

Xist-dependent imprinted X inactivation and the early developmental consequences of its failure

Maud Borensztein, Laurène Syx, Katia Ancelin, Patricia Diabangouaya, Christel Picard, Tao Liu, Jun-Bin Liang, Ivaylo Vassilev, Rafael Galupa, Nicolas Servant, et al.

▶ To cite this version:

Maud Borensztein, Laurène Syx, Katia Ancelin, Patricia Diabangouaya, Christel Picard, et al.. Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nature Structural and Molecular Biology, 2017, 10.1038/nsmb.3365. hal-03021809

HAL Id: hal-03021809 https://hal.science/hal-03021809

Submitted on 19 Nov 2021

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.

- Fully Xist-dependent initiation of imprinted X inactivation and the early developmental
 consequences of its failure.
- 3

4	Maud Borensztein ¹ , Laurène Syx ^{1,2} , Katia Ancelin ¹ , Patricia Diabangouaya ¹ , Christel
5	Picard ¹ , Tao Liu ⁴ , Jun-Bin Liang ⁴ Ivaylo Vassilev ^{1,2} , Rafael Galupa ¹ , Nicolas Servant ² ,
6	Emmanuel Barillot ² , Azim Surani ³ , Chong-Jian Chen ⁴ and Edith Heard ¹ .
7	
8	¹ Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, 26 Rue d'Ulm,
9	75248 Paris Cedex 05, France
10	² Institut Curie, PSL Research University, Mines Paris Tech, Bioinformatics and
11	Computational Systems Biology of Cancer, INSERM U900, F-75005, Paris, France
12	³ Wellcome Trust Cancer Research UK Gurdon Institute, Department of Physiology,
13	Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge
14	CB2 1QN, United Kingdom

⁴ Annoroad Gene Technology Co., Ltd, Beijing, China

In mammals, differences in sex-chromosome constitution between males (XY) and 16 females (XX) have led to the evolution of dosage compensation strategies, including 17 transcriptional silencing of one X chromosome in females¹. In mice, X-chromosome 18 inactivation (XCI) first initiates in the pre-implantation embryo. The non-coding Xist 19 RNA is expressed only from the paternal allele leading to paternal X (Xp) inactivation². 20 The Xp remains inactive in extra-embryonic tissues, but is reactivated in the inner cell 21 mass followed by random XCI in the embryo proper^{3,4}. Although Xist is essential for 22 random XCI, the precise requirements for imprinted XCI during pre-implantation 23 development have been debated⁵. Here we report that the initiation of imprinted XCI 24 absolutely requires Xist using single-cell RNA-sequencing (scRNAseq) of early mouse 25 embryos. Lack of paternal Xist leads to genome-wide transcriptional misregulation in 26 the early blastocyst, with failure to activate the extra-embryonic pathway that is 27 28 essential for post-implantation development. We also demonstrate that the expression dynamics of X-linked genes depends both on strain and parent-of-origin, as well as on 29 30 location along the X chromosome. This study demonstrates that dosage compensation failure has an impact as early as the blastocyst stage and reveals the genetic and 31 epigenetic basis for orchestrating the transcriptional silencing of the X chromosome 32 during early embryogenesis. 33

34

In early mouse embryos, XCI has been shown to be dynamic and its requirements, both in *cis* at the level of the X-inactivation center (Xic) and in *trans*, have been debated⁶. Imprinted XCI has been proposed to initiate *de novo*^{2,9} following the onset of zygotic genome activation (ZGA) and *Xist* expression. One study proposed that Xp inactivation is initially *Xist*independent and that *Xist* may only be required for early maintenance of silencing⁵, while another reported a lack of Xp gene silencing in the absence of *Xist*⁷. These studies were all based on the analysis of just a few genes, however. A recent transcriptomic study in female pre-implantation embryos derived from inter-specific Cast/B6 (CB) crosses⁸ revealed that paternal XCI indeed initiates between the 4-8-cell stage. However, whether initiation of Xplinked gene silencing relies on *Xist* RNA, or is influenced by strain- or parent-of-origin (*eg* imprinted X-linked genes) were not explored.

