



HAL
open science

Spatio-temporal X-linked gene reactivation and site-specific retention of epigenetic silencing in the mouse germline

Clara Roidor, Laurène Syx, Emmanuelle Beyne, Dina Zielinski, Aurélie Teissandier, Caroline Lee, Marius Walter, Nicolas Servant, Karim Chebli, Déborah Bourc'his, et al.

► **To cite this version:**

Clara Roidor, Laurène Syx, Emmanuelle Beyne, Dina Zielinski, Aurélie Teissandier, et al.. Spatio-temporal X-linked gene reactivation and site-specific retention of epigenetic silencing in the mouse germline. 2023. <hal-04266626>

HAL Id: hal-04266626

<https://hal.science/hal-04266626v1>

Preprint submitted on 2 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1 **Spatio-temporal X-linked gene reactivation and site-specific retention of epigenetic silencing in the**
2 **mouse germline.**

3
4 Clara Roidor^{1*}, Laurène Syx^{2*}, Emmanuelle Beyne¹, Dina Zielinski³, Aurélie Teissandier³, Caroline Lee^{4,5},
5 Marius Walter³, Nicolas Servant², Karim Chebli¹, Déborah Bourc'his³, M. Azim Surani^{4,5,6}, Maud
6 Borensztein^{1,4,5±}

7
8 ¹ IGMM, University of Montpellier, CNRS, Montpellier, France

9 ² Institut Curie, PSL Research University, INSERM U900, Mines ParisTech, Paris, France.

10 ³ Institut Curie, Paris Sciences et Lettres Research University, INSERM U934, CNRS UMR3215, Paris,
11 France.

12 ⁴ Wellcome Trust/Cancer Research UK Gurdon Institute, Henry Wellcome Building of Cancer and
13 Developmental Biology, Cambridge CB2 1QN, UK

14 ⁵ Physiology, Development and Neuroscience Department, University of Cambridge, Cambridge CB2 3EL,
15 UK.

16 ⁶ Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Puddicombe Way,
17 Cambridge Biomedical Campus, Cambridge CB2 0AW, UK.

18

19

20 * These authors contributed equally

21

22 ± Correspondence should be sent to maud.borensztein@igmm.cnrs.fr

23

24

25 Keywords: X-chromosome inactivation, primordial germ cells, reprogramming, scRNA-seq, mouse
26 embryogenesis

27

28

29 **Abstract**

30
31 Random X-chromosome inactivation (XCI) is a hallmark of female mammalian somatic cells. This
32 epigenetic mechanism, mediated by the long non-coding RNA *Xist*, occurs in the epiblast and is stably
33 maintained to ensure proper dosage compensation of X-linked genes during life. However, this silencing is
34 lost during primordial germ cell (PGC) development. Using a combination of single-cell allele-specific
35 RNA sequencing and low-input chromatin profiling in developing *in vivo* PGC, we provide unprecedented
36 detailed maps of gene reactivation. We demonstrated that PGC still carry a fully silent X chromosome on
37 embryonic day (E) 9.5, despite the loss of *Xist* expression. X-linked genes are then gradually reactivated
38 outside the *Xist* first-bound regions. At E12.5, a significant part of the inactive X chromosome (Xi) still
39 resists reactivation, carrying an epigenetic memory of its silencing. Late-reactivated genes are enriched in
40 repressive chromatin marks, including DNA methylation and H3K27me3 marks. Our results define the
41 timing of reactivation of the silent X chromosome a key event in female PGC reprogramming with direct
42 implications for reproduction.

43

44

45 **Introduction**

46

47 In mammals, while proper commitment and homeostasis of somatic lineages are central to individual
48 survival, correct establishment of the germline is crucial for functional gamete and species survival. In
49 mice, specification of germ cell lineage is initiated during embryonic post-implantation development at the
50 onset of gastrulation. Approximately 30-40 PGC become specified and are found at the base of the allantois
51 bud at E7.25^{1,2}. After E8.5, PGC undergo migration and proliferation, and reach the genital ridges between
52 E10 and E11. Throughout this period, germ cell proliferation and colonization continue until the PGC enters
53 meiotic prophase at E13.5 in the female embryos, a process that occurs only after birth in males³.
54 Establishment of the germline determines the production of oocytes and spermatozoa, and therefore their
55 competency to accomplish fertilization and transmit genetic and epigenetic information to the next
56 generation⁴. PGC differentiation is accompanied by repression of the somatic program, expression of
57 germline-specific genes, and extensive genome-wide epigenetic reprogramming including DNA
58 demethylation, loss of genomic imprints, and redistribution of histone marks^{1,5}. Epigenetic reprogramming
59 occurs when PGCs proliferate and migrate to colonize the future gonads. During the upstream development
60 of gonadal sex determination, PGC epigenetic reprogramming displays striking differences between XX
61 females and XY males, with reactivation of the inactive X chromosome⁶⁻⁹. This leads to an excess of X-
62 linked gene products, with the X:Autosome ratio exceeding 1 in female PGCs compared to males^{10,11}.

63 Excess X-linked genes could promote sexual dimorphism and meiosis progression through the direct or
64 indirect involvement of some X-linked genes in the process of sex-specific gonadal formation and Xi
65 reactivation itself^{8,12}.

66

67 Differences in sex chromosome content between males and females lead to gene dosage imbalance. This is
68 compensated in mammals by transcriptional silencing of one of the two X chromosomes in female somatic
69 cells^{13,14}. This epigenetic mechanism, called X-chromosome inactivation, represents an important paradigm
70 for chromosome-wide epigenetic regulation. Long non-coding RNA *Xist* plays a crucial role in the initiation
71 of XCI¹⁵⁻¹⁷. Its absence in early female embryos leads to lethality owing to both impaired dosage
72 compensation and extra-embryonic tissue development^{16,17}. Once expressed from the future Xi, *Xist* coats
73 in cis the most accessible regions, in 3D spatial proximity, the *Xist* ‘entry sites’, before spreading along
74 chromosome¹⁸⁻²⁰. Transposons, particularly LINE1 elements, have been proposed to facilitate this
75 heterochromatinization process^{21,22}. *Xist* then triggers the recruitment of specific factors involved in gene
76 silencing, which in turn induces the removal of active chromatin marks and recruitment of repressive
77 histone marks, such as H3K27me3, H3K9me2, and H2AK119ub²³. Finally, DNA methylation is recruited
78 to the promoters of inactivated genes to further lock the silent state of these genes and maintain them over
79 hundreds of cell divisions²⁴. Studies on the kinetics of XCI have shown that different genes followed
80 different speeds of silencing in the pre-implantation¹⁶ and post-implantation mouse embryos²⁵, and during
81 *in vitro* differentiation of mouse embryonic stem cells (mES)^{26,27}, except for a small subset of genes that
82 resist silencing (escapee, ~7 % in the mouse). Early silent genes seem to be more prone to lie inside *Xist*
83 ‘entry sites’ and close to the X-inactivation centre (Xic)¹⁶ from which *Xist* is transcribed, in gene-rich
84 regions, pre-bound by Polycomb²³ and close to LINE1 elements²⁷.

85

86 Although XCI has been extensively studied over the past 60 years, much less is known about how Xi
87 reactivation occurs and whether it mirrors XCI key events. Upon development, Xi reactivation occurs after
88 imprinted XCI in the inner cell mass (ICM) of the blastocyst^{28,29} (a rodent-specific event) and in PGC^{6,7}
89 after random XCI. It is also observed *in vitro* during female induced pluripotent stem cell derivation (iPSC)
90^{30,31}. Based on a combination of immunofluorescence, RNA-FISH, and RT-PCR in the germline, early PGC
91 carry an inactive X chromosome enriched in H3K27me3^{7,9,32}, with silent X-linked genes⁶. Reactivation is
92 initiated by the transcriptional extinction of *Xist* RNA (from E7.5, prior to and upon PGC migration), loss
93 of repressive H3K27me3 chromatin marks (from E9.5, during PGC migration and proliferation), and re-
94 expression of silenced X-linked genes (from E10.5, when PGC start to colonize the future gonads)^{6,7,32,33}.
95 *Xist* repression has been linked to pluripotency factors such as Nanog and Prdm14 in pluripotent mES³⁴,
96 ICM^{29,35} and germline⁹.

97 However, studies based on single-cell RNA sequencing (scRNA-seq) in post-implantation embryos have
98 shown progressive random XCI depending on lineage^{25,36}. Some epiblast cells can still carry two active X
99 chromosomes at the time of PGC specification. A recent study on Xi reactivation in *in vitro* PGC-like cells
100 suggested that Xi could only be moderately silenced in early PGC before full reactivation⁸. This study was
101 based on *in vitro* XCI in Epiblast-like cells differentiated from mES followed by Xi reactivation in induced
102 PGC-like cells.

103 Despite this knowledge, the events underlying X-linked gene reactivation in the germline remain unclear,
104 particularly at the level of the entire chromosome. The extent of XCI in early *in vivo* PGC and the kinetics
105 of gene reactivation are still unknown. To address these questions, we explored the precise kinetics of X-
106 linked gene expression and Xi chromatin change. We combined interspecific mouse crosses and scRNA-
107 seq, DNA methylation assay using Whole Genome Bisulfite (WGBS), and H3K27me3 histone mark
108 profiling using low-input allele-specific CUT&RUN. We investigated the transcriptional changes from
109 E8.5 to E12.5 PGC. We showed that X-linked genes are sequentially activated, as previously described for
110 the ICM of the blastocyst²⁹ but with different dynamics and requirements. In PGCs, we observed
111 reactivation dependency on *Xist* entry sites, DNA methylation levels, H3K27me3 enrichment, and genomic
112 location. This study provides important insights into the transcriptional, allelic, and chromatin dynamics of
113 Xi reactivation in germline cells. Our novel results emphasize the importance of studying epigenetic
114 reprogramming of the inactive X chromosome in the unique context of the germline, which is the most
115 relevant for developmental syndromes and human reproductive medicine.

116
117

118 **Results**

119

120 **Transcriptional analysis of *in vivo* PGC and soma cells by scRNA-seq**

121 To address the X-chromosome reactivation kinetics in PGC, at the chromosome-wide scale, we produced
122 high-quality, high-coverage scRNA-seq for female and male PGC, as well as surrounding female somatic
123 cells, as a control for the maintenance of X-chromosome inactivation. F1 interspecific embryos were
124 obtained by mating *Mus Musculus domesticus* (129) inbred and *Mus Musculus castaneus* (Cast) inbred
125 mice, and collected every day between E8.5 and E12.5 (**Figure 1A**). These sub-species have evolved for
126 more than 3 million years and carry an important number of well-characterized single nucleotide
127 polymorphisms (SNP) (mean of ~1 SNP / 650 bp on the X chromosome between 129 and Cast, Mouse
128 Genomes project)^{37,38}. To visualize PGC in our embryos, we utilized Green Fluorescent Protein (GFP)
129 transgene, in a 129 genetic background under the control of *Stella* or *Oct4* promoter. *Stella* gene (also
130 known as *Dppa3*) is expressed early during PGC development, at the time of emergence³⁹. We used it for
131 E8.5 embryos. After E8.5, we took advantage of the Oct4-GFP transgenic mice⁴⁰. *Oct4* also known as

132 *Pou5f1*, encodes a pluripotent transcription factor present in PGC, but not in surrounding somatic cells.
133 From E8.5 to E10.5, PGC cells were collected with the assistance of fluorescent active cell sorting based
134 on GFP expression. From E11.5, soma and PGC cells were collected after gonad dissection, based on their
135 size, and confirmed based on the presence of GFP under a microscope. Each single cell was then manually
136 picked, and poly adenylated mRNA was amplified⁴². We produced high-quality scRNA-seq libraries from
137 154 samples. Since we were interested in allelic expression, we decided to use a scRNA-seq method⁴¹
138 allowing high-depth of high-throughput sequencing for each single cell (**Supplemental Table 1**). Only
139 single cells that passed the quality controls (see Methods) were used for downstream analysis (n=137
140 libraries). We used Principal Component Analysis (PCA) to associate single cells based on their lineage
141 (PGC versus soma, **Figure 1B**). The cells were first associated based on their lineage, followed by their
142 embryonic stage. Next, we performed a correlation analysis based on the expression status of pluripotency,
143 soma, and Y-linked genes (**Figures 1B and C**; genes listed in **Figure 1C**). We classified the cells according
144 to their developmental stage and pluripotency/soma factor status. This clearly supports a strong repression
145 of the somatic program in PGC and confirmed the expression of well-known factors in PGC such as Stella,
146 Blimp1, Oct4, Nanog, and Dazl^{1,43}. The weight of these known factors in segregating the soma and PGC
147 lineages is shown in **Extended Figure 1A**. The length of the arrows is proportional to the implication of
148 the factor in PCA association. Without surprise, markers of soma point towards somatic cells and germline
149 towards PGC. With a closer look it is important to note that late PGC genes preferentially point towards
150 E12.5 PGC. We then asked the best 30 predictor genes of PCA clustering (**Extended Figure 1B**). Some
151 well-known factors were found to push towards PGC fate, such as Oct4, Tfap2c, and DND1^{1,44}. However,
152 other genes with interesting roles have also been identified, such as Zing Finger Protein Zfp985. The ZFP
153 family is important for protecting DNA methylation at imprinting loci and transposons⁴⁵. Recently, Zfp982
154 has been associated with the stemness state of mES through its potential control of Nanog and Stella⁴⁶.
155 Epithelial splicing regulatory protein 1, Esrp1, induces oocyte defects and female infertility if deleted from
156 E15⁴⁷. Our dataset could reveal novel PGC markers and are important candidates for PGC biology.

157
158 We then studied the sex of our single-cell samples to focus on sex specificities, X-chromosome reactivation,
159 and transcriptional changes during female PGC development. After clustering PGC and soma cells based
160 on the expression of well-known markers and confirmation by clustering analysis (**Figures 1C and D**), we
161 sexed all single cells based on *Xist* (in XX females) and Y-linked gene (in XY male) expression, allowing
162 attribution to each single cell towards a lineage and sex (**Figure 2A, Supplemental Table 1**). Sex was also
163 confirmed by the absence of SNPs from the X chromosomes in XY cells. By serendipity, an XO female
164 embryo was found at E12.5, and was used in the analysis (n= 3 PGC and 1 soma cells sequenced). We first
165 studied the differentially expressed genes (DEG) in migratory and colonizing female PGC by comparing

166 E9.5 versus E10.5, E11.5 versus E10.5 and E12.5 versus E11.5 (**Extended Figures 1C, 1D and 1E**).
167 Strikingly, very few DEG were found in migratory PGC (E10.5 versus E9.5). This is in accordance with a
168 previously published large scRNA-seq dataset of mouse male and female PGC⁴³. From E11.5, there was a
169 strong increase in DEG in female PGC, with expected upregulation of PGC genes such as *Dazl* and
170 *Stella/Dppa3*. In accordance with previous reports, there was significant gene repression and upregulation
171 at E12.5, compared to E11.5^{43,48}. Most DEG from the X-chromosome were upregulated, which is consistent
172 with X-chromosome reactivation occurring in female PGC (**Extended Figure 1E**).
173 Furthermore, soma and PGC cells strongly clustered on PCA by lineage then developmental stage (1st axis,
174 **Figure 1B**). We then plotted male and female cells on a PCA for each developmental stage (**Extended**
175 **Figure 2A-D**). Clustering by sex was confirmed from E12.5, based on gene expression (PCA, 1 000 most
176 differentially expressed genes) in post-migratory E12.5 PGC, at the onset of sex gonadal differentiation
177 (**Extended Figure 2D**). This was expected based on a previous large-scale scRNA-seq⁴³ and supported the
178 quality of our database for further analysis.

