarriage patients with skewed X chromosome inactivation. *Obstet Gynecol*.
437 2003;101:1236-42.
- 438 21. Lanasa MC, Hogge WA, Kubik C, Blancato J, Hoffman EP. Highly Skewed X-Chromosome
439 Inactivation Is Associated with Idiopathic Recurrent Spontaneous Abortion. *The American Journal of*
440 *Human Genetics*. 1999;65:252-4.

-
- 441 22. Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, *et al.* Single-Cell RNA-Seq
442 Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*.
443 2016;167:285.
- 444 23. Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopina J, Chen PY, *et al.* DNA
445 Demethylation Dynamics in the Human Prenatal Germline. *Cell*. 2015;161:1425-36.
- 446 24. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, *et al.* The Transcriptome and DNA Methylome Landscapes
447 of Human Primordial Germ Cells. *Cell*. 2015;161:1437-52.
- 448 25. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, *et al.* The DNA methylation landscape of human early
449 embryos. *Nature*. 2014;511:606-10.
- 450 26. Hou Y, Fan W, Yan L, Li R, Lian Y, Huang J, *et al.* Genome analyses of single human oocytes. *Cell*.
451 2013;155:1492-506.
- 452 27. Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, *et al.* Chromatin Accessibility Landscape in Human Early
453 Embryos and Its Association with Evolution. *Cell*. 2018;173:248-59.e15.
- 454 28. Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, *et al.* Chromatin analysis in human early development
455 reveals epigenetic transition during ZGA. *Nature*. 2018;557:256-60.
- 456 29. Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, *et al.* Genetic programs in human and mouse early
457 embryos revealed by single-cell RNA sequencing. *Nature*. 2013;500:593-7.
- 458 30. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, *et al.* Single-cell RNA-Seq profiling of human
459 preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol*. 2013;20:1131-9.
- 460 31. Haqq CM, King CY, Ukiyama E, Falsafi S, Haqq TN, Donahoe PK, *et al.* Molecular basis of
461 mammalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY.
462 *Science*. 1994;266:1494-500.
- 463 32. Payer B, Lee JT. X Chromosome Dosage Compensation: How Mammals Keep the Balance. *Annual*
464 *Review of Genetics*. 2008;42:733-72.
- 465 33. Wutz A. Gene silencing in X-chromosome inactivation: advances in understanding facultative
466 heterochromatin formation. *Nat Rev Genet*. 2011;12:542-53.
- 467 34. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI
468 sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X
469 chromosome inactivation. *Am J Hum Genet*. 1992;51:1229-39.
- 470 35. Bertelsen B, Tumer Z, Ravn K. Three new loci for determining x chromosome inactivation patterns.
471 *J Mol Diagn*. 2011;13:537-40.
- 472 36. Fisher EMC, Beer-Romero P, Brown LG, Ridley A, McNeil JA, Lawrence JB, *et al.* Homologous
473 ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible
474 implications for turner syndrome. *Cell*. 1990;63:1205-18.
- 475 37. Zinn AR, Alagappan RK, Brown LG, Wool I, Page DC. Structure and function of ribosomal protein
476 S4 genes on the human and mouse sex chromosomes. *Molecular and Cellular Biology*.
477 1994;14:2485-92.
- 478 38. Zinn AR, Bressler SL, Beer-Romero P, Adler DA, Chapman VM, Page DC, *et al.* Inactivation of the
479 Rps4 gene on the mouse X chromosome. *Genomics*. 1991;11:1097-101.
- 480 39. Watanabe M, Zinn AR, Page DC, Nishimoto T. Functional equivalence of human X- and Y-
481 encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nature Genetics*.
482 1993;4:268.