To investigate the extent and requirements of gene silencing during imprinted XCI in 46 early embryogenesis, we profiled the expression kinetics of genes on the Xp and Xm 47 chromosomes, using scRNAseq. F1 embryos were derived from inter-specific crosses, of 48 either wild-type (wt) or Xist paternally deleted mutant (Xist^{pat\Delta}) origin, between the 2-cell and 49 50 blastocyst (approximately 60-64-cell) stages. Reciprocal crosses between highly polymorphic Mus musculus castaneus (Cast/EiJ) and Mus musculus domesticus (C57BL6/J) strains, herein 51 referred to as Cast and B6 respectively, were used (Figure 1a) and a minimum of 5 embryos, 52 53 and 6 single cells per stage for BC and CB wt embryos (Supplementary Table 1). Of 24,499 referenced mouse genes, 15,581 were found expressed in at least one developmental stage, 54 55 including 580 X-linked genes.

We first assessed the extent to which transcriptomes of single cells were associated by 56 stage, sex or cross, by performing principal component analyses (PCA) and hierarchical 57 clustering (Figure 1b and Supplementary Figure 1). The primary source of variability between 58 all cells was developmental stage, as expected based on previous studies⁸, thus validating the 59 quality of our data. Single cell transcriptomes clustered to a lesser extent by cross (BC and 60 CB), and then by sex (XX and XY) (Supplementary Figure 1), with the differences between 61 the sexes reaching a minimum by the 32-cell and blastocyst stages, presumably due to dosage 62 compensation. 63

To assess the precise timing of dosage compensation in male and female embryos, we examined autosomal and X-linked transcripts at each stage in both sexes. According to

Ohno's law⁹ average X-linked gene expression should be equivalent to the expression of 66 67 autosomal genes. Furthermore, equal expression of X-linked genes between females and males is expected through XCI. We compared X:Autosomes (X:A) expression ratios in single 68 blastomeres of each sex (Figure 1c). Expected X:A ratios would be 1 in females and 0.5 in 69 males in the absence of any dosage compensation (ie no X overexpression compared to 70 autosomes, and no XCI). We found that the X:A ratios were significantly above the expected 71 ratios as early as the 4-cell stage ($p < 9*10^{-4}$ for males and females after 4-cell stage, t-test) and 72 continued to rise until the 32-cell stage, suggesting that there is a progressive increase in 73 expression of the X compared to autosomes at the same time as, or soon after ZGA. In 74 females, the X:A ratio rose to 1.58, by the 32 cell stages and then significantly dropped to 75 1.37 by the early blastocyst stage $(p=1.96*10^{-2})$ between 32-cell and blastocyst, Kruskal-76 Wallis (KW) test), presumably due to XCI by this stage (see below). This suggests that X:A 77 78 ratios in female blastocysts progressively reach 1, although even at the early blastocyst stage, they were still slightly higher compared to males $(p=2.03*10^{-3}, KW)$, in agreement with 79 previously published data¹⁰. 80

We next investigated allele-specific X-linked gene expression and the timing of XCI 81 in BC and CB female embryos. At the 2-cell stage, ZGA and massive degradation of the 82 maternal pool of mRNAs occur. Here, transcripts are maternally biased genome-wide as 83 expected given the residual maternal pool (Figure 1d). At subsequent stages, while autosomal 84 transcripts reached parity for both parental genomes, with a parental ratio in blastocysts of 0.5 85 in both crosses, X-linked transcripts displayed maternal skewing even at the 16-cell stage. By 86 the blastocyst stage global transcription of the Xp was significantly reduced in both crosses 87 $(p < 2.2*10^{-16}, KW)$ indicating that XCI was fairly complete, as previously reported^{7,8,11}. We 88 compared the kinetics of Xp silencing for 13 X-linked genes previously analyzed by nascent 89 RNA-FISH¹¹ and found that most (12/13) genes showed very similar patterns (Figure 1e and 90

Supplementary Figure 2), giving us confidence that our scRNAseq data, bioinformatics pipeline and expression thresholds were valid. The one gene (out of the 13) for which slightly earlier Xp silencing was found by scRNAseq compared to previous reports was *Atrx*. We confirmed that *Atrx* is inactivated on the Xp in most cells by the morula stage using RNA FISH with a gene-specific probe (Supplementary Figure 3a). We also confirmed its previously reported Xp reactivation in the blastocyst¹¹ (Supplementary Figure 2).