179

180 **X-chromosome reactivation initiates progressively upon PGC development**

181 To study X-chromosome reactivation, we analysed scRNA-seq in an allele-specific manner to determine
182 the parental origin of the transcripts. SNPs from F1 hybrid embryos were used to map informative reads to
183 either 129 or Cast genomes (see Methods). Each gene with informative SNP and expressed more than 2
184 Reads Per Retro-Transcribed length per million mapped reads (RPRT) were provided an allelic expression
185 ratio (reads from Cast divided by total informative reads). The parental origin of Xi was then determined
186 in each single cell based on the allelic ratio of all informative X-linked genes and the Xa allelic ratio
187 calculated as reads mapped on active X (Xa) divided by total reads, in following analysis (**Supplemental**
188 **Table 1**, see Methods). On the 94 female scRNA-seq, 64 cells carried an active X chromosome of Cast
189 origin (68 %) and 30 cells of 129 origin (32 %). Despite random choice of Xi, we observed a strongly
190 skewed silencing towards the 129 chromosome. This confirmed the X-chromosome controlling element
191 (Xce) effect in F1 hybrid female mice. Indeed, it has been well documented that an F1 hybrid background
192 could lead to skewed XCI depending on their Xce strength ($Xce^a < Xce^b < Xce^c < Xce^d < Xce^e$ with the strongest
193 Xce allele being the most resistant to silencing)⁴⁹. Using 129 and Cast strains associated with Xce^a and
194 Xce^c respectively, we confirmed that the 129 chromosome was preferentially chosen to be silent in the
195 expected proportions⁵⁰.

196

197 We then studied allelic expression of autosomal and X-linked genes during PGC development (**Figure 2B**).
198 Autosomal genes were biallelically expressed ($0.2 < \text{allelic ratio} < 0.8$), with parity between 129 and Cast
199 reads, in both males and females E9.5-E12.5 PGC. On the other hand, X-linked genes were strictly

200 expressed from the Xa (allelic ratio > 0.8) in E9.5 female PGC, except for a few genes, presumably the
201 escapees. Distribution of the X-linked gene allelic ratios was very similar in both female and male PGC,
202 highlighting complete XCI in E9.5 PGC. From E10.5, X-chromosome reactivation initiated in female PGC
203 (ratio<0.8) and progressed upon PGC development.

204 Because PGC could be heterogeneous in terms of developmental timing, even inside the same embryo, we
205 decided to order the female PGC cells by pseudotime ordering (**Figure 2C**). We used the 1st principal
206 component of our PCA (**Figures 1B** and **2A**) to order female PGC by pseudotime ordering. Despite a few
207 lagging cells, mainly at E11.5 and E12.5, most of the cells at the same developmental stage were clustered
208 together, without a clear distinction of sex (**Figure 2C** and **Extended Figure 2E**). We then studied the
209 percentage of reactivated X-linked genes in each female PGC (**Figure 2D**). We confirmed that Xi
210 reactivation was progressive, from fully silent E9.5 PGC to highly reactivated E12.5 PGC, following PGC
211 development (**Extended Figure 2E**).

212 Since the loss of *Xist* enrichment on Xi is the earliest known event during reactivation in the germline
213 (based on IF/RNA-FISH)^{6,10}, *Xist* expression levels were extracted in our scRNA-seq (**Extended Figure**
214 **2F**). We confirmed that *Xist* is not expressed in most PGC compared to female somatic cells, which restrains
215 *Xist* expression and inactivates the X chromosome. In migratory E9.5 female PGC, X-chromosome
216 reactivation has been initiated, despite the fact that most X-linked genes subject to XCI are still silenced.

217
218 Finally, to explore which pathways could drive Xi reactivation in the germline, we studied the correlation
219 and anti-correlation between genome-wide gene expression and the percentage of reactivated genes (allelic
220 ratio < 0.8) per female single cells (PGC and soma). Because X-chromosome reactivation occurs
221 concomitantly to PGC development, we found that the best 2 correlated genes were germ cell-specific genes
222 *Ddx4/Vasa* (R=0.78, q < 10⁻¹²) and *Dazl* (R=0.70, q < 10⁻⁷)¹. Gene ontology analysis of the top correlated
223 genes (p < 0.001) revealed that chromatin modifiers involved in DNA methylation, gene silencing, and
224 DNA modifications were overrepresented (**Extended Figure 3A**). This suggested that they may play a role
225 in both PGC and Xi reprogramming.

226
227 **Different genes reactivate at different kinetics along the inactive X chromosome**

228 Next, we investigated the kinetics of X-linked gene reactivation along the entire X chromosome. Heat maps
229 of X-linked gene activity were generated for E9.5 to E12.5 female PGC (**Figure 3**). We focused on well-
230 expressed genes to avoid confounding effects due to PCR bias and molecular loss in scRNA-seq method.
231 Polymorphic genes expressed at RPRT>2 in at least 3 developmental stages were included in the heatmap
232 and ordered by genomic position (**Figure 3 left**). E9.5 PGC displayed 92 % (164 out of 179) of
233 monoallelically expressed genes (allelic ratio ≥ 0.8) and 8 % of biallelically expressed genes (0.2 < allelic

234 ratio < 0.8), 8 of which are well-known escapees, such as *Kdm6a*, *Kdm5c*, *Ddx3x*, *Eif2s3x*, *Utp14a*, *Zrsr2*,
235 *181003007Rik*, *Pqbp1*⁵¹. Interestingly, the proportion of biallelically expressed genes increased through
236 the development of female PGC to reach 59 % (112 out of 189) at E12.5 PGC. This indicated the strong
237 reactivation of X-linked genes during PGC development. However, an important portion of the X-linked
238 genes was silenced at E12.5.

239 We classified the genes into different classes with respect to the timing of reactivation (**Figure 4A** and
240 **Method, Figure 3 right**). At E9.5, all genes were silenced, except for the escapee class (n=8 out of 198
241 genes). Early genes (n=29 out of 198) were reactivated from E10.5, intermediate genes (n=55 out of 198
242 genes) from E11.5, and late genes (19 out of 198) from E12.5. At E12.5, 76 out of 198 genes were still
243 silenced (monoallelically expressed from the Xa) and belonged to the very late-reactivated class.

244 Differences in reactivation kinetics were not explained by different expression levels (**Figure 4B**). Early
245 and escapee genes tended to be more highly expressed in E10.5 and E11.5, compared to still-silent genes.
246 This could be explained by the fact that the Xi allele was also transcribed for early reactivated and escapee
247 genes compared to the other classes of genes. Consistently, at E12.5, very late reactivated genes tended to
248 be less expressed than those in the other classes.

249 A closer examination of the reactivation heatmap, ordered by genomic position, showed several regions of
250 reactivation along the entire X chromosome. Because close genomic proximity to escapees could favour
251 early reactivation in mouse iPSC³¹, we tested the distance of our different reactivation class genes to the
252 closest escapee (**Extended Figure 3B**). No link was detected between the differential reactivation kinetics
253 and the distance from escapees in our *in vivo* female PGC. We then tested whether distance to *Xist* locus
254 could be a significant parameter for reactivation kinetics (**Extended Figure 3C**). We found a strong bias
255 for very late-reactivated genes to be localised closer to the *Xist* genomic locus compared to intermediate
256 (p=0.02, KW test) and early (p=0.004, KW test) reactivated genes. Escapee were found further to the *Xist*
257 locus (p=0.0045, KW test). We concluded that there was a strong correlation between close proximity to
258 *Xist* locus and longer germline silencing.

259

260 **Context matters for the kinetics of X-linked gene reactivation**

261 We aimed to test the consistency of the reactivation kinetics of X-linked genes in different developmental
262 contexts (ICM vs. PGC). We compared the classes of reactivation that belong to common X-linked genes
263 during imprinted Xi reactivation in ICM and random Xi reactivation in PGC (**Figure 4C**)²⁹. Few
264 similarities were found, except for escapees in PGC, who also escaped imprinted XCI. We then compared
265 the reactivation classes in PGC to a recent study on PGC-like cells (PGCLC) (**Figure 4D**)⁸. The *in vitro*
266 PGCLC model recapitulates early PGC specification, including Xi reactivation in the female cells.
267 However, very few X-linked genes exhibited similar kinetics. *In vitro*, the cells underwent partial XCI;

268 consequently, X-linked genes could be more prone to early reactivation. Late and very late reactivated
269 genes in PGCLC could be genes that were properly silenced *in vitro* before undergoing reactivation. Thus,
270 these genes had similar reactivation kinetics in our PGC *in vivo*. These data led us to hypothesize that late
271 and very late reactivated genes in PGCLC, similar to PGC, could have been the first genes to be inactivated.

272

273 **Late-reactivated genes lay into *Xist* entry sites**

274 To test the correlation between early XCI and late X reactivation, we compared the kinetics of reactivation
275 in PGC to silencing in differentiating mES cells (**Figure 4E**)²⁶. We found that the early silenced genes
276 belonged to the very late reactivation class in PGC. Being an early silent gene can influence the speed of
277 reactivation a few days later.

278 Because the genes that resisted reactivation at E12.5, are closer to *Xist* locus (**Extended Figure 3C**) and
279 could be the first silenced genes during random XCI (**Figure 4E**), this prompted us to question the
280 relationship between X-linked gene reactivation and *Xist* entry sites. *Xist* entry sites are genomic regions of
281 the chromosome that are the first bound by *Xist* RNA upon initiation of XCI^{18,20}. It is believed that *Xist*
282 exploits the 3D conformation of the chromosome to first bind the regions in 3D spatial proximity to its
283 transcription site and then initiate silencing before spreading across the entire chromosome^{20,52}.
284 Furthermore, we have previously shown that the genes lying inside these 3D accessible regions were more
285 prone to early silencing *in vivo*¹⁶. (**Figure 4F**). Here, we found that X-linked genes located within the *Xist*
286 entry regions (Transcription Start Site TSS inside the predicted regions) showed more resistance to early
287 reactivation than other genes. Genes outside the *Xist* entry sites showed the earliest reactivation and
288 strongest allelic expression from Xi upon PGC development. Thus, the genes inside the *Xist* entry sites may
289 correspond to genes that are more resistant to reactivation, carrying a stronger epigenetic memory of their
290 silencing *in vivo*.

291

292 **Resistance to reactivation could be partially explained by enrichment in chromatin repressive marks**

293 Very late-reactivated genes appeared to correlate with the first silenced regions of the Xi. Because repeated
294 sequences of the genome, mainly LINE-1 elements, have been proposed to help silence propagation and
295 facultative heterochromatinization of the Xi, we tested the enrichment of transposons in and outside *Xist*
296 entry sites (**Figure 5A**). We confirmed that the X chromosome is highly enriched in LINE-1 compared to
297 control autosomal regions. However, we did not observe any significant enrichment of repeats inside *Xist*
298 entry sites compared with the rest of the X chromosome, except for a slight enrichment of short interspersed
299 nuclear elements (SINEs).

300

301 Different chromatin environments could explain the differential kinetics of reactivation of X-linked genes
302 in the PGC, as previously demonstrated in the ICM of blastocysts ²⁹. Both repressive histone marks, such
303 as H3K27me3 and DNA methylation, are enriched on the Xi upon random XCI ¹⁴. We studied DNA
304 methylation by whole-genome bisulfite sequencing (WGBS) in female E6.5 epiblasts, when XCI takes
305 place, and in publicly available datasets of female E10.5 ⁵³ and E12.5 ⁵⁴ PGC (**Figure 5B**). However,
306 because XCI is random, the population of PGC was heterogeneous in terms of parental origin of the Xi (*i.e.*
307 50 % of the cells silence the paternal X chromosome and 50 % the maternal chromosome), and cell
308 population-based assays on mosaic female embryos are not informative for deciphering between Xa and
309 Xi. Thus, we considered for the following analysis that a significant enrichment of DNA methylation at X-
310 linked gene promoters had a higher probability of coming from the Xi rather than the Xa. Very late
311 reactivated genes showed slight enrichment in DNA methylation at their TSS compared to early-reactivated
312 genes at E6.5. From E10.5, the DNA methylation erasure was nearly complete, and differences seemed to
313 be lost (**Figure 5B**). Although very late reactivated genes were initially more enriched in DNA methylation
314 than the early reactivated ones, it is not clear why these genes still resist reactivation at E12.5, when DNA
315 demethylation is complete.

316
317 We then decided to study the repressive histone mark H3K27me3, which is enriched on the inactive X
318 chromosome and confers resistance to early reactivation of some X-linked genes in the ICM ²⁹. To
319 overcome the mosaicism in the PGC population and decipher between Xa and Xi, we used triple transgenic
320 female mice (*Xist^{lox/lox}; Zp3-Cre; Oct4(APE)eGFP*, on a C57B16/J genetic background) crossed with Cast
321 males (**Figure 5C**). This allowed us to collect polymorphic female embryos and sort pure PGC populations
322 based on GFP (under the promoter control of the pluripotency factor OCT4) with non-random X
323 inactivation. The *Xist* deletion occurred in the maternal germline and allowed the transmission of a *Xist*^{KO}
324 B6 allele, which cannot be inactivated; Xi being always the Cast allele. Low-input allele-specific
325 CUT&RUN against H3K27me3 marks was performed on sorted GFP+ female PGC at E11.5 and E12.5
326 (**Figure 5C**). Statistically enriched broad domains of H3K27me3 were identified on X chromosomes
327 (**Figure 5D**). We found a higher number of peaks at E11.5 on the X chromosome than at E12.5, in which
328 most of the peaks were conserved from E11.5. The reads were then mapped to the parental genomes to
329 determine the parental origin of the reads. As expected, most peaks were found on the inactive X
330 chromosome at both E11.5 and E12.5 (**Figure 5D and Extended Figure 4**). We then intersected H3K27m3
331 enrichment with reactivation classes. Enrichment in K27me3 was mainly found at E11.5, (20 % of the late-
332 reactivated and 45 % of very late-reactivated genes). At E12.5, H3K27me3 enrichment was lost in late
333 genes when they were transcriptionally reactivated (**Figure 5E**). The enrichment level is also more
334 important in very late reactivated genes, with most of the signal coming from the inactive Cast chromosome

335 **(Figures 5F-G).** In contrast, early-reactivated genes were depleted in H3K27me₃, in accordance with their
336 biallelic expression status. Tracks showing global and allele-specific H3K27me₃ enrichment were
337 produced for the early reactivated gene *Med14*, late *Dlg3* and *Pjal* genes, and very late *Med12* (**Figure**
338 **5H**). H3K27me₃ enrichment was not detected at *Med14* gene location. Both *Dlg3* and *Pjal* were
339 statistically enriched at E11.5, but not at E12.5, once they were reactivated. Finally, *Med12* gene was
340 strongly enriched at both E11.5 and E12.5, and showed no reactivation at E12.5, based on scRNA-seq. We
341 observed that some late reactivated genes could carry an epigenetic memory of their silencing (H3K27me₃),
342 which was lost concomitantly to transcriptional reactivation.

343

344

345 **Discussion**

346

347 In this study, we performed a comprehensive allele-specific single-cell transcriptomic analysis of migratory
348 and colonizing female PGC, combined with low-input epigenomics. We demonstrated that female early
349 PGC carry a fully inactive X chromosome, which undergoes progressive reactivation in parallel with PGC
350 development. Although it was previously established that the inactive X chromosome experiences
351 transcriptional reactivation in PGC, our study provides the first detailed map of X-chromosome activities
352 *in vivo*. We showed that different genes followed different kinetics of reactivation along the chromosome,
353 with early versus late reactivated genes. We provide evidence for the involvement of genomic location, 3D
354 spatial proximity to the *Xist* locus, and H3K27me₃ chromatin modification in resistance to early
355 reactivation. Together, these investigations open a way for a better understanding of the *in vivo*
356 requirements for female epigenetic reprogramming in general, stem cell biology, and reproduction.