-
- 483 40. Andrés O, Kellermann T, López-Giráldez F, Rozas J, Domingo-Roura X, Bosch M. RPS4Y gene family
484 evolution in primates. *BMC Evolutionary Biology*. 2008;8:142-.
- 485 41. Rosner A, Paz G, Rinkevich B. Divergent roles of the DEAD-box protein BS-PL10, the urochordate
486 homologue of human DDX3 and DDX3Y proteins, in colony astogeny and ontogeny. *Dev Dyn*.
487 2006;235:1508-21.
- 488 42. Vakilian H, Mirzaei M, Sharifi Tabar M, Pooyan P, Habibi Rezaee L, Parker L, *et al*. DDX3Y, a
489 Male-Specific Region of Y Chromosome Gene, May Modulate Neuronal Differentiation. *Journal of*
490 *Proteome Research*. 2015;14:3474-83.
- 491 43. Cooper DW. Directed genetic change model for X chromosome inactivation in eutherian
492 mammals. *Nature*. 1971;230:292-4.
- 493 44. Huynh KD, Lee JT. Inheritance of a pre-inactivated paternal X chromosome in early mouse
494 embryos. *Nature*. 2003;426:857-62.
- 495 45. Okamoto I, Arnaud D, Le Baccon P, Otte AP, Disteché CM, Avner P, *et al*. Evidence for de novo
496 imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature*.
497 2005;438:369-73.
- 498 46. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X
499 inactivation during early mouse development. *Science*. 2004;303:644-9.
- 500 47. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic Dynamics of Imprinted X
501 Inactivation During Early Mouse Development. *Science*. 2004;303:644.
- 502 48. Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, *et al*. Reactivation of the
503 paternal X chromosome in early mouse embryos. *Science*. 2004;303:666-9.
- 504 49. Dasoula A, Kalantaridou S, Sotiriadis A, Pavlou M, Georgiou I, Paraskevaidis E, *et al*. Skewed
505 X-Chromosome Inactivation in Greek Women with Idiopathic Recurrent Miscarriage. *Fetal Diagnosis*
506 *and Therapy*. 2008;23:198-203.
- 507 50. Tarin JJ, Garcia-Perez MA, Hermenegildo C, Cano A. Changes in sex ratio from fertilization to birth
508 in assisted-reproductive-treatment cycles. *Reprod Biol Endocrinol*. 2014;12:56.
- 509 51. Pollier J, Rombauts S, Goossens A. Analysis of RNA-Seq data with TopHat and Cufflinks for
510 genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol Biol*.
511 2013;1011:305-15.
- 512 52. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
513 *Bioinformatics*. 2009;25:1105-11.
- 514 53. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing
515 data. *Bioinformatics*. 2015;31:166-9.
- 516 54. Shahriyari L. Effect of normalization methods on the performance of supervised learning
517 algorithms applied to HTSeq-FPKM-UQ data sets: 7SK RNA expression as a predictor of survival in
518 patients with colon adenocarcinoma. *Brief Bioinform*. 2017.
- 519 55. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*.
520 2010;11:R106.
- 521 56. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using
522 DAVID bioinformatics resources. *Nature Protocols*. 2008;4:44.
- 523 57. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform.
524 *Bioinformatics*. 2010;26:589-95.

525 58. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
526 population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987-93.

527 59. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, *et al.* dbSNP: the NCBI database
528 of genetic variation. *Nucleic Acids Res.* 2001;29:308-11.

529 60. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from
530 high-throughput sequencing data. *Nucleic Acids Research*. 2010;38:e164-e.

531

532

533 **Figure legend**

534

535 **Figure1. Global transcriptome profiling of male and female embryos reveals**
536 **differences during development.**

537 A) Samples of various development stages included in this study: trophectoderm (TE);
538 epiblast (EPI) and primitive endoderm (PE).

539 B) Table showing the number of male and female cells and embryos used in this
540 study within each embryonic stage. The embryos of 4-cell stage are classified as
541 neither male nor female.

542 C) Two-dimensional t-SNE results of all cells represented by the expression of total
543 genes; Various colors are used to indicate the embryonic day for male (triangle) and
544 female (dot) embryos.

545 D-E) Genome-wide expression per chromosome in the male (light blue) and female
546 (pink) embryos at E3 and E7 stage. Chromosomal RPKM values are calculated as
547 chromosomal reads per kilobase of transcript per million reads mapped (Methods).

548 The significant differences between sexes are defined as $P < 10^{-5}$ (two-sided MWW)
549 and marked with red stars.