We established an in vivo chromosome-wide map of X-linked gene activity between 97 the 4-cell and blastocyst stages. Of the 580 X-linked genes expressed in our scRNAseq, we 98 focused on the 164 (BC cross) and 134 (CB cross) most highly expressed and informative 99 genes (RPRT>4 and expressed in at least 25% of the cells of each stage and cross with a 100 minimum of 2), for which we could establish allelic expression profiles with confidence 101 (Supplementary Figure 3b and Figure 2 for the 125 common genes between BC and CB 102 103 crosses, see Online Methods for allelic expression threshold details). A striking switch from biallelic (grey, pale pink or pale blue) expression at the 4-cell stage, to monoallelic, maternal 104 105 (red) expression at the blastocyst stage can be observed for most X-linked genes. Several genes underwent only partial or no XCI (escapees) and will be discussed later. As expected, 106 Xist expression was exclusively of paternal origin throughout (Figure 2 and Supplementary 107 Figure 3b). Another gene showing only paternal expression was *Fthl17f*, part of the ferritin, 108 109 heavy polypeptide-like 17 family (also known as Gm5635), which has previously been reported to be exclusively paternally expressed and imprinted¹². By the blastocyst stage 110 Fthl17f expression was no longer detectable, presumably due to XCI (Figure 2 and 111 Supplementary Figure 3b). 112

We categorized genes into different groups with respect to their timing of XCI for each cross (*early* \leq *16-cell; intermediate* \leq *32-cell; late* = *blastocyst*; Figure 3a, SI Table 2 and Online Methods). Even by the 8-cell stage, XCI is complete for a few genes (*eg Rnf12*,

Pnma5, Figure 2 and Supplementary Figure 3b). By the blastocyst stage, Xp reached a very 116 similar state of inactivation in both BC and CB crosses (respectively 83.5% and 84.3% of the 117 164 and 134 X-linked informative and expressed genes are either silenced or maternally 118 biased at the blastocyst stage, Figures 2, 3a and 3b). However, when comparing gene 119 expression in embryos derived from BC and CB crosses (125 common genes), marked 120 differences were seen between crosses, with only 71.2% (89 of 125 genes) of X-linked genes 121 falling into the same or a similar category between BC and CB crosses (eg early/mid or 122 late/biased). The degree of consistency in silencing kinetics between crosses was evaluated if 123 no more than one developmental stage separated the same gene between BC and CB crosses 124 125 (Supplementary Table 2 and Online Methods for classification details). Furthermore, several genes show strain-specific escape, some of which have previously been described¹³ or 126 reported to escape in a tissue-specific fashion at later stages of development or in somatic 127 tissues $(eg \ Ddx3x, \ Idh3g)^{13-15}$. Several genes remain biallelically expressed, even at the 128 blastocyst stage and tend to show escape independent of strain (Figure 3b and Supplementary 129 Table 3). Many of these also show escape in somatic tissues¹⁶ (*eg Eif2s3x, Kdm5c, Utp14a*). 130