357

358 X-chromosome reactivation occurs during the reprogramming of female primordial germ cells.
359 Consequently, this leads to an excess of X-linked gene products in female PGCs compared to males¹⁰. In
360 mice, this happens transiently at the onset of sex-specific gonadal differentiation (E9.5-<E15.5) in the
361 female germline and could be crucial for normal gonadal development and meiosis, as well as for sex-
362 specific reprogramming. This could promote sexual dimorphism. Patients with sex-chromosome
363 aneuploidy, such as Turner 45,XO, and Klinefelter 47, XXY syndromes, often present infertility and
364 hypogonadism. Maternally inherited sex-chromosome aneuploidy could arise from the presence of a non-
365 reactivated X, potentially detrimental to homologous chromosome pairing and segregation in meiosis¹².

366 Our transcriptomic analysis of XX female, XO female, and XY male PGC highlighted differentially
367 expressed genes, which could be involved in germline formation and/or sex-specific differences. The X
368 chromosome is enriched in factors required for oogenesis (*e.g.* *Fmr1*, *Zfx*) and chromatin modifications and

369 transcription (e.g., histone demethylases Kdm6a, Kdm5a, mediator complex Med14). Importantly, we
370 showed in this study that all these factors are either not subject to XCI or are early-reactivated in female
371 PGC. *Med14* codes for a co-unit of Mediator complex, involved in transcription regulation, and is prone to
372 early reactivation in both mouse iPS³¹ and human breast cells depleted for XIST⁵⁵. *Smc1a* gene has recently
373 been shown to be involved in the remodelling and reactivation of Xi in mouse iPSC⁵⁶. *Kif4* knock-down in
374 oocytes is detrimental to meiosis⁵⁷. *Zfx* is a well-known dose sensitive gene. Its absence in mouse leads to
375 infertility owing to the reduced number of germ cells⁵⁸. Together, this raises the importance of studying
376 these genes with dosage imbalance in PGC. An appropriate dosage of X-linked genes, whose functions are
377 linked to chromatin processes, transcription, and gametogenesis, could be important for female
378 gametogenesis and X-chromosome reactivation.

379
380 Our detailed mapping of X-linked gene reactivation kinetics highlights differential behaviours along the
381 entire X chromosome. The differential kinetics of reactivation are dependent on the developmental context
382^{11,29,31}. Understanding the resistance to early reactivation and the underlying mechanism is important for
383 understanding epigenetic reprogramming on the X- and genome-wide levels. Surprisingly, we showed that
384 40 % of the Xi remained silent at E12.5. These genes lay into regions in close 3D proximity to *Xist* locus
385 (*Xist* entry sites) and could be the first genes to be silenced upon XCI. We believe that these first-silenced
386 genes could be the first targets of *Xist* because of their 3D accessibility⁵². They would become less accessible
387 to the transcriptional machinery and are more enriched in repressive marks, carrying an epigenetic memory
388 of their silencing. In support of this hypothesis, our results indicated that the latest reactivated genes at
389 E12.5 are still enriched in H3K27me3 on their silent allele. These repressive marks are lost concomitantly
390 with gene reactivation.

391 In conclusion, in PGC, we observed a reactivation dependency on *Xist* RNA loss, DNA methylation level,
392 enrichment in transposable elements, and the proximity of X-linked genes to the first regions coated by *Xist*
393 RNA. Together, these investigations open a way for a better understanding of the in vivo requirements for
394 female epigenetic reprogramming in general, stem cell biology, and reproduction.

395

396

397 **Methods**

398

399 Mouse husbandry

400 The care and use of animals are strictly applying European and National Regulation for the Protection of
401 Vertebrate Animals used for Experimental and other Scientific Purposes (Directive 2010/63/EU and French
402 decree R.214-103). All husbandry and experiments involving mouse scRNA-seq were authorised by the

403 UK Home Office Project Licenses PPL80/2637 and PE596D1FE and were carried out in a Home Office
404 designated facility (Wellcome Trust Cancer Research Gurdon Institute). Chromatin experiments were
405 authorized by the French ethics committee number 36 under agreement F3417216 and carried out in the
406 pathogen-free Animal Care Facility of IGMM (facility licence #G34-172-16). Researchers carrying out
407 regulated procedures on living mice held a personal licence from either the UK (C.L., M.B., and M.A.S.)
408 or France (C.R., M.B., K.C., and D.B.).

409 Mice were housed under a 12h light/12h dark cycle at 22 ± 2 °C ambient temperature, with free access to
410 food and water. All embryos were derived from natural mating. Noon on the day of observation of the
411 vaginal plugs was scored as embryonic day (E) 0.5. Embryos were harvested every 24 h between E8.5 and
412 E12.5. Collected embryos were included in the analyses only if they showed normal morphology according
413 to their developmental stages. No statistical method was used to determine the sample size.

414 Male and female hybrid embryos were obtained by breeding *Mus musculus domesticus* 129S1/SvImJ *Stella-*
415 *eGFP* transgenic line³⁹ (at E8.5) or *Mus musculus domesticus* 129S1/SvImJ *GOF-ΔPE-18* transgenic line⁴⁰
416 (from E9.5 onwards) with *Mus musculus castaneus* (CAST) (**Figure 1A**). *Xist*^{-/+} *Zp3*-CRE; *Oct4*-eGFP
417 female embryos (**Figure 5C**) were obtained by mating *Mus musculus domesticus* C57Bl6/J *Xist*^{fllox/fllox}; *Zp3*-
418 CRE; *Oct4*-eGFP females^{17,59,60} with *Mus Musculus Castaneus* males.

419

420 Sexing of the embryos

421 The sex of the embryos was characterized based on the morphology of the gonads from E11.5. Before
422 E11.5, sex was characterized in single-cell RNA-seq datasets by studying the expression of *Xist* and Y-
423 linked genes, as well as the presence or absence of polymorphisms (SNPs) on the X chromosome. For the
424 WGBS, and CUT&RUN experiments, the sex of the embryos was determined by PCR using genomic DNA
425 and *Ube1* primers (*Ube1*-Forward *TGGATGGTGTGGCCAATG*; *Ube1*-Reverse
426 *CACCTGCACGTTGCCCTT*).

427

428 Collection of PGC

429 After dissection of embryos at the location of the PGC, according to embryonic stage, and sexing of the
430 embryos, samples were resuspended in 200 μL of 0.25 % trypsin and incubated at 37 °C for 3 min. Trypsin
431 was inactivated with serum and a single-cell solution was obtained by vigorous up-and-down. For scRNA-
432 seq experiments from E10.5, cells were manually picked based on their GFP and size and washed in PBS-
433 acetylated BSA (**Figure 1A**). For the other stages and CUT&RUN experiments, cells were collected by
434 fluorescence-associated cell sorting (FACS ARIA© and S3e Cell Sorter Bio-Rad©) and processed quickly
435 for scRNA-seq or low-input CUT&RUN.

436

437 Single cell RNA sequencing and bioinformatic analysis

438 Single PGC were washed thrice with PBS/acetylated BSA (Sigma) before being manually transferred
439 within the minimum amount of liquid into PCR tubes. We either directly prepared the cDNA amplifications
440 or kept the single cells at -80°C (less than 2 months) for future preparation. Poly(A)⁺ mRNA extracted
441 from each single cell was reverse-transcribed from 3'-UTRs and amplified according to a previously
442 described protocol^{42,61}. Care was taken to process only embryos and PGC of the highest quality based on
443 morphology and amplification yield. A total of 140 single cells were processed and quality control (QC)
444 was performed as previously described in ¹⁶.

445 Single-cell libraries were prepared from 137 samples that passed QC, according to the manufacturer's
446 protocol (Illumina). Sequencing to produce single-end 50-bp reads was then performed on an Illumina
447 HiSeq 4000 instrument (**Supplementary Table 1**).

448 Quality controls, filtering of raw data, mapping, and SNP calling have been described previously ^{8, 9}.
449 Briefly, the mouse mm10 genome was downloaded from Sanger database. To study allele-specific gene
450 expression, reads were processed according to Borensztein et al¹⁶. SNPs between the 129 and Cast strains
451 were extracted from the VCF file and used to reconstruct the Cast genome. After the removal of the common
452 exonic SNPs between *Xist* and *Tsix*, 20,220,776 SNPs were retained. The number of paternal and maternal
453 reads were counted at each SNP position. The threshold used to call a gene informative was five reads
454 mapped per single SNP, with a minimum of eight reads mapped on SNPs per gene, to minimize disparity
455 with low-polymorphic genes. The allele-specific origin of the transcripts (allelic ratio) was calculated as
456 the total number of reads mapped to the Cast genome divided by the total number of reads for each gene:
457 allelic ratio = Cast reads/(Cast + 129) reads. For X-linked gene, we modified the allelic ratio to: allelic ratio
458 = Xa reads/(Xa + Xi) reads. In the case of a 129 Xi, the allelic ratio became 1-[Cast reads/(Cast+129) reads].
459 Genes were thus classified into two categories: inactivated genes: allelic-ratio value ≤ 0.20 or ≥ 0.80 , and
460 biallelically expressed genes: allelic-ratio value > 0.20 or < 0.80 .

461
462 *Estimation of gene expression levels.* Given that our RNA reverse transcription allowed sequencing only
463 up to an average of 3 kb from the 3' UTR, half of the expressed genes were only partially covered (less than
464 50% of the gene size on average). To estimate transcript abundance, read counts were normalized based on
465 the amplification size of each transcript (RPRT) rather than the size of each gene (RPKM) (see details in
466 Borensztein et al.⁴²).

467
468 *Principal component analysis, hierarchical clustering, and differentially expressed genes in volcano plots.*
469 Only genes with an RPRT value > 1 in at least 25% of the single cells of at least one developmental stage
470 (with a minimum of two cells) were retained for downstream analysis, as previously described in ¹⁶. With

471 the Benjamini–Hochberg correction, genes with an adjusted P value lower than $\alpha = 0.05$ were called as
472 differentially expressed.

473
474 *Heatmap generation for X-chromosome allelic gene expression.* For allelic ratio heatmaps, data from
475 informative genes were analysed at each developmental stage only if the gene was expressed (RPRT >2) in
476 at least 25% of the single blastomeres (with a minimum of two cells) (**Figure 3**). To follow the kinetics of
477 expression, we focused only on genes expressed in at least three different stages. The mean allelic ratio of
478 each gene is represented for the different stages of the female PGC.

479
480 *Global gene expression correlation with X-chromosome reactivation.* Correlation and anti-correlation
481 between gene expression levels (autosomes and X chromosomes) and the percentage of X-linked gene
482 reactivation (allelic ratio <0.8 for X-linked genes) were measured using Pearson’s correlation and
483 Benjamini–Hochberg correction. Gene ontology analysis was performed for the top-correlated genes (q-
484 value <0.05). Genes with RPRT < 2 were considered to be unexpressed (RPRT = 0 and allelic ratio = NA).

485
486 *Definition of X-linked gene reactivation classes.* We automatically assigned X-linked genes to the
487 reactivation classes.

- 488 - Early reactivation: expressed on both chromosomes at stage E10.5. Allelic ratio ≥ 0.8 or NA at E9.5;
489 allelic ratio < 0.8 at E10.5; allelic ratio < 0.8 at E11.5; allelic ratio < 0.8 or NA at E12.5.
- 490 - Intermediate reactivation: expressed on both chromosomes at stage E11.5. Allelic ratio ≥ 0.8 or NA
491 at E9.5; allelic ratio ≥ 0.8 or NA at E10.5; allelic ratio < 0.8 at E11.5; allelic ratio < 0.8 or NA at
492 E12.5.
- 493 - Late reactivation: expressed on both chromosomes at stage E12.5. Allelic ratio ≥ 0.8 or NA at E10.5;
494 allelic ratio ≥ 0.8 at E11.5; allelic ratio < 0.8 at E12.5.
- 495 - Very Late reactivation: silenced at stage E12.5. Allelic ratio ≥ 0.8 at E10.5 Allelic ratio ≥ 0.8 at
496 E11.5 and E12.5 OR allelic ratio < 0.8 at E11.5 and ≥ 0.8 at E12.5.
- 497 - Escapees: always biallelic. Allelic ratio < 0.8 at all stages (**Figure 4A**).

498
499 Low-input CUT&RUN

500 After PGC were collected, the cells were directly pelleted at 4 °C for 5 min. The CUT&RUN protocol was
501 modified from Skene *et al*^{64,65} and Dura *et al*⁶⁶ to accommodate a low number of PGC (5 000- 15 000 per
502 sample, **Supplementary Table 1**). Briefly, cells were split according to the number of required antibody
503 profiles, and Nuclear Extraction Buffer (20 mM HEPES-KOH, 10 mM KCl, 0.5 mM spermidine, 0.1 %
504 Triton X-100, 20 % Glycerol, Complete EDTA-free protease inhibitor cocktail) was gently added to the

505 cell solution and incubated on ice for 5 min. Cells and concanavalin A beads were incubated for 10 min at
506 room temperature (RT) on a rotating wheel. Cells were then collected on magnets, resuspended in Blocking
507 Buffer (20 mM HEPES-KOH, 150 mM NaCl, 0.5 mM spermidine, 0.1 % BSA), 2 mM EDTA, and 1×
508 Complete EDTA-free protease inhibitor cocktail), and incubated at RT for 5 min. Cells were then washed
509 and incubated with H3K27me3 antibody (1:200 dilution, Cell signalling 36B11#9733, control with IgG
510 rabbit Sigma) for 2h30 at 4 °C on a rotating wheel and then washed twice. Samples were then incubated
511 with 1:400 Protein A-MNase fusion protein (gift from the Dominique Helmlinger lab, CRBM France) for
512 1 h at 4 °C followed by two washes. Cells were then resuspended in 150 µL Wash Buffer and cooled in an
513 ice-water bath for 5 min before the addition of a final concentration of 100 mM CaCl₂. Targeted digestion
514 was performed for 30 min on ice. The samples were then incubated for 20 min at 37 °C to release the cleaved
515 chromatin fragments. After centrifugation at 16,000 × g for 5 min, supernatants were transferred to new
516 low-binding tubes. Following the addition of 20 % SDS and 20 mg ml⁻¹ Proteinase K, the samples were
517 incubated 30 min at 70 °C. DNA was purified using phenol/chloroform, followed by chloroform extraction
518 and precipitation with 20 mg ml⁻¹ glycogen and three volumes of 100 % ethanol at 20 °C. The DNA pellet
519 was washed with 85% ethanol, centrifuged, and resuspended in low Tris-EDTA.
520 Library preparation was performed according to manufacturer's instructions (NEBNext® Ultra™ II DNA
521 Library Prep Kit for Illumina) using the following modified library amplification program: 98 °C for 30 s
522 (98 °C for 10 s, 65 °C for 15 s) × 15 cycles, 65 °C for 5 min, hold at 4 °C. Average library size and quality
523 control were performed using a Fragment Analyzer (High Sensitivity NGS kit) and qPCR (Roche Light
524 Cycler 480). CUT&RUN libraries were sequenced on a NovaSeq 6000 (Illumina) from Biocampus, MGX
525 platform, using a paired-end 150-bp run. E11.5, and E12.5 H3K27me3 CUT&RUN were performed on one
526 replicate of five and two pooled female embryos, respectively.