550 F) Gene Ontology enrichment results of differentially expressed genes at each stage,
551 representing GO terms for biological processes (red bubble) and molecular function
552 (blue bubble). Summarization of most significant results is listed. x-axis: z-score;
553 y-axis: negative logarithm of the adjusted p-value (provided by DAVID); area of a
554 circle: gene number assigned to the term.

555

556 **Figure 2. Distinct behavior of sex chromosomes during the process of genome**
557 **activation.**

558 A) Heatmap showing the expression of detected Y-linked genes in the early stages. The mean
559 value within each stage is calculated and scaled to z-score range in $[-4,4]$;

560 B) Boxplot of the two most highly expressed genes on the Y chromosome indicates high
561 expression in male cells from the 8-cell stage. The boxes with blue color represent the
562 expression of *RPS4X1* and *DDX3Y*, respectively. The mean values of all other Y-linked genes
563 except these two markers are calculated within each cell and drawn as gray boxes.

564 C) Hierarchical clustering for E3 male (blue) and female (red) embryos using the expression of
565 *RPS4Y1* shows a consistent classification pattern. The embryos starting with "Y" are from Yan
566 et al., and the others are from Petropoulos *et al.*

567 D) Heatmap of the average expression of X-linked genes in males and females during the
568 process of genome activation (from E2 to E4). These genes are sorted by their genomic

569 location and the expression is scaled to z-score. E3_M: male embryos at E3; E3_F: female
570 embryos at E3; E4_M: male embryos at E4; E4_F: female embryos at E4.

571 E) Histogram of two representative genes on the X chromosome with biallelic expression in
572 female embryos after EGA. The exact reads number supporting each allele is marked above
573 each bar, and the heatmap under bars present its expression in each cell (with a range from
574 blue to red to show the increase of expression). The two informative loci are *rs41307260* and
575 *rs5963597*.

576

577 **Figure 3. Dosage compensation of the X chromosome firstly occurring in**
578 **trophectoderm.**

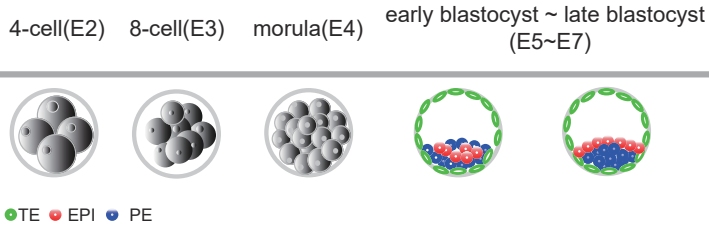
579 A) t-SNE results using the expression of X-linked genes during all stages. Various colors indicate
580 the embryonic day for male (triangle) and female (dot) embryos.

581 B) The two -dimension clustering results of E7 embryos, with lineage assignments of EPI (blue),
582 PE (red) and TE (green) for males (triangle) and females (dot).

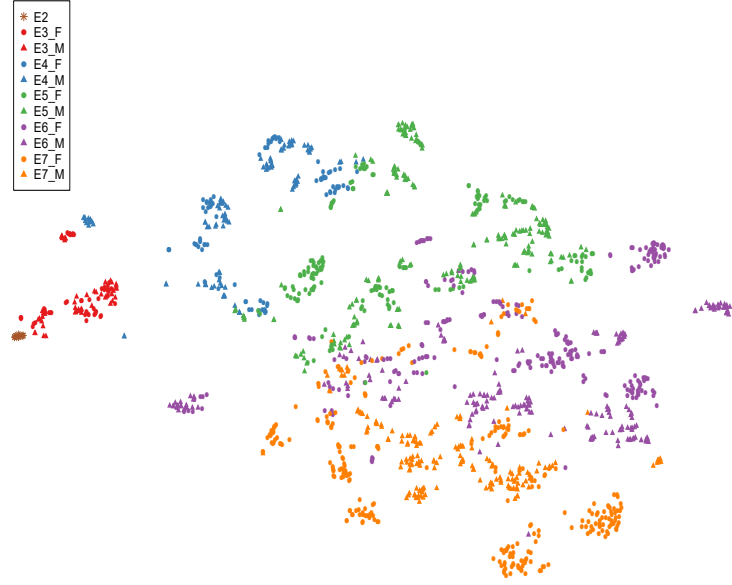
583 C) Boxplots of X chromosome RPKM sum for individual cells from E5 to E7 blastocysts, with
584 lineage of EPI, PE and TE; an example of chr10 is presented here as negative control of
585 autosomes (chrA); p value, two-sided MWW.