We then assessed whether gene-silencing kinetics was correlated with genomic 131 position along the X chromosome. We first focused on the 71.2% (n=89 genes) of genes with 132 correlated kinetics between crosses and the strain-specific genes (n=48). Although early and 133 intermediate silenced genes do tend to lie closer to the Xic compared to late silenced genes 134 (Figure 3c), gene silencing does not appear to occur as a simple linear gradient from the Xic 135 according to our allele-specific expression heatmap, with presence of escapees close to the 136 Xic (Figure 2). Rather, we noted several regions across the X chromosome, containing early-137 silenced genes, (eg Pnma5, Kif4, Magt1) from which silencing appeared to "spread" locally 138 (Figure 2). A recent study in ES cells showed that Xist RNA initially binds to specific 139 genomic regions (Xist "entry" sites) along the X chromosome, dependent on 3D proximity to 140

the *Xist* locus¹⁷. This binding has been hypothesized to silence genes locally and to then spread along the rest of the X chromosome by Engreitz *et al*¹⁷. We found that X-linked genes lying within (8 and 11 genes respectively in 32-cell and blastocysts), or close to (20 and 23 respectively in 32-cell and blastocysts) these regions showed the earliest silencing and strongest maternal imbalance (p=0.02 and p=0.03, KW, respectively in 32-cell and blastocysts, Figure 3d). Thus, we show that *Xist* RNA "entry" sites as defined in ESCs¹⁷ could correspond to XCI initiation sites *in vivo*, during imprinted XCI.

These findings suggest that Xist RNA plays a key early role in triggering gene 148 silencing during imprinted XCI. This contrasts with a previous report suggesting that 149 initiation of imprinted XCI could be is *Xist*-independent⁵. Indeed, although $Xist^{pat\Delta}$ females 150 die around E10.5, with major growth delay¹⁸, mutant and wt females appear morphologically 151 indistinguishable during pre-implantation development (data not shown). To evaluate whether 152 XCI is established in the absence of *Xist* expression, as previously reported⁵, we examined X-153 linked gene expression profiles in single cells of pre-implantation female embryos carrying a 154 paternal Xist deletion $(Xist^{pat\Delta})^{18,19}$. Xist is normally expressed exclusively from the Xp in pre-155 implantation embryos² (Figure 2). Transcriptomes of single blastomeres from hybrid F1 156 embryos (Cast females crossed with Xist^{matA} B6 males) were compared to CB wt embryos 157 between the 8-cell stage (when XCI normally initiates for some genes) to the blastocyst stage. 158 159 We found similar X:A expression ratios between female mutants and controls up to the 32cell stage (Figure 4a). However at the blastocyst stage, X:A ratios remained much higher in 160 mutants compared to wt embryos where this ratio normally decreases due to XCI ($p=1.77*10^{-10}$ 161 ⁴, KW). This demonstrates that Xp silencing is not initiated in $Xist^{pat\Delta}$ female blastocysts. 162 Bioinformatics analysis on the Xist-mutant single cell transcriptomes was used to produce an 163 allele-specific expression heatmap (see Online Methods) and, as expected, given the absence 164 of apparent dosage compensation in the mutants, we found that X-linked genes remained 165

significantly biallelically expressed in $Xist^{pat\Delta}$ embryos (Figure 4b). Only 1.6% genes (2 out of 122 genes: Rgn and Tktl1) showed maternal monoallelic expression in $Xist^{pat\Delta}$ mutant blastocysts, compared to 84.3% in CB controls. One of these, Tktl1, has been hypothesized to be imprinted²⁰. Moreover, Fthl17f, a well-known imprinted gene was aberrantly expressed in $Xist^{pat\Delta}$ blastocysts, suggesting a lack of Xp silencing.

We thus found no evidence for *Xist*-independent XCI (Supplementary Figure 5a), even 171 for the X-linked genes previously proposed to be silenced independently of $Xist^5$ (11 out of 14 172 assayed, only Rnf12, Abcb7 and Atrx were found to be Xist-dependent in the Kalantry et al 173 study). Three of the genes assayed by Kalantry et al (Abcb7, Fmr1 and Pgk1) showed a slight 174 maternal bias at the 16-cell or 32-cell stages in the $Xist^{pat\Delta}$ cells (left column, Supplementary 175 Figure 5a). However this is probably due to variability in their parental-origin expression, also 176 observed in CB controls (Abcb7 and Fmr1, Supplementary Figures 3 and 5a) or low parental-177 origin information (Pgk1, data not shown) rather than to Xp silencing. Instead, our data is in 178 agreement with the Namekawa et al study⁷ where *Xist*-dependent Xp silencing was proposed 179 to occur based on nascent RNA-FISH on 2-cell to blastocyst stage embryos, although their 180 study was only based on 8 genes, 4 of which were in common with ref 5. The discrepancies 181 between these previous studies were likely due to technical differences. The scRNAseq 182 analysis we provide here represents chromosome-wide evidence for Xist-dependent gene 183 silencing during pre-implantation embryogenesis. 184