527

528 Allele-specific CUT&RUN bioinformatic analysis

529 FastQC (v0.11.9) and MultiQC (v1.13)⁶⁷ were used to control CUT&RUN data quality. UMIs were used
530 to confirm the homogeneous yield of library amplification and sequencing for each sample. Paired-end
531 reads with at least one undefined UMI were discarded using Cutadapt v4.1⁶⁸ and seqkit v2.3⁶⁹. The reads
532 were then trimmed using Trim Galore(v0.6.6¹⁶)
533 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/⁷⁰; options "--length 20 --illumina --
534 2colour 20"). Trimmed reads were mapped with Bowtie2 (v2.4.2)⁷¹ (options "--end-to-end --very-sensitive
535 --reorder") to the mm10 reference genome, modified as followed: autosomal, sexual and mitochondrial
536 chromosomes are kept, and 20 668 274 SNPs positions (0.76% of genome size) related to *Mus musculus*
537 *Castaneus* strain are N-masked using SNPsplit (v0.6)⁷²
538 ([16](https://ftp.ebi.ac.uk/pub/databases/mousegenomes/REL-1505-</p></div><div data-bbox=)

539 SNPs_Indels/strain_specific_vcfs/CAST_EiJ.mgp.v5.snps.dbSNP142.vcf). SAMtools (v1.11)⁷³ was used
540 to sort and convert the data formats. PCR duplicates were removed using GATK MarkDuplicates, and reads
541 were unmapped with or without primary alignment discarded (SAMtools options “-F 0x04 -F 0x100 -F
542 0x800”). The coverage for each developmental stage was calculated from the bam file using the deepTools
543 bamCoverage tool (v3.5.1, normalization by scale factor, bin=10 nt)⁷⁴.
544 Peak calling of H3K27me3 histone marks was performed for E11.5 and E12.5 female samples using
545 immunoglobulin G (IgG) as input for each stage (E11.5 IgG from male PGC, E12.5 IgG from female PGC)
546 with MACS2 (v2.2.7.1, options “-f BAMPE –broad –broad-cutoff 0.1”)⁷⁵.
547 Allele-specific reads (*i.e.* reads from B6: Xa and Cast: Xi) were sorted using SNPsplit. The d-score
548 parameter (reads Cast / reads B6 + reads Cast) was calculated using featureCounts from the Rsubread R
549 package(v2.12.3)⁷⁶. H3K27me3 enrichment heatmaps show the log₂ fold change in the coverage difference
550 between the H3K27me3 mark and the IgG control for each developmental stage, normalized by the
551 sequencing depth using deepTools bamCompare (bin= 10 nt). computeMatrix and plotHeatmap.

552

553 Whole Genome Bisulfite sequencing of female Epiblast

554 C57Bl6/J Epiblasts were manually dissected from extra-embryonic tissues of E6.5 embryos, followed by
555 sex determination by PCR on the extra-embryonic tissue (see section sexing of the embryos). Whole-
556 Genome Bisulfite sequencing libraries from 2 E6.5 female replicates were prepared as described by
557 Smallwood et al⁷⁷. The WGBS was analysed as described previously⁷⁸. Briefly, reads generated in this study
558 or recovered from the available datasets were treated as follows: the first eight base pairs of the reads were
559 trimmed using the FASTX-Toolkit v0.0.13 (hannonlab.cshl.edu/fastx_toolkit/index.html). Adapter
560 sequences were removed with Cutadapt v1.3 (code.google.com/p/cutadapt/)⁶⁸ and reads shorter than 16 bp
561 were discarded. The cleaned sequences were aligned to the mouse reference genome (mm10) using Bismark
562 v0.12.5 70 with Bowtie2-2.1.0 71 and the default parameters. Only the reads that mapped uniquely to the
563 genome were conserved. Sequencing statistics can be found in the Figure Legend and/or the main text.
564 Methylation calls were extracted after duplicate removal. Only CG dinucleotides covered by a minimum of
565 ten reads were conserved for the remainder of the analysis.

566

567 DNA methylation and transposon data analysis

568 DNA methylation data at E10.5 and E12.5, were downloaded from DRA000607⁵³ and GSE76971⁷⁹
569 respectively. The raw data were cleaned using Trim Galore v0.4.4⁷⁰. The cleaned reads were aligned to the
570 mouse reference genome assembly (GRCm38/mm10) using Bismark v0.18.2⁸⁰ with Bowtie2-2.2.9⁷¹
571 allowing for one mismatch in the seed alignment. Only reads that mapped uniquely to the genome were

572 retained, and methylation calls were extracted after duplicate removal, considering only CpG dinucleotides
573 covered by a minimum of five reads.

574 For the control regions of transposon enrichment, a set of 25 random regions (number of *Xist* entry sites),
575 with the same length as the median *Xist* entry sites) was bootstrapped 1 000 times either genome-wide or
576 on the X chromosome, using the R package *regioneR* (v1.10.0, Gel B, Diez-Villanueva A, Serra E,
577 Buschbeck M, Peinado MA, Malinverni R (2016). “*regioneR*: an R/Bioconductor package for the
578 association analysis of genomic regions based on permutation tests.” *Bioinformatics*, **32**(2), 289-291). The
579 number of retrotransposons overlapping *Xist* entry sites and random regions was calculated and normalized
580 to a 10 kb window. DNA methylation levels were estimated using a window of -1 kb to +100 bp from the
581 Transcriptional Start Site (TSS) for each gene. In all the cases, a permutation test was performed using the
582 *Regioner* package.

583

584 Statistics.

585 Statistical significance was evaluated using Kruskal–Wallis followed by Dunn’s correction and t-tests. *P*
586 values are provided in the figure legends and/or the main text.

587

588 **Data availability**

589 All sequencing data will be deposited in GEO and made publicly available after publication of this article
590 in a peer-reviewed scientific journal.

591

592 **Authors’ contribution**

593 M.B. conceived the study with M.A.S. and performed scRNA-seq experiments. C.R. performed chromatin
594 analysis, handled mouse colonies, and collected embryos with the help and supervision of K.C., L.S., E.B.,
595 and D.Z. performed bioinformatics analysis, supervised by N.S. and M.B.. A.T. performed the repeats and
596 WGBS bioinformatics analysis. C.L. contributed to the animal husbandry and sample collection. M.W. and
597 D.B. conducted WGBS experiments. M.B., M.A.S., and D.B. secured the funding. M.B. wrote the original
598 manuscript. with inputs from co-authors.

599

600 **Acknowledgments**

601 We acknowledge the Surani, Bourc’his, and Heard labs for insightful discussions, Kay Harnish, Cambridge
602 Stem Cell Institute Genomics Facility, and the MGX-Montpellier GenomiX platform for advice and deep
603 sequencing of the libraries. We thank the pathogen-free barrier animal facility of Gurdon Institute and the
604 PCEA and ZEFI of IGMM, UMR5535, and Montpellier. We are grateful to Joan Barau, Dominique
605 Helmlinger, and Lorraine Bonneville for their advice and reagents regarding the low-input CUT&RUN.

606 This work was supported by a CNRS-INSERM ATIP-Avenir grant, the FRM AJE202005011598 and an
607 ANR (the French National Research Agency) under the "Investissements d'avenir" programme with the
608 reference ANR-16-IDEX-0006 » to M.B.; Wellcome Trust funding 096738 and 092096 and Cancer
609 Research UK program C6946/A14492 to M.A.S., by a PhD fellowship from La Ligue nationale contre le
610 cancer to C.R. and FRM SPE20150331826 and a Marie Skłodowska-Curie Individual Fellowship (H2020-
611 MSCA-IF-2015—no. 706144) to M.B. MGX acknowledges financial support from France Génomique
612 National infrastructure, funded as part of “Investissement d’Avenir” program managed by Agence
613 Nationale pour la Recherche (contract ANR-10-INBS-09).

614

615 **Competing interests:**

616 Authors declare no competing interests-

617

618

619 **References**

- 620
- 621 1. Leitch, H. G., Tang, W. W. C. & Surani, M. A. Primordial Germ-Cell Development and Epigenetic
- 622 Reprogramming in Mammals. in *Current Topics in Developmental Biology* vol. 104 149–187 (Elsevier,
- 623 2013).
- 624 2. Tang, W. W. C., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M. A. Specification and epigenetic
- 625 programming of the human germ line. *Nature Reviews Genetics* **17**, 585–600 (2016).
- 626 3. Kurimoto, K. & Saitou, M. *Germ cell reprogramming. Current Topics in Developmental Biology* vol. 135
- 627 (Elsevier Inc., 2019).
- 628 4. Kurimoto, K. & Saitou, M. Epigenome regulation during germ cell specification and development from
- 629 pluripotent stem cells. *Current Opinion in Genetics & Development* **52**, 57–64 (2018).
- 630 5. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line.
- 631 *Nature* **452**, 877–81 (2008).
- 632 6. Sugimoto, M. & Abe, K. X chromosome reactivation initiates in nascent primordial germ cells in mice.
- 633 *PLoS genetics* **3**, e116 (2007).
- 634 7. Chuva de Sousa Lopes, S. M. *et al.* X chromosome activity in mouse XX primordial germ cells. *PLoS*
- 635 *genetics* **4**, e30 (2008).
- 636 8. Severino, J. *et al.* Controlled X-chromosome dynamics defines meiotic potential of female mouse in
- 637 vitro germ cells. *The EMBO Journal* **41**, e109457 (2022).
- 638 9. Mallol, A., Guirola, M. & Payer, B. PRDM14 controls X - chromosomal and global epigenetic
- 639 reprogramming of H3K27me3 in migrating mouse primordial germ cells. *Epigenetics & Chromatin* **12**,
- 640 1–10 (2019).
- 641 10. Sangrithi, M. N. *et al.* Non-Canonical and Sexually Dimorphic X Dosage Compensation States in
- 642 the Mouse and Human Germline. *Developmental Cell* **40**, 1–13 (2017).
- 643 11. Naik, H. C. *et al.* Lineage-specific dynamics of erasure of X-upregulation during inactive-X
- 644 reactivation. 2020.12.23.424181 Preprint at <https://doi.org/10.1101/2020.12.23.424181> (2023).
- 645 12. Heard, E. & Turner, J. Function of the Sex Chromosomes in Mammalian Fertility. *Cold Spring*
- 646 *Harb Perspect Biol* 1–18 (2011).
- 647 13. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–
- 648 3 (1961).
- 649 14. Loda, A., Collombet, S. & Heard, E. Gene regulation in time and space during X-chromosome
- 650 inactivation. *Nat Rev Mol Cell Biol* **23**, 231–249 (2022).
- 651 15. Namekawa, S. H., Payer, B., Huynh, K. D., Jaenisch, R. & Lee, J. T. Two-step imprinted X
- 652 inactivation: repeat versus genic silencing in the mouse. *Molecular and cellular biology* **30**, 3187–
- 653 3205 (2010).
- 654 16. Borensztein, M. *et al.* Xist -dependent imprinted X inactivation and the early developmental
- 655 consequences of its failure. *Nature structural & molecular biology* **24**, 226–233 (2017).
- 656 17. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are
- 657 defective in dosage compensation but not spermatogenesis. *Genes & Development* **11**, 156–166
- 658 (1997).
- 659 18. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread
- 660 across the X chromosome. *Science (New York, N.Y.)* **341**, 1237973 (2013).

- 661 19. Pandya-Jones, A. *et al.* A protein assembly mediates Xist localization and gene silencing. *Nature*
662 (2020) doi:10.1038/s41586-020-2703-0.
- 663 20. Strehle, M. & Guttman, M. Xist drives spatial compartmentalization of DNA and protein to
664 orchestrate initiation and maintenance of X inactivation. *Curr Opin Cell Biol* **64**, 139–147 (2020).
- 665 21. Chaumeil, J., Baccon, P. L., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a
666 repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–
667 2237 (2006).
- 668 22. Chow, J. C. *et al.* LINE-1 activity in facultative heterochromatin formation during X chromosome
669 inactivation. *Cell* **141**, 956–69 (2010).
- 670 23. Zylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell*
671 **176**, 182-197.e23 (2019).
- 672 24. Monk, M. Methylation and the X chromosome. *Bioessays* **4**, 204–208 (1986).
- 673 25. Cheng, S. *et al.* Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition
674 and X Chromosome Dynamics during Early Mouse Development. *Cell Reports* **26**, 2593–2607 (2019).
- 675 26. Marks, H. *et al.* Dynamics of gene silencing during X inactivation using allele-specific RNA-seq.
676 *Genome biology* **16**, 149 (2015).
- 677 27. Sousa, L. B. de A. e *et al.* Kinetics of Xist -induced gene silencing can be predicted from
678 combinations of epigenetic and genomic features. *Genome research* (2019) doi:doi:
679 10.1101/gr.245027.118.
- 680 28. Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos. *Science*
681 (*New York, N.Y.*) **303**, 666–9 (2004).
- 682 29. Borensztein, M. *et al.* Contribution of epigenetic landscapes and transcription factors to X-
683 chromosome reactivation in the inner cell mass. *Nature Communications* **8**, 1–14 (2017).
- 684 30. Pasque, V. *et al.* X chromosome reactivation dynamics reveal stages of reprogramming to
685 pluripotency. *Cell* **159**, 1681–1697 (2014).
- 686 31. Janiszewski, A. *et al.* Dynamic reversal of random X-Chromosome inactivation during iPSC
687 reprogramming. *Genome research* **29**, 1659–1672 (2019).
- 688 32. de Napoles, M., Nesterova, T. & Brockdorff, N. Early loss of Xist RNA expression and inactive X
689 chromosome associated chromatin modification in developing primordial germ cells. *PLoS One* **2**,
690 e860 (2007).
- 691 33. Mallol, A., Guirola, M. & Payer, B. PRDM14 controls X - chromosomal and global epigenetic
692 reprogramming of H3K27me3 in migrating mouse primordial germ cells. *Epigenetics & Chromatin* **12**,
693 1–10 (2019).
- 694 34. Navarro, P. *et al.* Molecular coupling of Xist regulation and pluripotency. *Science (New York,*
695 *N.Y.)* **321**, 1693–5 (2008).
- 696 35. Payer, B. *et al.* Tsix RNA and the Germline Factor, PRDM14, Link X Reactivation and Stem Cell
697 Reprogramming. *Molecular Cell* **52**, 1–14 (2013).
- 698 36. Wang, M., Lin, F., Xing, K. & Liu, L. Random X-chromosome inactivation dynamics in vivo by
699 single-cell RNA sequencing. *BMC Genomics* **18**, 90 (2017).
- 700 37. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation.
701 *Nature* **477**, 289–94 (2011).

- 702 38. Frazer, K. A. *et al.* A sequence-based variation map of 8 . 27 million SNPs in inbred mouse
703 strains. *Nature* **448**, 1050–1053 (2007).
- 704 39. Payer, B. *et al.* Generation of stella-GFP Transgenic Mice : A Novel Tool to Study Germ Cell
705 Development. *Genesis* **44**, 75–83 (2006).
- 706 40. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of
707 embryonal cells. *Development* **122**, 881–94 (1996).
- 708 41. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods* **6**,
709 (2009).
- 710 42. Borensztein, M., Syx, L., Servant, N. & Heard, E. Transcriptome Profiling of Single Mouse
711 Oocytes. *Methods in Molecular Biology* **1818**, 51–65 (2018).
- 712 43. Mayère, C. *et al.* Single-cell transcriptomics reveal temporal dynamics of critical regulators of
713 germ cell fate during mouse sex determination. *The FASEB Journal* **35**, e21452 (2021).
- 714 44. Yamaji, M. *et al.* DND1 maintains germline stem cells via recruitment of the CCR4-NOT complex
715 to target mRNAs. *Nature* **543**, 568–572 (2017).
- 716 45. Bertozzi, T. M., Elmer, J. L., Macfarlan, T. S. & Ferguson-Smith, A. C. KRAB zinc finger protein
717 diversification drives mammalian interindividual methylation variability. *Proc. Natl. Acad. Sci. U.S.A.*
718 **117**, 31290–31300 (2020).
- 719 46. Dehghanian, F., Bovio, P. P., Hojati, Z. & Vogel, T. ZFP982 confers mouse embryonic stem cell
720 characteristics by regulating expression of Nanog, Zfp42 and Dppa3. 2020.06.03.131847 Preprint at
721 <https://doi.org/10.1101/2020.06.03.131847> (2020).
- 722 47. Yu, L. *et al.* Loss of ESRP1 blocks mouse oocyte development and leads to female infertility.
723 *Development* **148**, dev196931 (2021).
- 724 48. Zhao, Z.-H. *et al.* Single cell RNA sequencing reveals the landscape of early female germ cell
725 development. *FASEB Journal* 2020.05.09.085845 (2020) doi:10.1101/2020.05.09.085845.
- 726 49. Cattanach, B. M. & Williams, C. E. Evidence of non-random X chromosome activity in the mouse.
727 *Genetics Research* **19**, 229–240 (1972).
- 728 50. Calaway, J. D. *et al.* Genetic Architecture of Skewed X Inactivation in the Laboratory Mouse. *PLoS*
729 *Genet* **9**, e1003853 (2013).
- 730 51. Berletch, J. B. *et al.* Escape from X Inactivation Varies in Mouse Tissues. *PLOS Genetics* **11**,
731 e1005079 (2015).
- 732 52. Pandya-Jones, A. & Plath, K. The ???Inc??? between 3D chromatin structure and X chromosome
733 inactivation. *Seminars in Cell and Developmental Biology* **56**, 35–47 (2016).
- 734 53. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies
735 gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
- 736 54. Hill, P. W. S. *et al.* Epigenetic reprogramming enables the primordial germ cell-to-gonocyte
737 transition Europe PMC Funders Group. *Nature* **555**, 392–396 (2018).
- 738 55. Richart, L. *et al.* XIST loss impairs mammary stem cell differentiation and increases
739 tumorigenicity through Mediator hyperactivation. *Cell* **185**, 2164-2183.e25 (2022).
- 740 56. Generoso, S. F. *et al.* Cohesin controls X chromosome structure remodeling and X-reactivation
741 during mouse iPSC-reprogramming. *Proceedings of the National Academy of Sciences* **120**,
742 e2213810120 (2023).