586 D) Integrative genome view (IGV) of the methylation level near to the marker region
587 (chrX:48,693,322-48,693,661) within *PCSK1N* of embryos during development. The height of
588 bars shows the percentage of methylation at each loci, ranging from 0% to 100%.

A



C

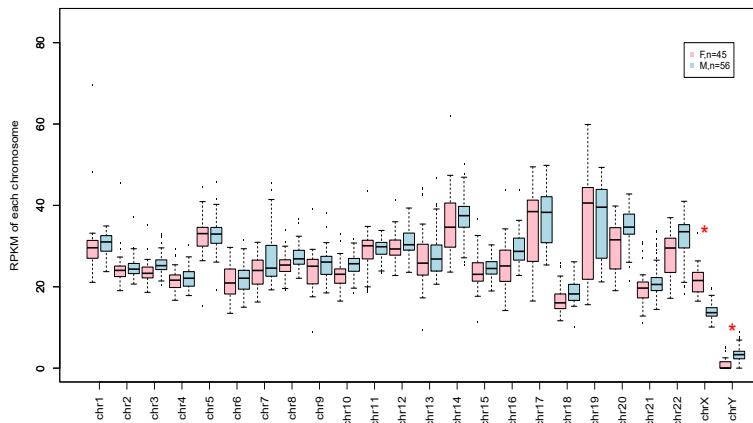


B

Data Stage	Yan <i>et al</i>		Petropoulos <i>et al</i>		Combined	
	Female	Male	Female	Male	Female	Male
E2(4-cell)	12(3)		-	-	12(3)	
E3(8-cell)	12(2)	8(1)	33(6)	48(7)	45(8)	56(8)
E4(Morula)	0	16(2)	92(9)	98(7)	92(9)	114(9)
E5 (Early blastocyst)	-	-	171(10)	206(14)	171(10)	206(14)
E6 (Late blastocyst)	8(1)	22(2)	235(10)	180(8)	243(11)	202(10)
E7 (Late blastocyst)	-	-	290(10)	176(7)	290(10)	176(7)