The transcriptome of $Xist^{pat\Delta}$ embryos also provides us with a unique opportunity to explore the molecular defects that occur in the absence of paternal XCI. A genome-wide differentially expressed (DE) gene analysis was performed in *wt* and $Xist^{pat\Delta}$ embryos (Supplementary Table 4). Expression profiles of single blastomeres of controls and mutants were still found to cluster according to developmental stage by PCA (data not shown). However, at the 8-cell and 32-cell stages, a surprisingly elevated number of DE autosomal 191 genes (FDR<0.05) was found in $Xist^{pat\Delta}$ embryos compared to wt (Supplementary Figure 5b). 192 By the blastocyst stage, when paternal XCI is normally complete in wt females, 30% of the 193 total up-regulated genes in $Xist^{pat\Delta}$ embryos were found to be X-linked, corroborating an XCI 194 defect in the absence of Xist. DE genes included Tsix (the antisense transcript to Xist) which is 195 normally not expressed from the Xp at the 32-64 cell stage²¹ (Figures 4b and d and 196 Supplementary Figure 5b). The absence of Xist on the paternal X thus releases paternal Tsix197 repression in *cis* (without affecting the maternally imprinted Xist/Tsix alleles).

We explored the degree to which transcriptomes were perturbed in the Xist mutant 198 embryos using Ingenuity Pathway Analysis software. We found that many aberrantly down-199 regulated genes in $Xist^{pat\Delta}$ female blastocysts were associated with extra-embryonic tissue 200 pathways, embryonic growth and cell viability (Figure 4c and Supplementary Table 5). Key 201 extra-embryonic development genes that were aberrantly down-regulated included Tead4 202 $(\text{trophectoderm})^{22}$, Sox17²³ (primitive endoderm PrE and ExE), Bmp4²⁴ (trophectoderm TE 203 and PrE), $Arid3a^{25}$ (TE specification) and $Socs3^{26}$ (placental development) (Figure 4d). To 204 confirm the aberrant decrease of Sox17-positive cells in the PrE in $Xist^{pat\Delta}$ females, we 205 performed immunofluorescence on late blastocysts (Supplementary Figure 6a, c, e and g). In 206 $Xist^{pat\Delta}$ females, fewer cells express Sox17 compared to their male littermates and the 207 intensity of fluorescence of Sox17 is slightly decreased (Supplementary Figure 6g), which 208 corroborates the decrease in mRNA expression that we observe by scRNAseq. 209

Importantly, in addition to aberrant down regulation or repression of extra-embryonic genes, we also found abnormal overexpression of several pluripotency genes including *Prdm14, Esrrb* and *Tcl1* in *Xist*^{*pat* Δ} embryos. This suggested an inappropriate activation or lack of repression of such factors in the absence of XCI (Figure 4d). This is relevant to our recent findings showing that the presence of two active X chromosomes delays exit from pluripotency in ESCs, by preventing down-regulation of key genes, such as *Prdm14* or