- 743 57. Camlin, N. J., McLaughlin, E. A. & Holt, J. E. Kif4 Is Essential for Mouse Oocyte Meiosis. *PLoS One*
744 **12**, e0170650 (2017).
- 745 58. Luoh, S.-W. *et al.* Zfx mutation results in small animal size and reduced germ cell number in male
746 and female mice. *Development* **124**, 2275–2284 (1997).
- 747 59. Lan, Z. J., Xu, X. & Cooney, A. J. Differential oocyte-specific expression of Cre recombinase
748 activity in GDF-9-iCre, Zp3cre, and Msx2Cre transgenic mice. *Biol Reprod* **71**, 1469–1474 (2004).
- 749 60. Szabo, P. E., Hübner, K., Schöler, H. & Mann, J. R. Allele-specific expression of imprinted genes in
750 mouse migratory primordial germ cells. *Mechanisms of development* **115**, 157–160 (2002).
- 751 61. Tang, F., Lao, K. & Surani, M. A. Development and applications of single-cell transcriptome
752 analysis. *Nature methods* **8**, (2011).
- 753 64. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of
754 DNA binding sites. *eLife* **6**, 1–35 (2017).
- 755 65. Prakash, S. A. & Barau, J. Chromatin Profiling in Mouse Embryonic Germ Cells Embryonic germ
756 cells by CUT&RUN. in *Epigenetic Reprogramming During Mouse Embryogenesis: Methods and*
757 *Protocols* (eds. Ancelin, K. & Borensztein, M.) 253–264 (Springer US, 2021). doi:10.1007/978-1-0716-
758 0958-3_17.
- 759 66. Dura, M. *et al.* DNMT3A-dependent DNA methylation is required for spermatogonial stem cells
760 to commit to spermatogenesis. *Nat Genet* **54**, 469–480 (2022).
- 761 67. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for
762 multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
- 763 68. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
764 *EMBnet.journal* **17**, 10–12 (2011).
- 765 69. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File
766 Manipulation. *PLOS ONE* **11**, e0163962 (2016).
- 767 70. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. *FelixKrueger/TrimGalore:*
768 *v0.6.7 - DOI via Zenodo.* (Zenodo, 2021). doi:10.5281/zenodo.5127899.
- 769 71. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–
770 359 (2012).
- 771 72. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes
772 with known SNP genotypes. *F1000Res* **5**, 1479 (2016).
- 773 73. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
- 774 74. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
775 *Nucleic Acids Res* **44**, W160–W165 (2016).
- 776 75. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
- 777 76. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
778 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- 779 77. Smallwood, S. a *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic
780 heterogeneity. *Nature Methods* **11**, 817–20 (2014).
- 781 78. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc’his, D. An epigenetic switch ensures
782 transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**,
783 e11418 (2016).

784 80. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
785 applications. *Bioinformatics* **27**, 1571–1572 (2011).
786
787

788 **Figure Legends**

789

790 **Figure 1**

791 **Single-cell RNA sequencing of polymorphic primordial germ cells during embryonic development.**

792 (A) Schematic illustration of single-cell transcriptomic experiments including mouse breeding (129 × cast),
793 harvested embryonic stages, and single-cell collection. At E8.5, PGC were sorted according to the
794 expression of *Stella*-GFP. *Stella* gene, also known as *Dppa3*, is expressed in early PGC. From E9.5, *Oct4*-
795 *GFP* (GOF-ΔPE-18 line)⁴⁰ marker was used to differentiate between PGC and soma. PGC, primordial germ
796 cells; FACS, fluorescence-activated cell sorting. (B) Principal component analysis (PCA) of single PGC
797 and soma based on the 1 000 most variable genes in the transcriptomic datasets. The different stages are
798 denoted by different colours. The rounds represent PGC, and triangles represent GFP-negative somatic cells
799 in the proximity of the PGC. The number of cells analysed per stage and further details of the scRNA-seq
800 samples are shown in Supplementary Table 1. (C) Hierarchical clustering and Pearson's distance of scRNA-
801 seq samples based on germline, soma, and sex-specific gene expression variation using Pearson's
802 correlation. Cells were clustered first by lineage (PGC and soma), then by stage (E8.5–E12.5), and then by
803 sex for the E11.5 and E12.5 stages. n = 137 single-cell samples. (D) Expression levels of 26 known genes
804 expressed in developing PGC or soma, and sex-specific genes (*Xist* and Y-linked genes) in the 137 single-
805 cell samples were used to classify cells according to their lineage, as shown. The cells were ordered
806 according to hierarchical clustering in C. PGC primordial germ cells.

807

808 **Figure 2**

809 **Progression of X-chromosome reactivation in developing PGC.** (A) Principal component analysis

810 (PCA) based on scRNA-seq data from PGC and soma between E8.5-E12.5. X-chromosome composition
811 of each single cell is represented on the PCA, such as the XY male, XX female, and XO female cells. The
812 XO cells originate from a single embryo and were found by serendipity. PCA is based on the 1 000 most
813 variable genes, as described in **Figure 1B**. The details of each cell are listed in **Supplementary Table 1**.
814 (B) Allele-specific expression ratios for genes on autosomes and X chromosomes in female and male single
815 PGC from E9.5 to E12.5. The allelic ratio represents the number of reads mapped to Cast genome, divided
816 by the total number of 129 and Cast reads. For X-linked genes, we measured the allelic ratio as the parental
817 genome from which the Xi is originated, divided by the total number of 129 and Cast reads mapped for
818 each gene ($X_a \text{ counts} / \text{total } X_a + X_i \text{ counts}$). A gene was considered biallelically expressed when $0.2 <$
819 $\text{allelic ratio} < 0.8$. Box plots represent medians (centre lines) with lower and upper quartiles (box limits).
820 Whiskers represent $1.5 \times$ the interquartile range. Outliers are represented by dots. The number of cells
821 analysed per stage, and the parental origin of the Xi, are shown in **Supplementary Table 1**. (C) and (D)

822 Pseudotime representation of scRNA-seq data, based on the first principal component for female PGC
823 between E9.5 and E12.5. In **(D)**, the percentage of reactivated X-linked genes per single cell is provided by
824 a colour gradient as shown in the key. A gene is called reactivated if its allelic ratio is <0.8 , which represents
825 an expression from the Xi of at least 20 %

826

827 **Figure 3**

828 **Kinetics of reactivation of X-linked genes over the entire X chromosome in developing PGC during**
829 **reversal of random XCI.** The mean of the allele-specific expression ratios, per embryonic stage, for each
830 informative and expressed X-linked gene in female PGC are represented as heatmaps from E9.5 to E12.5,
831 with strict monoallelic Xa expression (ratio >0.8) in red and strict monoallelic Xi expression (ratio <0.2) in
832 blue. Color gradients is used in between these two values, as indicated in the key. Genes are ordered by
833 genomic position (left) and reactivation kinetics class (right). *Xist* expression was always below RPRT < 2
834 and its genomic location has been added to the heatmap for information (green arrow). $n = 198$ informative
835 X-linked genes, with a RPRT expression >2 , expressed in at least 3 out of 4 developmental stages. White
836 box, data not available (below threshold).

837

838 **Figure 4**

839 **Differential timing of X-linked gene reactivation is associated with timing of silencing and**
840 **chromosomal location in regards to *Xist* early sites.** **(A)** X-linked genes are clustered based on their
841 reactivation kinetics as early (expressed from the Xi at E10.5; allelic ratio <0.8 at E10.5), intermediate
842 (expressed from the Xi from E11.5), late (expressed from the Xi from E12.5), very late (not reactivated at
843 E12.5), and escapee (not undergoing XCI). The allelic ratio of each gene represents the fraction of Xa
844 expression, with the number of reads mapped on the Xa genome divided by the total number of reads
845 mapped. This is shown for stages E9.5 to E12.5 for all female PGC. $n = 207$ X-linked genes. Box plots are
846 as in Figure 2B. Further information is provided in **Methods**. **(B)** Expression level of X-linked genes in the
847 different reactivation-timing classes in female PGC (mean of each single gene). Expression of each gene
848 represents the total number of reads mapped, normalized by the covered gene length and is represented at
849 E9.5, E10.5, E11.5 and E12.5, as a function of the reactivation classes. Further information is provided in
850 **Methods**. $n = 207$ X-linked genes. Box plots are as in Figure 2B. Comparison of reactivation classes
851 between the inner cell mass of the blastocyst²⁹ **(C)** and the *in vitro* PGC-like cell system⁸ **(D)**. **(E)** Silencing
852 classes of informative X-linked genes in mouse embryonic stem cells (mESC) compared with their
853 reactivation classes in PGC²⁶. **(F)** *Xist* ‘entry’ sites are regions of the X chromosome showing early
854 accumulation of *Xist* RNA upon initiation of X-chromosome inactivation, and thought to be the closest to
855 *Xist* locus in 3D spatial proximity. Allelic expression of X-linked genes classified on the basis of their

856 relative position to *Xist* entry sites (as identified during XCI induction in ESC¹⁸): inside (TSS located inside
857 a *Xist* entry site), next to (TSS located less than 100 kb away from an entry site) and outside (over 100 kb
858 from an entry site). $p=0.05$ for E11.5 by Kruskal-Wallis test followed by Dunn's correction. $n = 207$ X-
859 linked genes. Box plots are as in **Figure 2B**. The numbers of cells analysed per stage is shown in
860 **Supplementary Table 1**.

861

862 **Figure 5**

863 **Contribution of repressive chromatin marks to resistance to early reactivation.** (A) Number of total
864 transposons, LINEs, and SINES, overlapping with *Xist* entry sites, control regions of the X chromosome
865 and control regions of autosomes. Set of control regions have been generated randomly 1000 times (see
866 section **Methods**). A t-test was performed to compare number of repeats in *Xist* entry sites compared to
867 controls. (B) Whole-Genome Bisulfite of E6.5 female epiblast cells compared to public datasets of
868 female PGC at E10.5 (DRA000607 in DDBJ database)⁵³ and E12.5 (GSE76971 in GEO database)⁵⁴. DNA
869 methylation level were estimated using a window of -1kb to +100bp to the TSS for each gene. (C)
870 Schematic illustration of the mouse breeding between C57Bl6/J *Xist* flox/flox; *Zp3*-Cre; *Oct4*-eGFP females
871 and Castaneus males in order to obtain female polymorphic embryos, with non-random XCI and fluorescent
872 PGC. PGC were isolated by FACS with the Oct4-eGFP reporter at E11.5 and E12.5. Following PGC
873 sorting, low-input CUT&RUN was completed for H3K27me3 marks (see **Methods**). (D) Distribution of
874 H3K27me3 CUT&RUN peaks in E11.5 and E12.5 female PGC. The Venn diagram shows H3K27me3
875 broad domain overlapping in X chromosomes between E11.5 and E12.5 female PGC (left), and between
876 the active (B6, Xa) and inactive (Cast, Xi) chromosomes, right. (E) Percentage of early ($n=29$), intermediate
877 ($n=55$), late ($n=20$), very late ($n=70$), and escapee ($n=8$) X-linked genes significantly enriched in
878 H3K27me3 repressive marks in both E11.5 and E12.5 female PGC (at least one H3K27me3 broad domain
879 per gene). (F) H3K27me3 enrichment (fold change compared to IgG and normalized by library size and
880 peak length) in the different X-linked gene reactivation classes. Kruskal Wallis test followed by Dunn's
881 post hoc test was performed to compare all classes. Each point represents a gene, in red the mean +/- sem.
882 E11.5: p -value $< 0,0001$, early and intermediate versus very late and E12.5: p -value = 0,0020 intermediate
883 versus very late. (G) enrichment (fold change compared to IgG and normalized by library size and peak
884 length) weighted by the d-score (allelic ratio, see **Methods**) at E11.5 (p -value = 0,0004, early and
885 intermediate versus very late), E12.5 (p -value = 0,0038, intermediate versus very late). Statistical test
886 (mean, SEM). Each point represents a gene, in red the mean +/- sem. E11.5: p -value $< 0,0001$, early and
887 intermediate versus very late and E12.5: p -value = 0,0020 intermediate versus very late. (H). Integrative
888 Genomics Viewer plot of representative genes from early, late and very late reactivation classes. Tracks
889 depict global and allele-specific H3K27me3 enrichment for *Med14* early reactivated gene, *Dlg3* and *Pjal*

890 late reactivated genes and *Med12* very late reactivated gene. Global enrichment tracks are in dark blue and
891 allele-specific tracks are overlaid with global enrichment tracks (Cast Xi reads in light blue; B6 Xa reads
892 in red). Dark blue boxes and highlighted grey area are significant H3K27me3 broad domains. Location is
893 given in mm10, with gene isoforms extracted from Integrative Genome viewer and UCSC.

894

895 **Extended Figure 1 Differential gene expression upon female PGC development and key markers of** 896 **soma and PGC**

897 (A). Principal Component Analysis (PCA) based on 26 known markers of PGC or soma (**Figure 1C** and
898 **D**). (B) PCA of the 30 most differentially expressed genes (DEGs) that contributed to lineage segregation.
899 (C-D) Volcano plots represent differentially expressed genes (DEG) between the two developmental stages
900 of female PGC. A few transcriptional changes have been observed in migratory PGC. Changes arise once
901 PGC colonize the gonads. Some examples of DEG are highlighted. Red dots represent upregulated genes,
902 and green dots represent downregulated genes. X-linked genes are shown in orange. They showed a
903 statistically greater enrichment in upregulated genes at E12.5, compared to E11.5, owing to X reactivation.

904

905 **Extended Figure 2 Clustering of PGC by sex.**

906 (A-D) PCA per developmental stage with sample names, lineage soma versus PGC, and sex information.
907 Cells were clustered by lineage. From E11.5, the cells clustered by sex. (E) Pseudotime representation of
908 the scRNA-seq data based on the first principal component for XX females (pink), XO females (red), and
909 XY males (blue). (F) Level of *Xist* expression and degree of reactivation in each single cell. Each dot
910 represents a single cell. Most female soma exhibit high *Xist* expression and a low number of biallelically
911 expressed genes. Genes with RPRT expression < 2 were considered to be unexpressed.