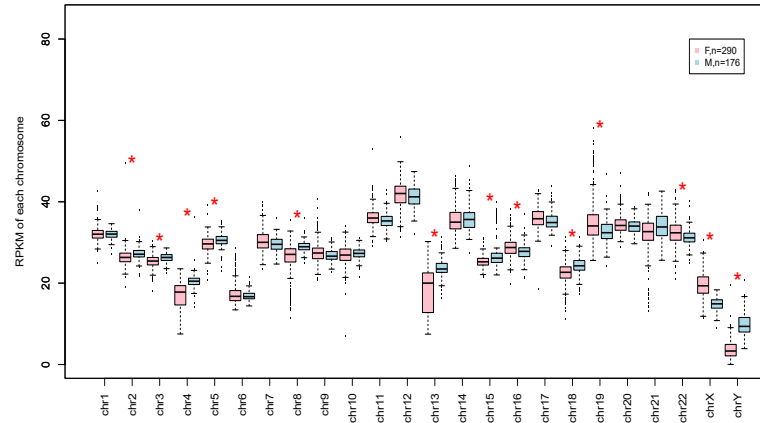
D

Expression of all chromosomes at E3

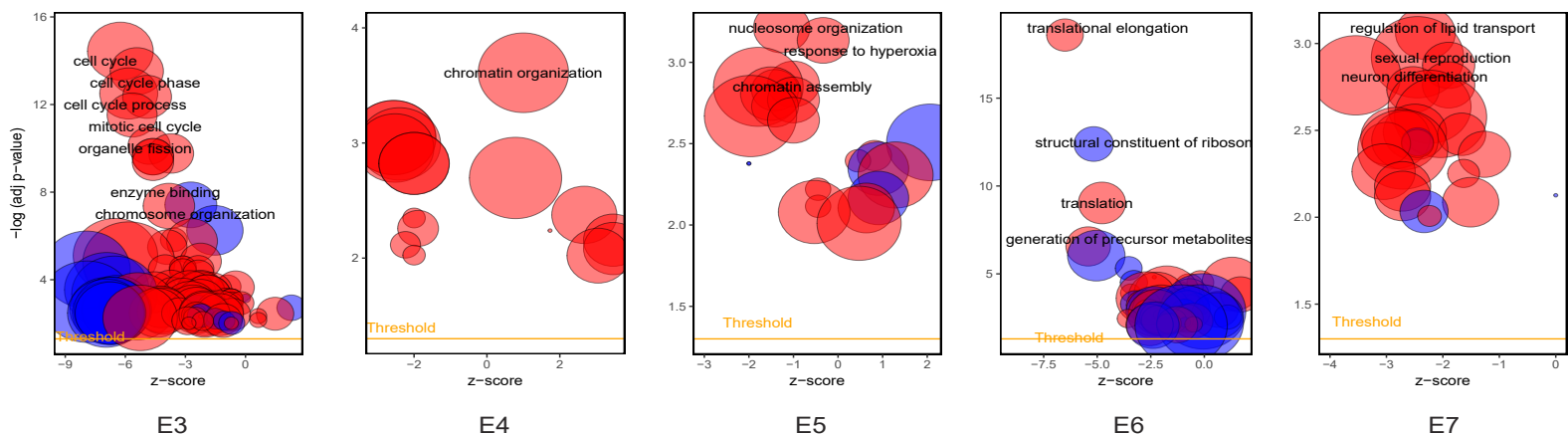


E