Esrrb¹⁹. Moreover, aberrant over-expression of Prdm14, Esrrb and Tcl1 was observed in Xist 216 ^{/-} female ESCs induced to differentiate¹⁹. Intriguingly, the most significantly up-regulated 217 gene (10 log fold change) in $Xist^{pat\Delta}$ female blastocysts was the imprinted *Rhox5* gene, also 218 known as Pem-1. Rhox5 is a member of the reproductive X-linked Hox (Rhox) cluster, and is 219 expressed exclusively in the male germ line and in female (but not male) pre-implantation 220 embryos (Xp only)²⁷. Following implantation, its expression switches to the maternal allele 221 and becomes restricted to extra-embryonic tissues²⁷. The human RHOXF1 gene that is 222 hypothesized to be related to the murine $Rhox5^{28}$ shows similar sex-specific and lineage-223 specific expression in human pre-implantation embryos²⁹. Importantly, previous *in vitro* 224 studies demonstrated that over-expression of *Rhox5* can block differentiation of ESCs by 225 preventing exit from pluripotency^{30,31}. We validated *Rhox5* up-regulation at the protein level 226 using immunofluorescence and found that $Xist^{pat\Delta}$ female blastocysts show significantly 227 higher Rhox5 staining, particularly in the polar trophectoderm and inner cell mass region of 228 the embryo, compared to wt blastocysts (Supplementary Figure 6b, d, f and g). Quantification 229 230 of Rhox5 immunofluorescence showed a significant increase in Rhox5 protein levels (p=0.0171, Kolmogorov-Smirnov KS test, Supplementary Figure 6h) and in the number of 231 cells stained by Rhox5 antibody (Supplementary Figure 6f). This correlates well with our 232 scRNAseq data. 233

We conclude that even by the early blastocyst stage, the lack of initiation of Xp inactivation in Xist^{patA} embryos leads to inappropriate down-regulation of several key genes involved in extra-embryonic development, overexpression of several pluripotency genes and massive overexpression of Rhox5. Together some or all of these defects must ultimately result in compromised extra-embryonic development and redirection towards what appears to be a more pluripotent state, or at least a state from which further differentiation is perturbed. Importantly, aberrant induction of maternal *Xist* and Xm inactivation in extraembryonic tissues of blastocysts carrying a maternal *Tsix* deletion demonstrates that the presence of two active X chromosomes at the blastocyst stage can still be rescued in some females, and suggests that the major defect associated with a lack of paternal XCI is initially in the extra-embryonic lineage³².

In conclusion, we have demonstrated here the key role that Xist RNA plays in 245 initiating imprinted XCI. Our scRNAseq has enabled us to identify the molecular defects in 246 developmental pathways that emerge from this absence of dosage compensation and result in 247 lethality a few days later. We also reveal the influence of chromosomal location, as well as 248 genetic background and parent-of-origin on XCI kinetics. Our finding that Xist's predicted 249 250 initial binding sites on the X chromosome correspond to the earliest regions silenced, between the 8-16 cell stage, with evidence for local spreading in *cis* at the 32-blastocyst stage will 251 enable exploration of the local features that underlie the spread of silencing along the X 252 253 chromosome in an in vivo context. Our study demonstrates the critical requirement for accurate X-chromosome gene dosage during early embryo development and uncovers some 254 255 of the key pathways and factors that are affected in the absence of XCI. Future dissection of these pathways and their relationship to X-linked gene dosage should provide a better 256 understanding of the important role that small changes in RNA and protein levels can play, 257 258 not only development but also in disease.

260 **References**

- Lyon, M. F. Gene action in the X-chromosome of the mouse (Mus musculus L.).
 Nature 190, 372–3 (1961).
- 263 2. Okamoto, I. *et al.* Evidence for de novo imprinted X-chromosome inactivation
 264 independent of meiotic inactivation in mice. *Nature* 438, 369–373 (2005).
- 265 3. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of
- imprinted X inactivation during early mouse development. *Science* **303**, 644–9 (2004).
- Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos.
 Science 303, 666–9 (2004).
- 269 5. Kalantry, S., Purushothaman, S., Bowen, R. B., Starmer, J. & Magnuson, T. Evidence
- of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation.
 Nature 460, 647–651 (2009).
- Galupa, R. & Heard, E. X-chromosome inactivation: New insights into cis and trans
 regulation. *Curr. Opin. Genet. Dev.* **31**, 57–66 (2015).
- 274 7. Namekawa, S. H., Payer, B., Huynh, K. D., Jaenisch, R. & Lee, J. T. Two-step
- 275 imprinted X inactivation: repeat versus genic silencing in the mouse. *Mol. Cell. Biol.*276 **30**, 3187–205 (2010).
- 277 8. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-Cell RNA-Seq Reveals
- Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science (80-.).*343, 193–196 (2014).
- - Brockdorff, N. & Turner, B. M. Dosage compensation in Mammals. *Cold Spring Harb Perspect Biol* 7:a019406, (2015).
 - 10. Nguyen, D. K. & Disteche, C. M. Dosage compensation of the active X chromosome in
 mammals. *Nat. Genet.* 38, 47–53 (2006).
 - 11. Patrat, C. et al. Dynamic changes in paternal X-chromosome activity during imprinted