912

913 **Extended Figure 3 Contribution to kinetics of X-chromosome reactivation**

914 (A) Representation of the Gene ontology analysis of Biological process performed on the best correlated
915 genes with X-linked gene reactivation (p-value < 0.001). Correlation and anti-correlation between gene
916 expression levels (autosomes and X chromosomes) and the percentage of X-linked gene reactivation (allelic
917 ratio > 0.2 for X-linked genes) were measured using Pearson's correlation and the Benjamini-Hochberg
918 correction. The 20 best enrichment classes (based on fold enrichment) were represented by their p-values.

919 (B) Distance to escapee loci. Distribution of genomic distances to escapees (Mb) for different X-linked
920 gene reactivation classes. Transcription Start Site (TSS) of each gene was used to measure the distance
921 from the closest escaping gene. Non-significant by Kruskal-Wallis test. Boxplots represent the medians
922 with lower and upper quartiles. (C) Distance to *Xist* genomic locus. Distribution of genomic distances to
923 the *Xist* locus (Mb) for different X-linked gene reactivation classes. The Transcription Start Site (TSS) of

924 each gene was used to measure the distance from the *Xist* locus. Very late reactivated genes were
925 significantly closer to the *Xist* locus than early reactivated ($p=0.0041$), intermediate-reactivated ($p=0.0201$),
926 and escapee ($p=0.0045$) genes, according to the Kruskal-Wallis test. Boxplots represent the medians with
927 lower and upper quartiles.

928

929 **Extended Figure 4**

930 H3K27me3 enrichment by low-input CUT&RUN in E11.5 and E12.5 female PGC. Profile plots and their
931 corresponding heat maps around the transcription start sites (TSS) and transcription end sites (TES) of X-
932 linked genes for H3K27me3 repressive histone marks. Enrichment was extracted for the E11.5 and E12.5
933 CUT&RUN experiments within a region spanning ± 1 kb around TSS and TES. The blue-to-red gradient
934 indicates low to high enrichment in the corresponding regions ranked by reactivation classes.

935

936 **Supplementary Table 1**

937 Summary of sequenced data including information of scRNA-seq and CUT&RUN datasets.

938

939

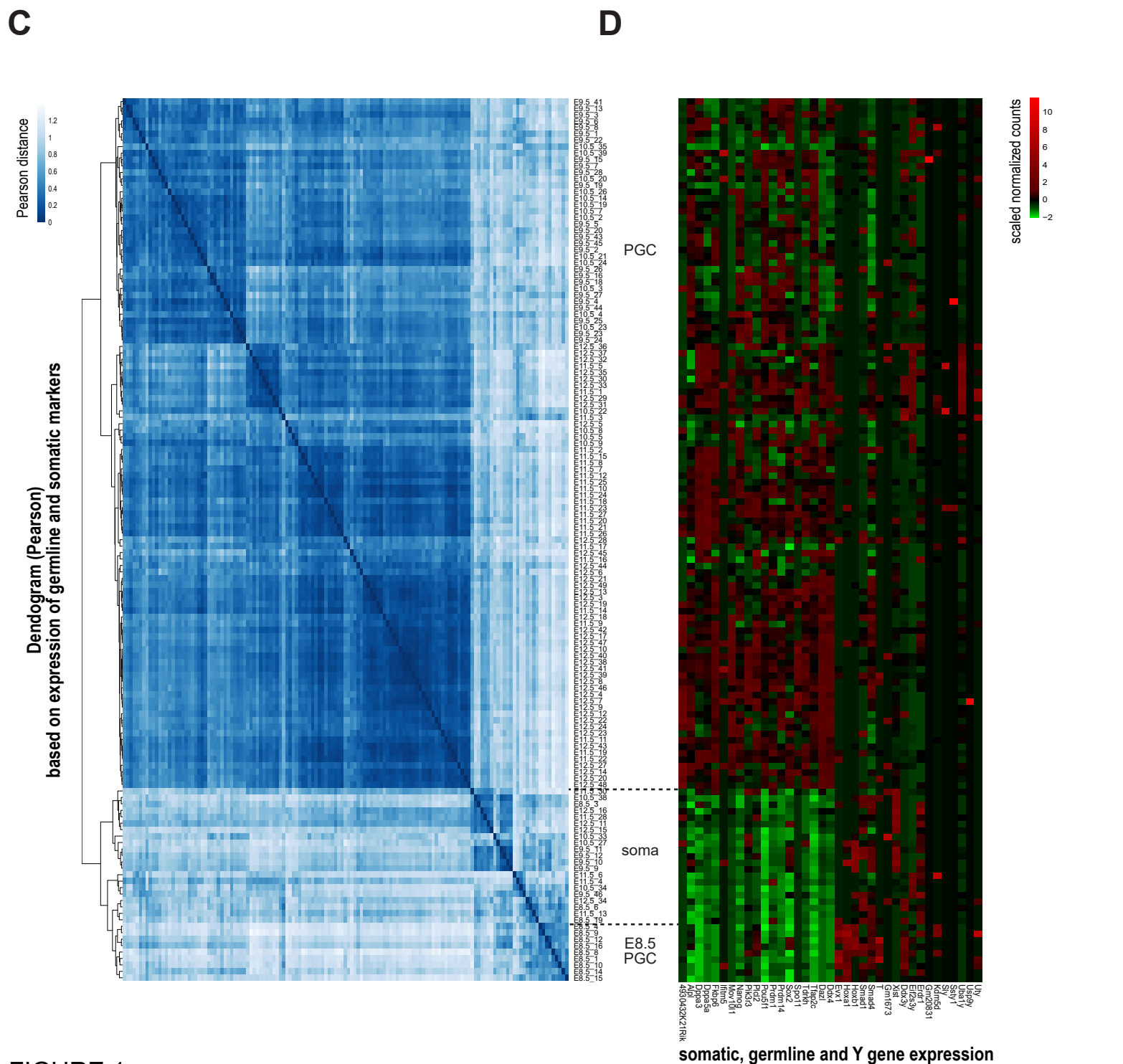
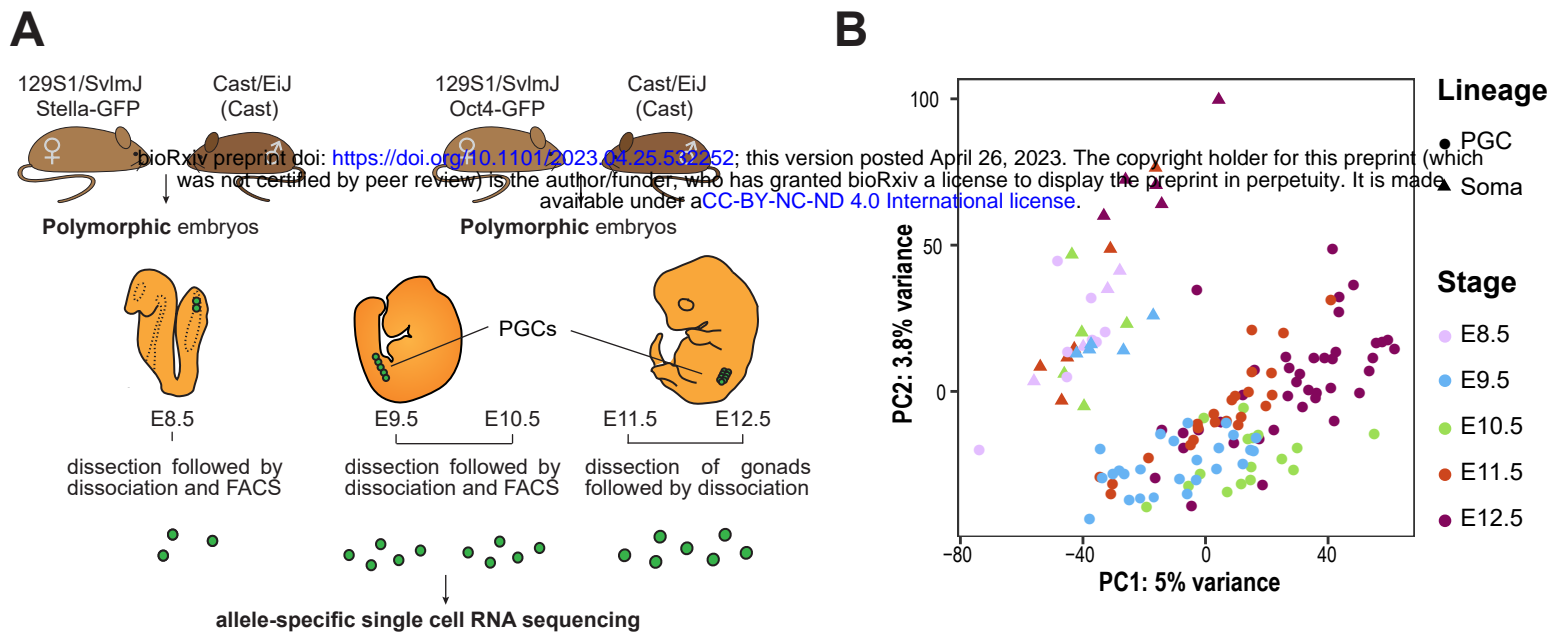


FIGURE 1

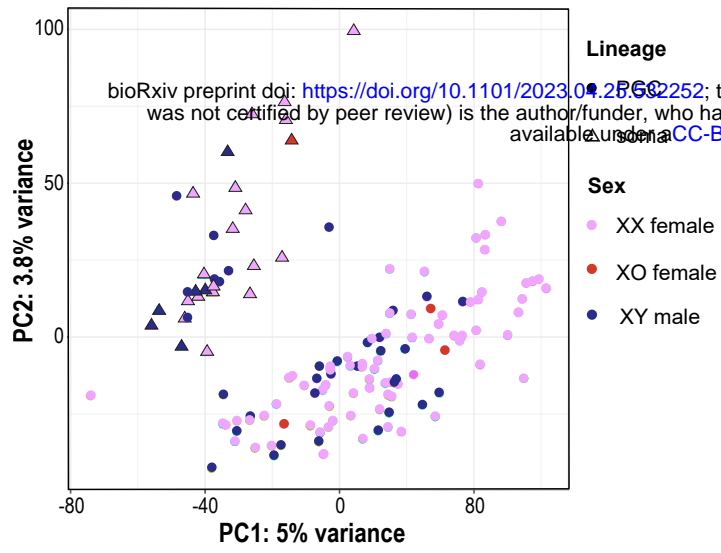
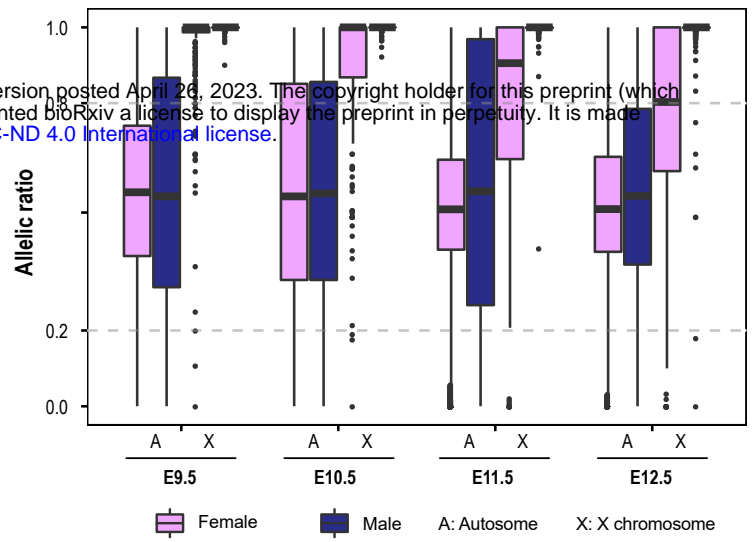
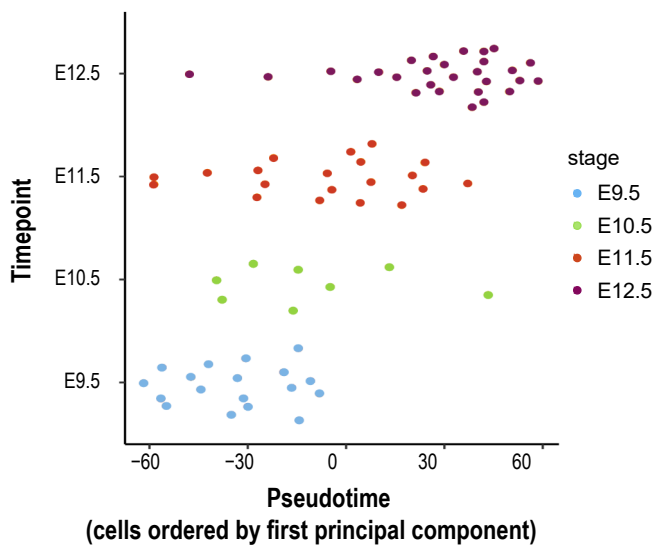
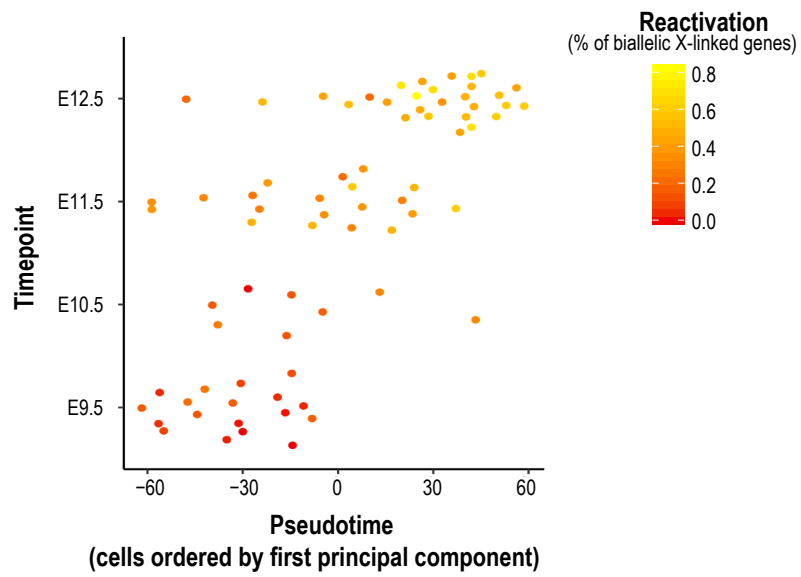
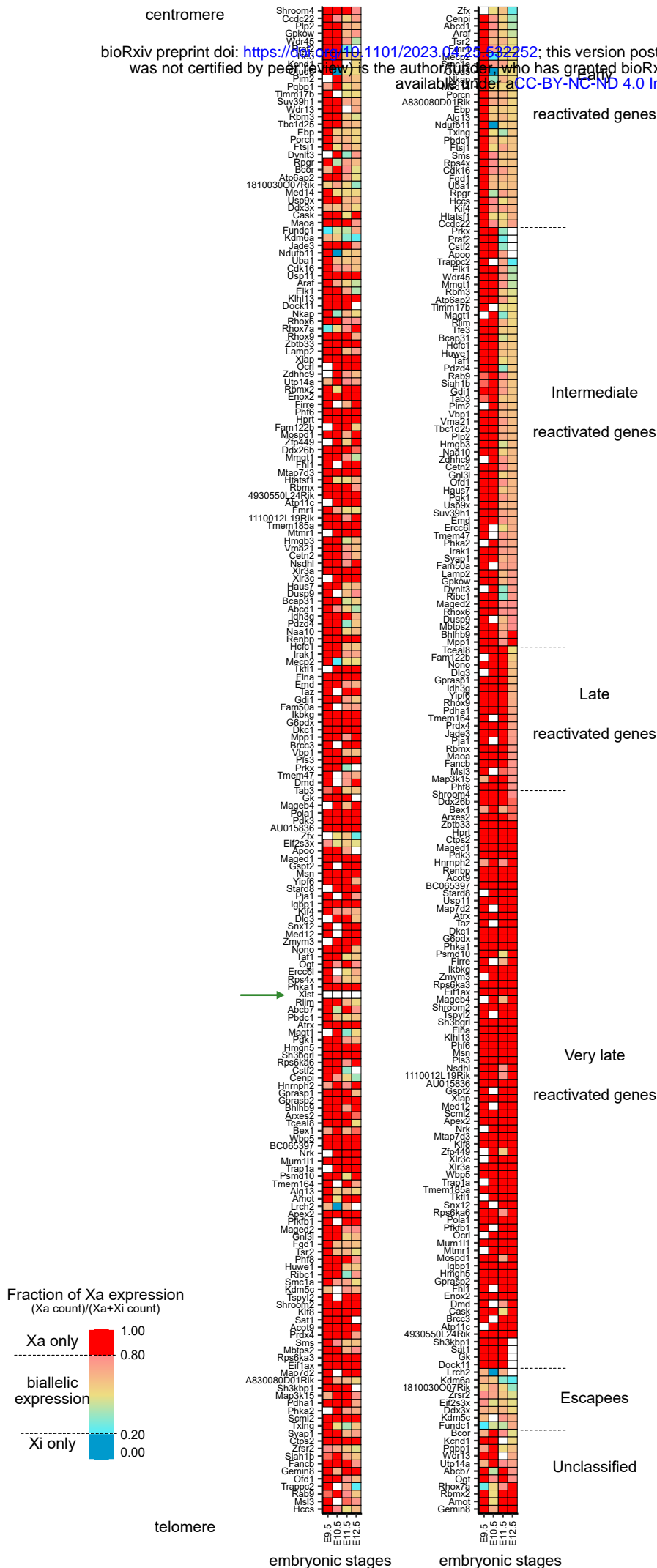
A**B****C****D**