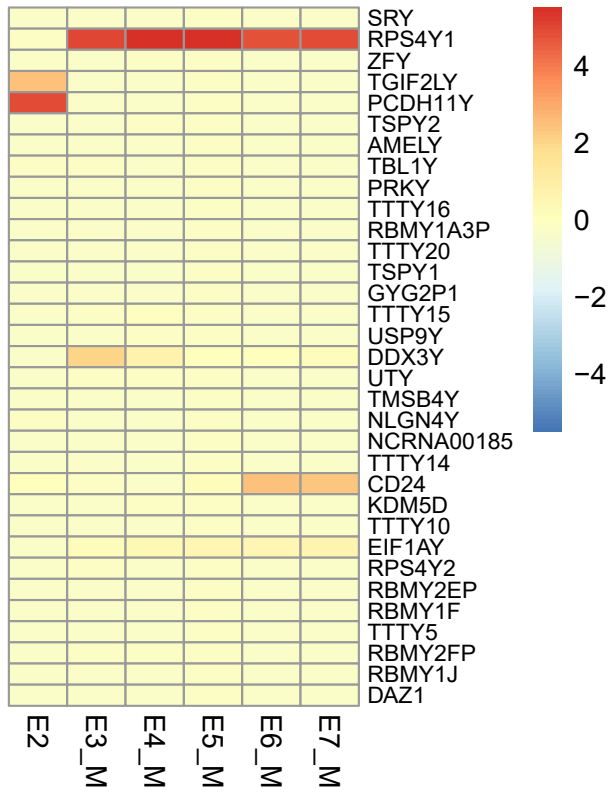
Expression of all chromosomes at E7



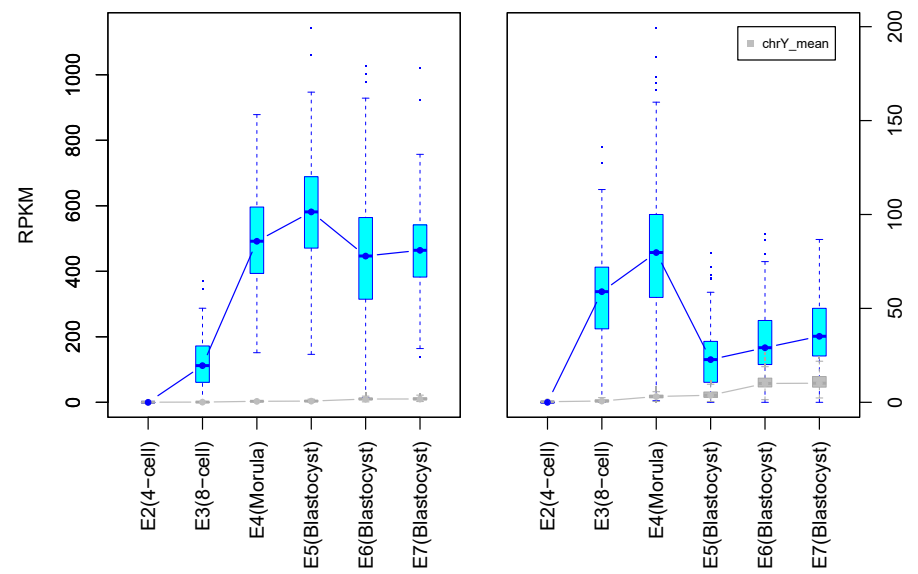
F



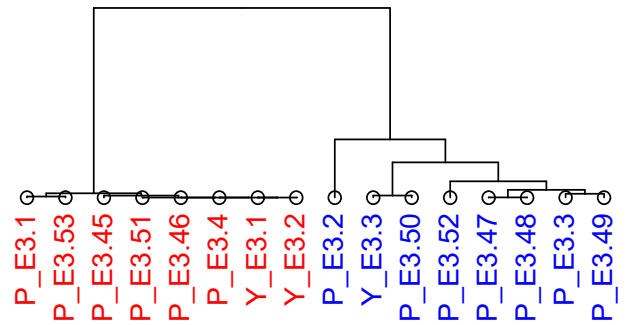