- 285 X-chromosome inactivation in mice. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5198–203
 286 (2009).
- 12. Kobayashi, S. *et al.* The X-linked imprinted gene family Fth117 shows predominantly
 female expression following the two-cell stage in mouse embryos. *Nucleic Acids Res.*38, 3672–3681 (2010).
- Calabrese, J. M. *et al.* Site-Specific Silencing of Regulatory Elements as a Mechanism
 of X Inactivation. *Cell* 151, 951–963 (2012).
- 292 14. Berletch, J. B. *et al.* Escape from X Inactivation Varies in Mouse Tissues. *PLOS Genet.*293 11, e1005079 (2015).
- Marks, H. *et al.* Dynamics of gene silencing during X inactivation using allele-specific
 RNA-seq. *Genome Biol.* 16, 149 (2015).
- 16. Balaton, B. P. & Brown, C. J. Escape Artists of the X Chromosome. *Trends Genet.* 32,
 348–359 (2016).
- Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture
 to spread across the X chromosome. *Science* 341, 1237973 (2013).
- 300 18. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient
- mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* 11,
 156–166 (1997).
- Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the
 pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* 14, 203–216
 (2014).
- Nesbitt, A. M. Genomic imprinting of the X-linked gene transketolase-like 1 in mouse
 and human. *ProQuest Diss. Theses A&I.* 884624661, (2010).
- Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the Xinactivation centre. *Nat. Genet.* 21, 400–4 (1999).

- 310 22. Nishioka, N. et al. Tead4 is required for specification of trophectoderm in pre-
- 311 implantation mouse embryos. *Mech. Dev.* **125**, 270–283 (2008).
- 312 23. Niakan, K. K. et al. Sox17 promotes differentiation in mouse embryonic stem cells by
- 313 directly regulating extraembryonic gene expression and indirectly antagonizing self-
- 314 renewal. *Genes Dev.* **24**, 312–326 (2010).
- Graham, S. J. L. *et al.* BMP signalling regulates the pre-implantation development of
 extra-embryonic cell lineages in the mouse embryo. *Nat. Commun.* 5, 5667 (2014).
- 317 25. Rhee, C. *et al.* Arid3a is essential to execution of the first cell fate decision via direct
- embryonic and extraembryonic transcriptional regulation. *Genes Dev.* 28, 2219–2232
 (2014).
- Takahashi, Y. *et al.* SOCS3: An essential regulator of LIF receptor signaling in
 trophoblast giant cell differentiation. *EMBO J.* 22, 372–384 (2003).
- 322 27. Kobayashi, S. *et al.* Comparison of gene expression in male and female mouse
- blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation
 stages. *Curr. Biol.* 16, 166–172 (2006).
- 28. Li, Q., O'Malley, M. E., Bartlett, D. L. & Guo, Z. S. Homeobox gene Rhox5 is
- regulated by epigenetic mechanisms in cancer and stem cells and promotes cancer
 growth. *Mol Cancer* 10, 63 (2011).
- 29. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome
 Dynamics in Human Preimplantation Embryos. *Cell* 165, 1012–1026 (2016).
- 330 30. Fan, Y., Melhem, M. F. & Chaillet, J. R. Forced expression of the homeobox-
- containing gene Pem blocks differentiation of embryonic stem cells. *Dev Biol* 210,
 481–96. (1999).
- 333 31. Cinelli, P. *et al.* Expression profiling in transgenic FVB/N embryonic stem cells
 334 overexpressing STAT3. *BMC Dev. Biol.* 8, 57 (2008).