FIGURE 2

ordered by
genomic position

ordered by
reactivation timing

bioRxiv preprint doi: <https://doi.org/10.1101/2023.04.26.532252>; this version posted April 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



reactivated genes

Intermediate

reactivated genes

Late

reactivated genes

Very late

reactivated genes

Escapes

Unclassified

FIGURE 3

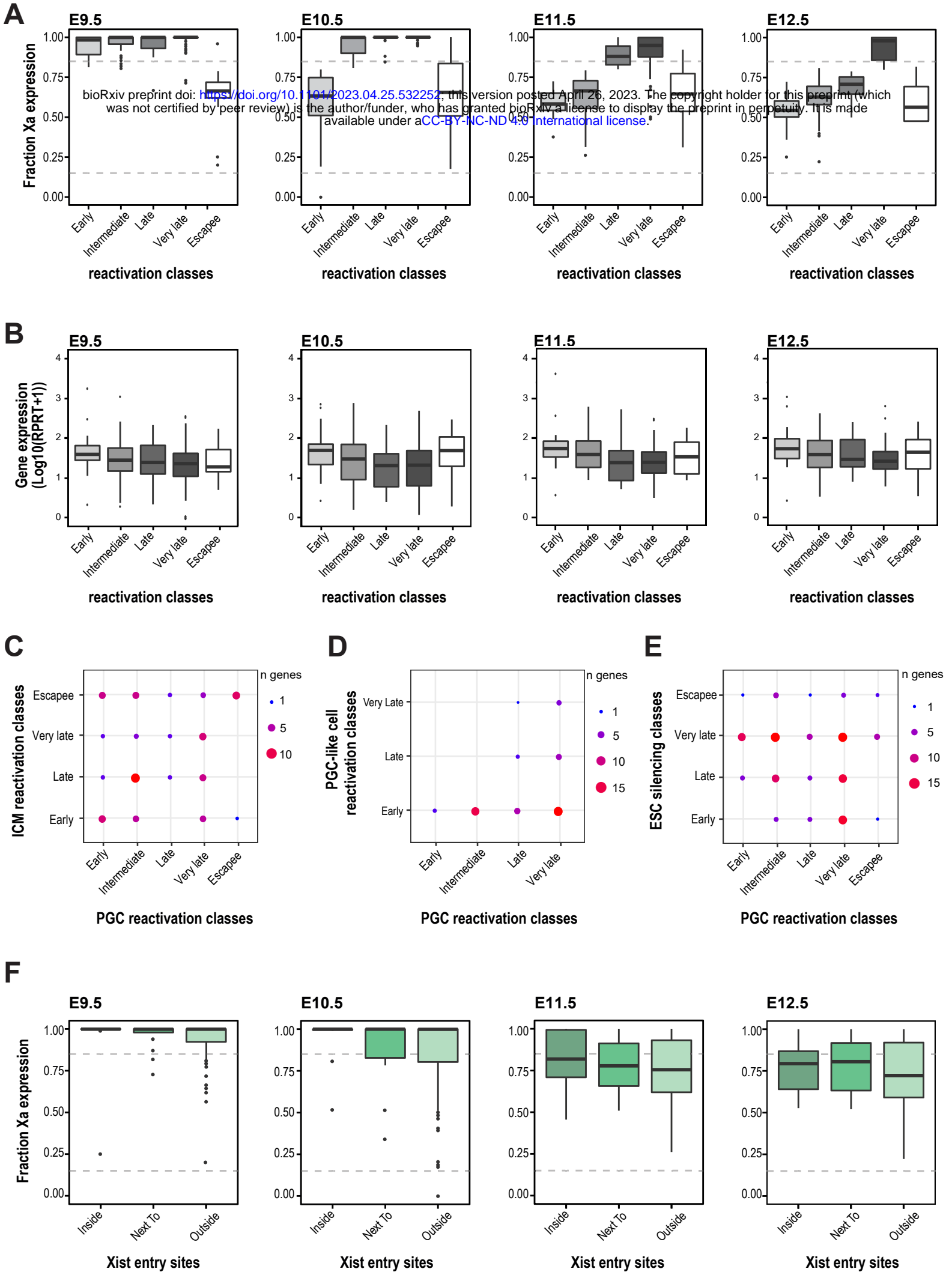


FIGURE 4

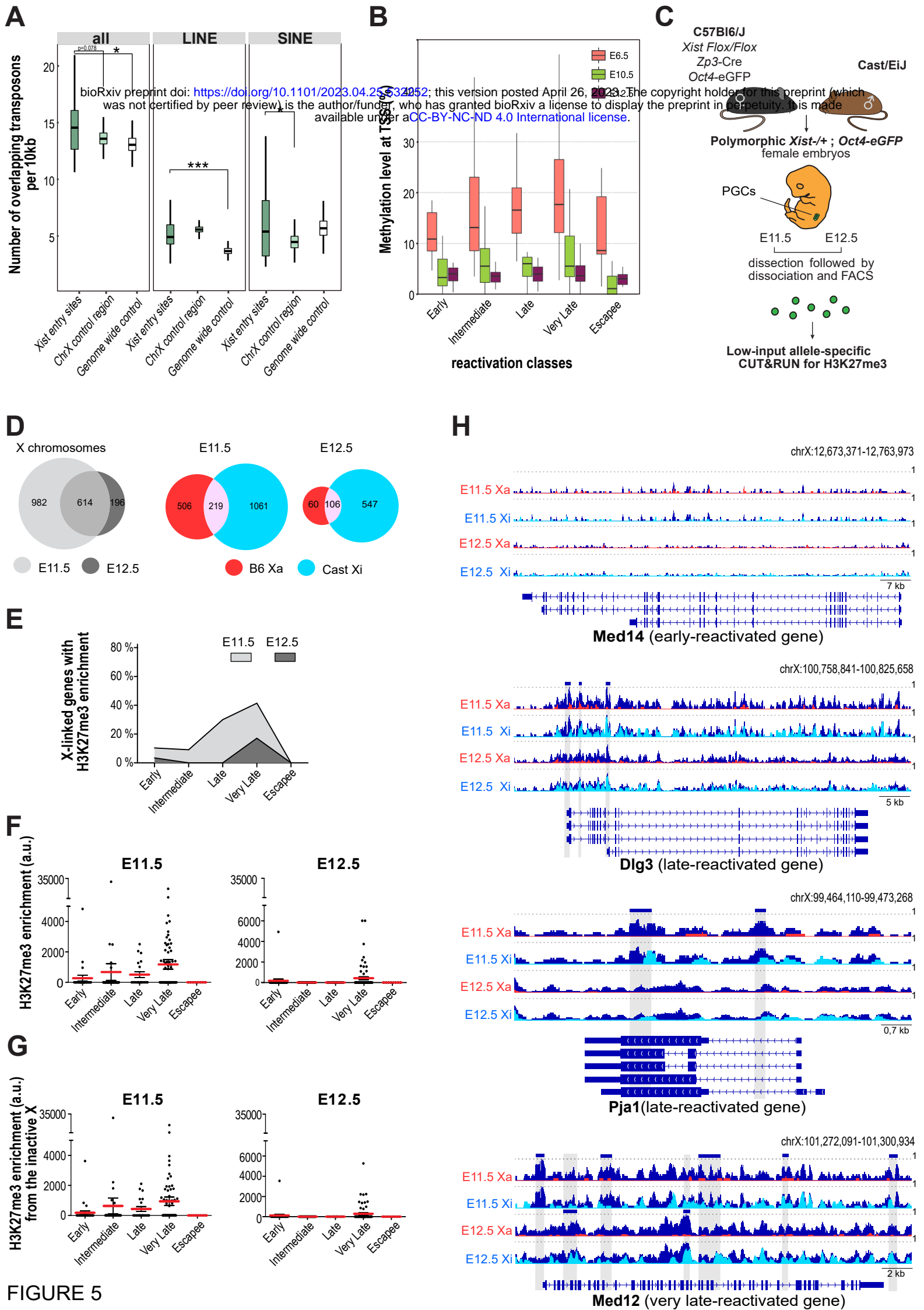
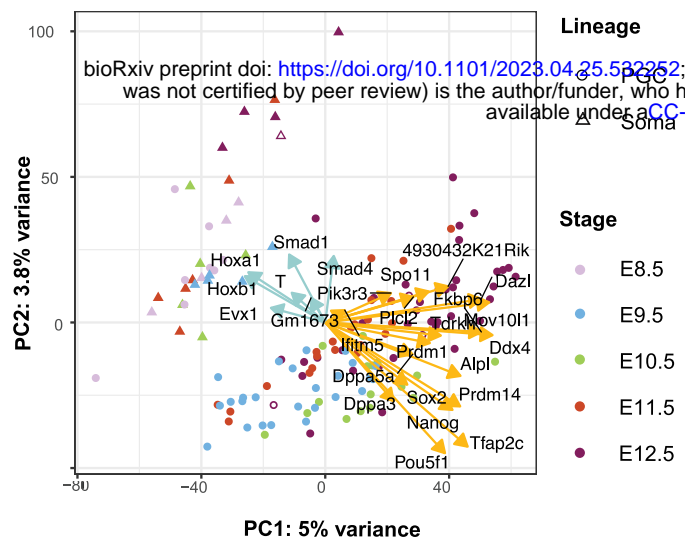
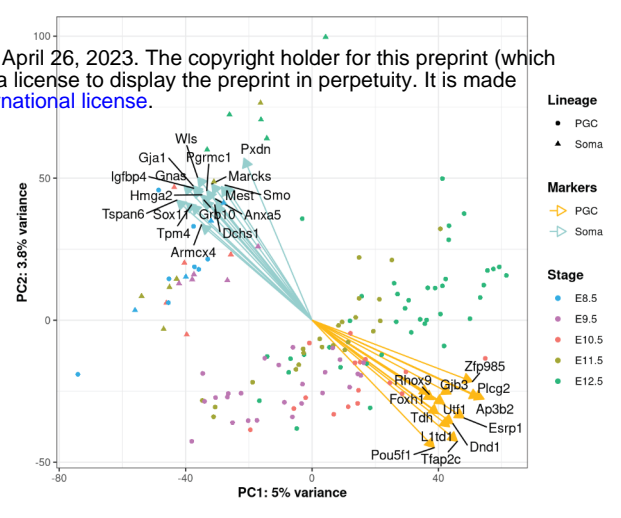
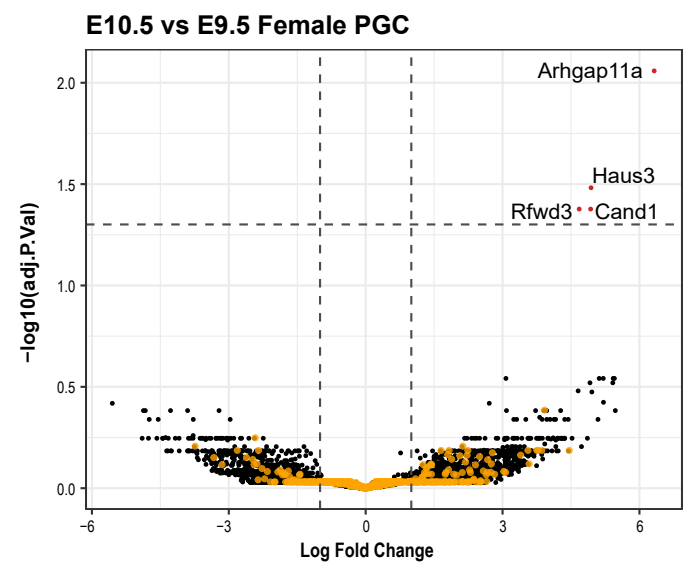
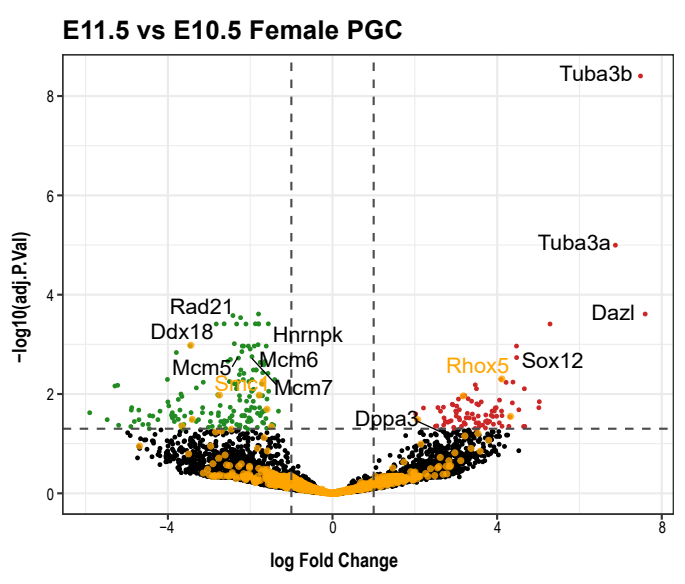
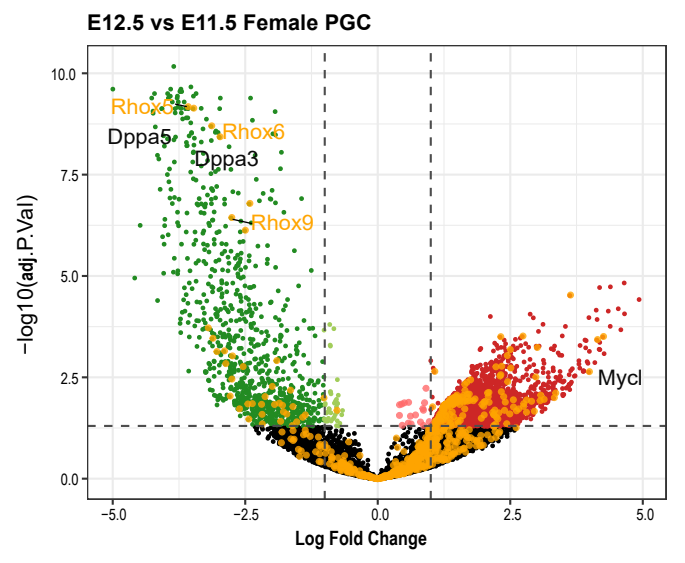