A



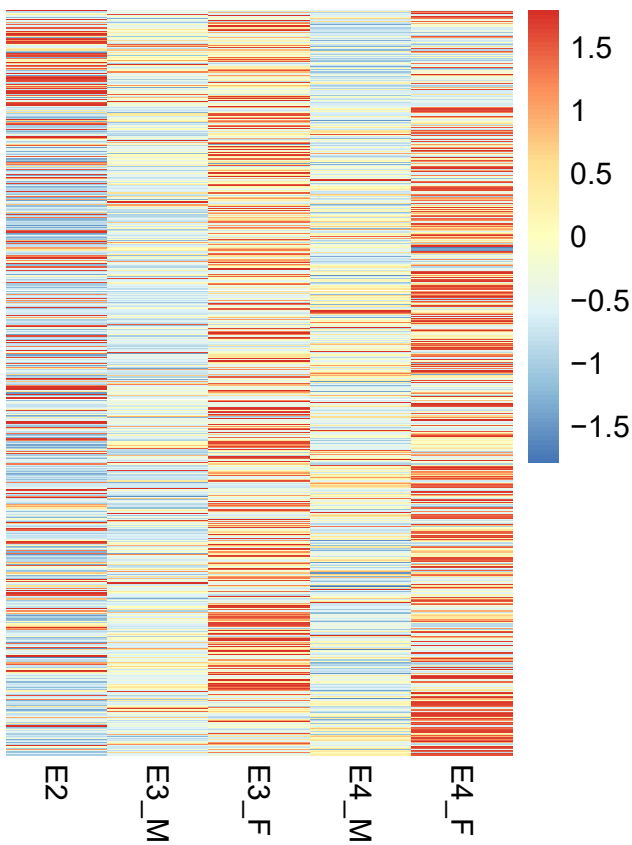
B



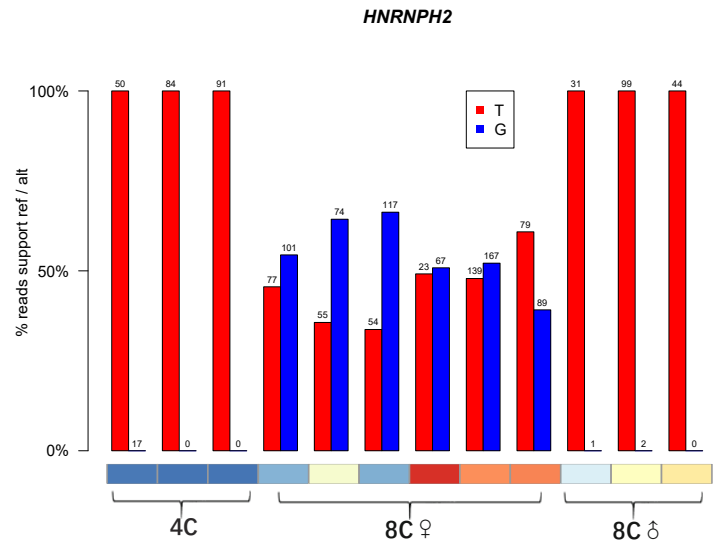
C



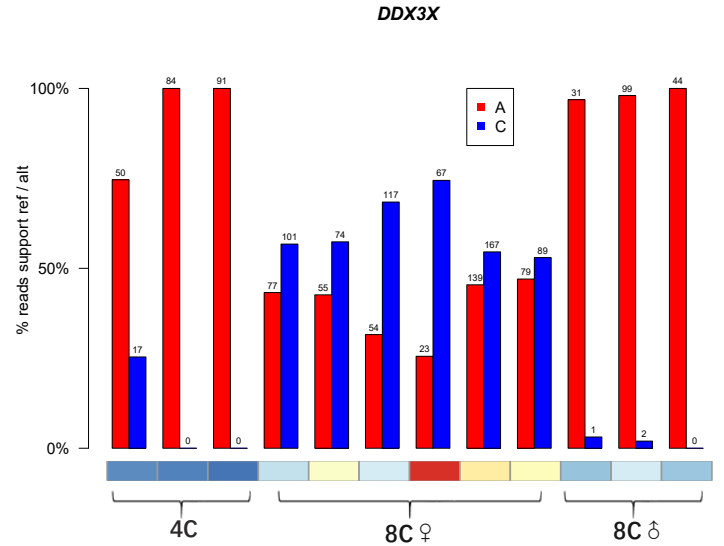
D