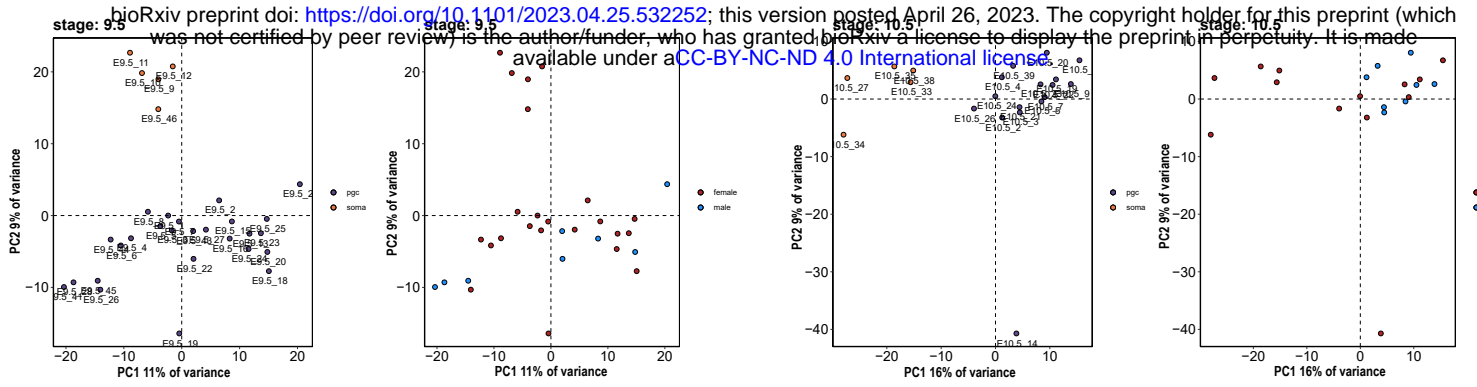
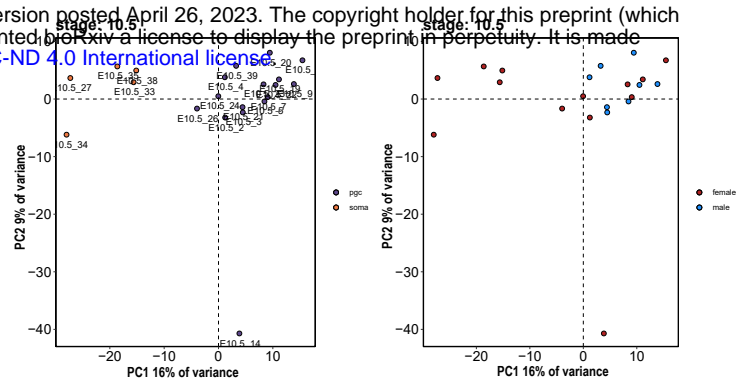
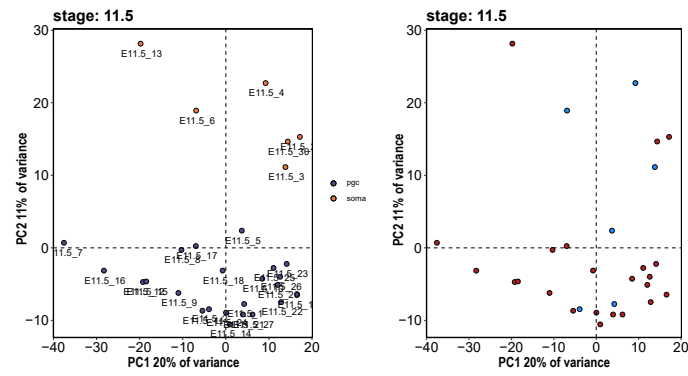
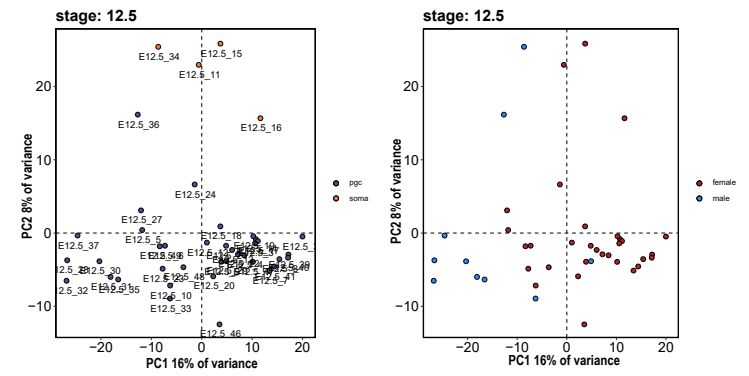
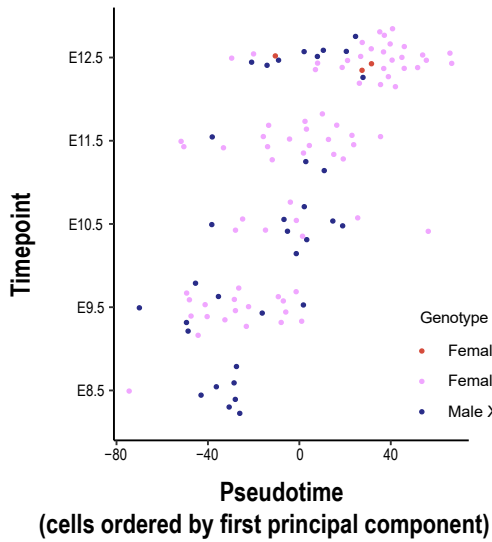
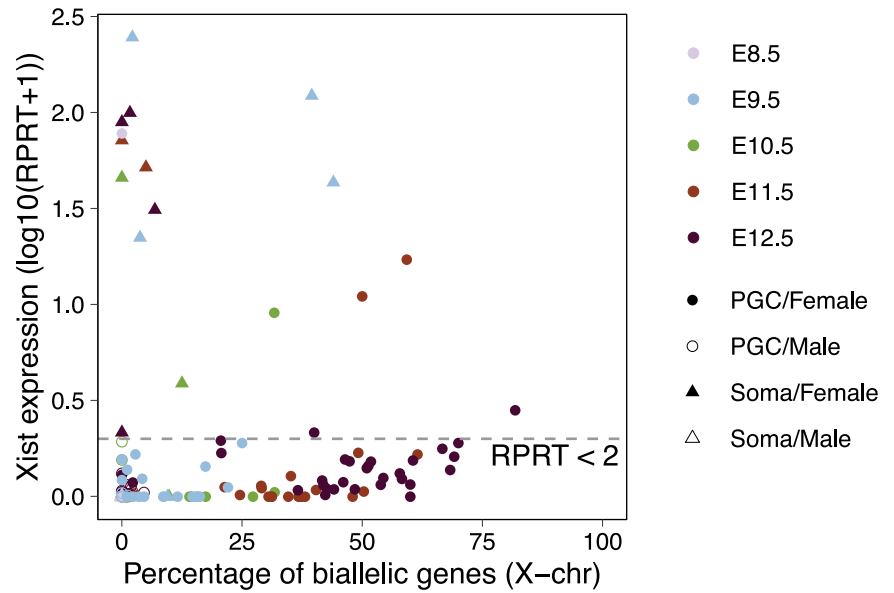
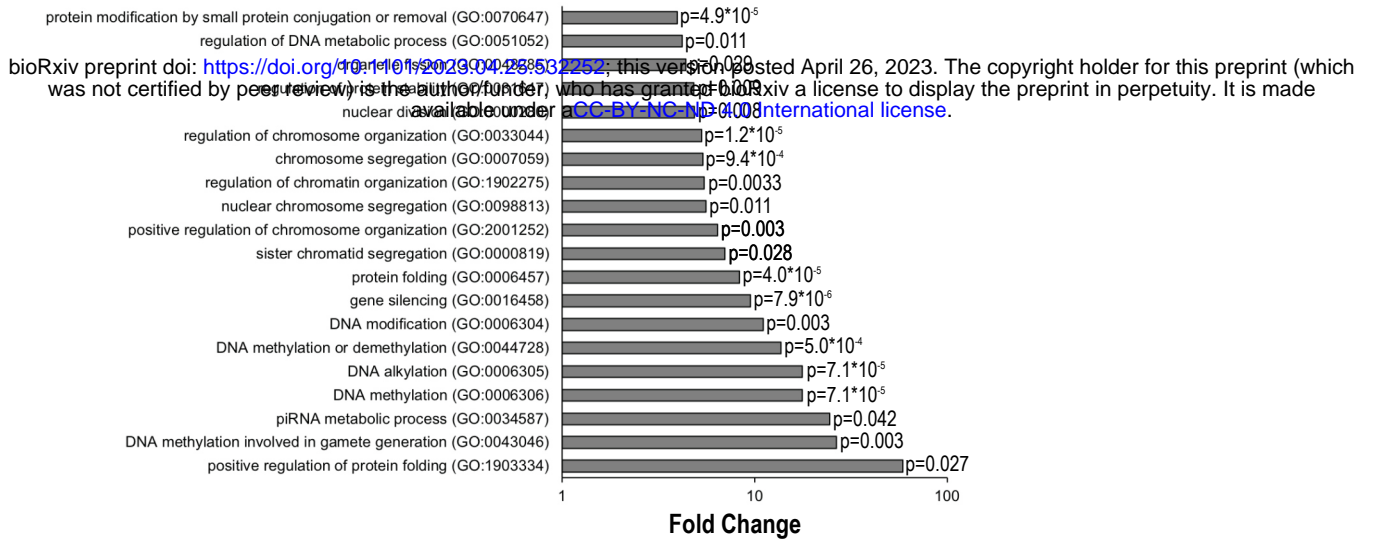
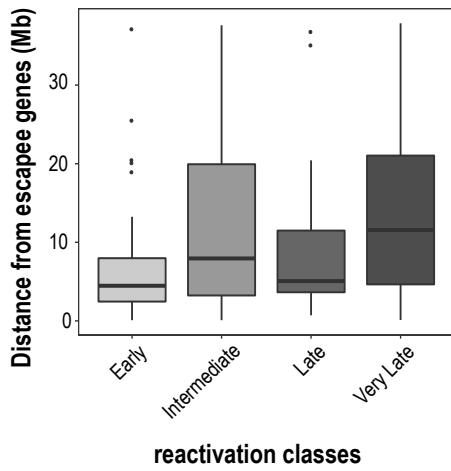
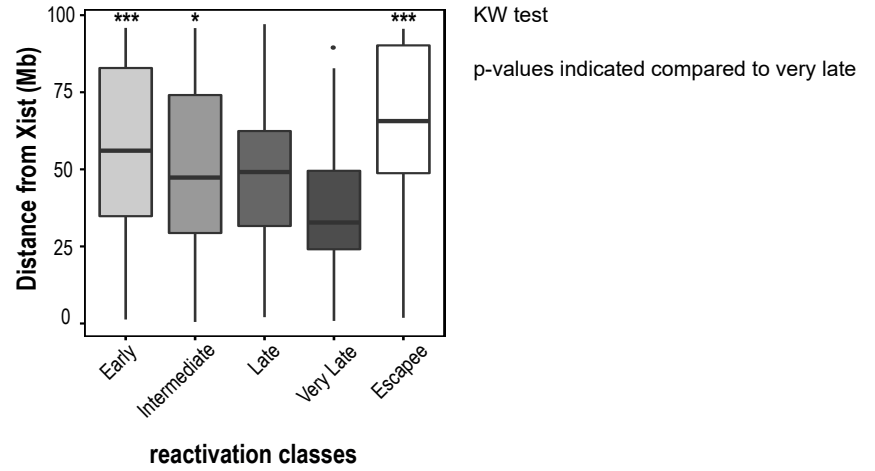


FIGURE 5

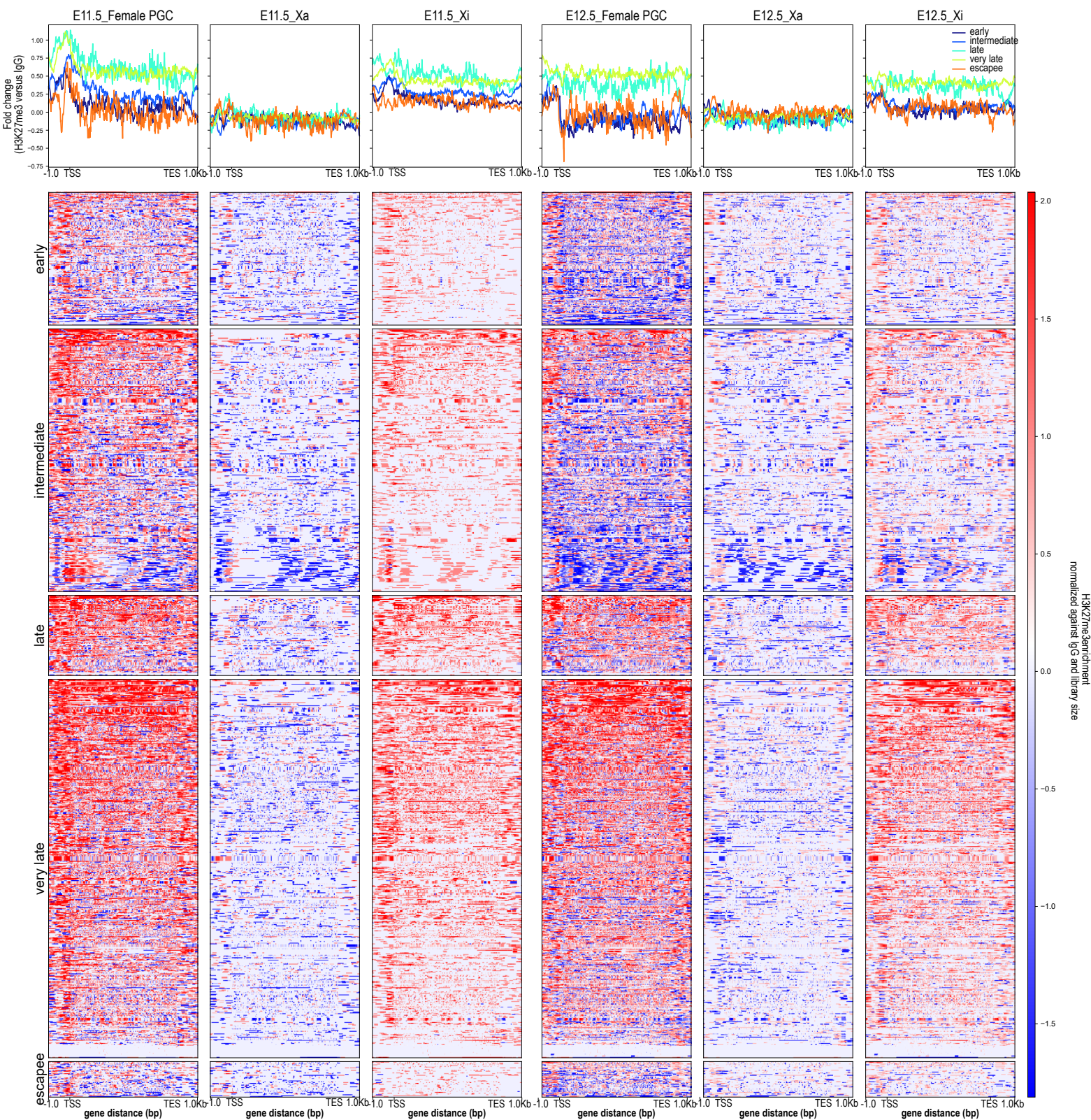
A**B****C****D****E**

A**B****C****D****E****F**

A**B****C**

A

bioRxiv preprint doi: <https://doi.org/10.1101/2023.04.25.532252>; this version posted April 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).



EXTENDED FIGURE 4