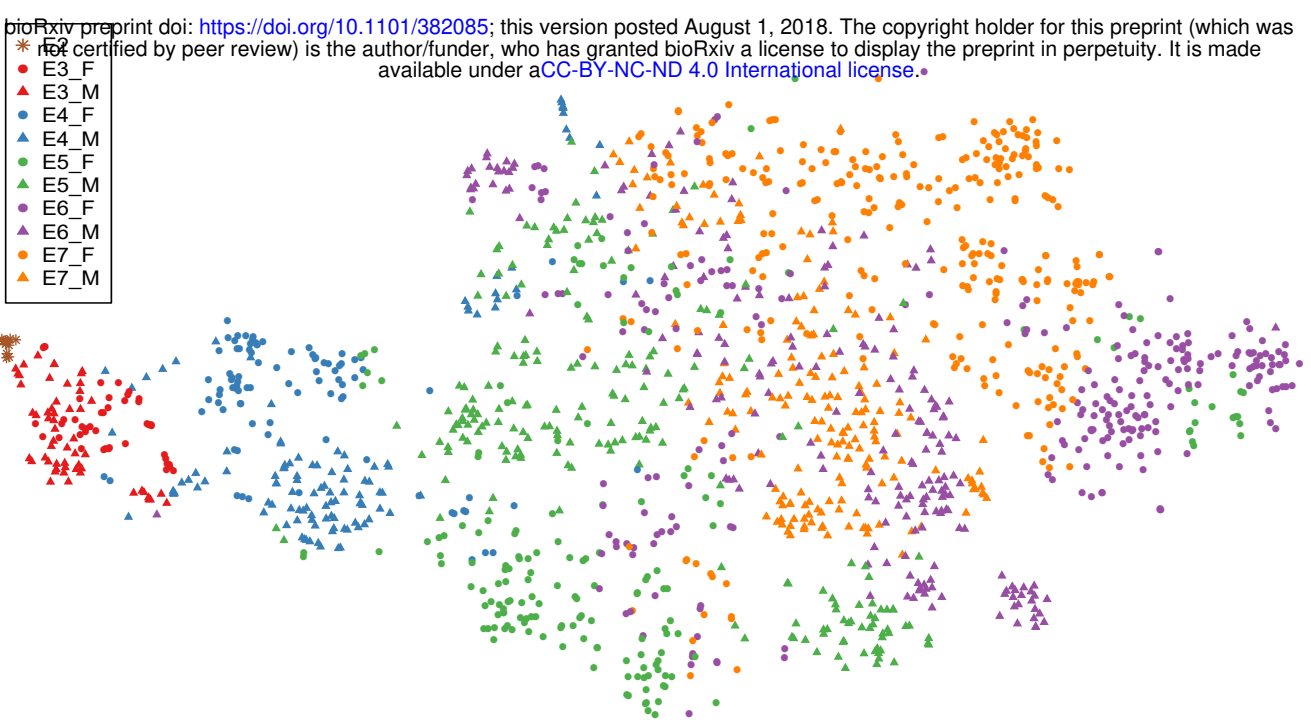
E



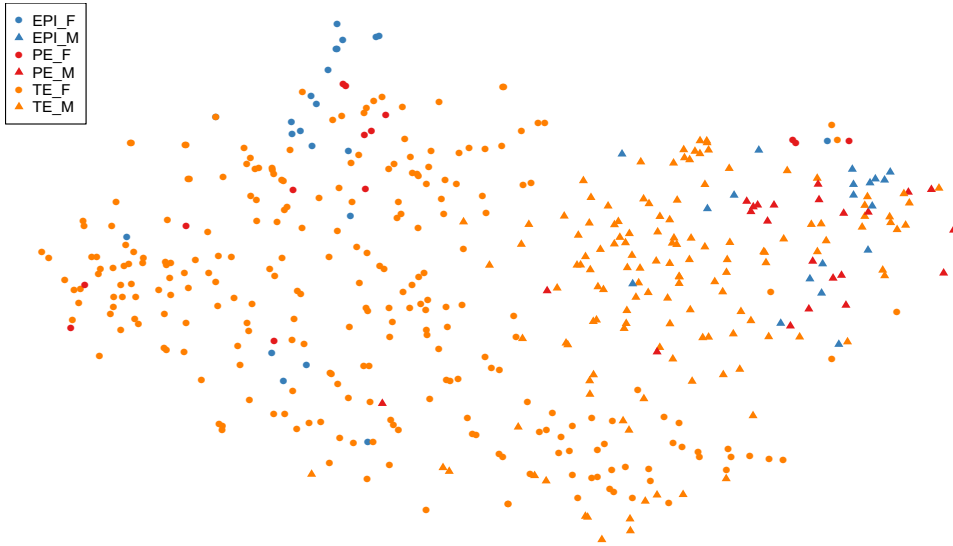
F



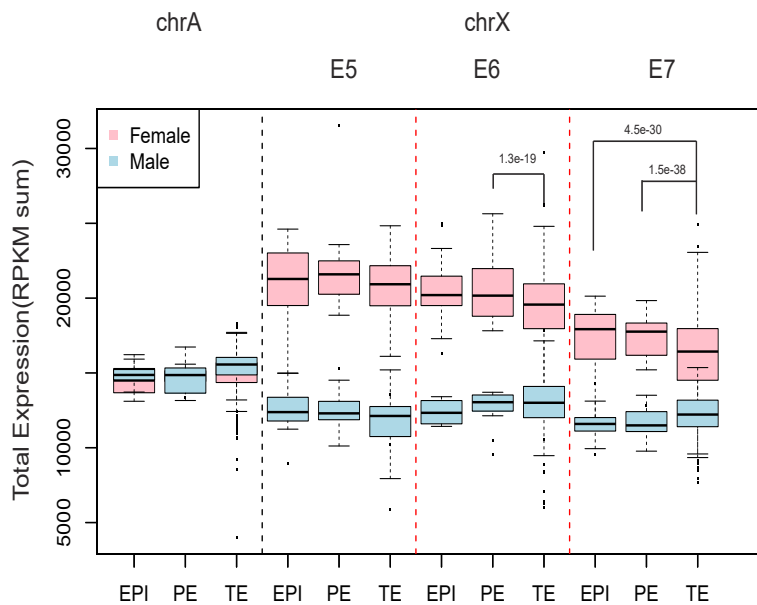
A



B



C



D