- 335 32. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome
- 336 inactivation in mice by Tsix. *Development* **128**, 1275–1286 (2001).

338 **Supplementary Information** is linked to the online version of the paper at

339 www.nature.com/nature.

340

Acknowledgements We thank S. Bao and N. Grabole for experimental help in single 341 blastomere RNA sequencing and M. Guttman for sharing the Xist "entry" site coordinates. 342 We are grateful to P. Gestraud and V. Sibut respectively for the help in statistical and IPA 343 pathway analysis. We thank the pathogen-free barrier animal facility of the Institut Curie and 344 J. Iranzo for help with the animals and the Cell and Tissue Imaging Platform - PICT-345 IBiSA (member of France-Bioimaging) of the Genetics and Developmental Biology 346 Department (UMR3215/U934) of Institut Curie for help with light microscopy. We 347 acknowledge E. Schulz, E. Nora, I. Okamoto and the members of E.H. laboratory for help, 348 feedbacks and critical input. This work was funded by a fellowship of Région Ile-de-France 349 350 (DIM STEMPOLE) to M.B., the Paris Alliance of Cancer Research Institutes (PACRI-ANR) to LS and ERC Advanced Investigator award (ERC-2010-AdG – No. 250367), EU FP7 grants 351 SYBOSS (EU 7th Framework G.A. no. 242129) and MODHEP (EU 7th Framework G.A. no. 352 259743), La Ligue, Fondation de France, Labex DEEP (ANR-11-LBX-0044) part of the 353 IDEX Idex PSL (ANR-10-IDEX-0001-02 PSL) and ABS4NGS (ANR-11-BINF-0001) to E.H 354 and France Genomique National infrastructure (ANR-10-INBS-09) to EH, NS, EB. 355

356

357 Author Contributions

M.B., A.S. and E.H. conceived the study. M.B. performed most of the experiments. K.A. and
P.D., C.P., M.B. and R.G. performed respectively the IF and the RNA-FISH experiments.
T.L., J.B.L. and C.C.C. performed the single cell transcriptome library preparation and
sequencing. L.S., M.B., C.C.C., I.V. N.S. and E.B. defined the data processing and
bioinformatics analysis. L.S. built the computational pipeline for scRNAseq and analyzed the

363 data with M.B. M.B. and E.H. wrote the paper.

364

Author Information ScRNAseq data are deposited in Gene Expression Omnibus under accession number GSE80810. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for material should be addressed to E.H. (edith.heard@curie.fr).

Figure 1: Single cell RNA sequencing of early hybrid embryos and dosage compensation
 mechanisms.

(a) Schematic illustration of the single cell experiment and the harvested stages during pre implantation mouse development. Time windows showing the persistence of maternal mRNA
 pool, activation of zygotic gene expression and Xp inactivation are indicated.

(b) Principal component analysis (PCA) of single oocytes and pre-implantation blastomeres (2C to blastocysts) based on scRNA data. BC and CB Females and BC and CB males are respectively represented as \Box , \blacksquare and \blacktriangle , \bigstar . Different stages are designed by different colors. n= 6 to 30 cells per stage (details of each single cell is listed in Supplementary Table 1).

(c) Dosage compensation during preimplantation development. Level ratio of X chromosome
expression is measured over autosomal one (XX:AA or X:AA) between 2-cell stage to
blastocyst embryos, by Dunn's test (Kruskal-Wallis), p<0.001 to **.

(d) Allele-specific gene ratios from autosomes (plain red or yellow) and from X
chromosomes (dash red or yellow) in female single blastomeres (2-cell to blastocyst) from
BC and CB crosses. Allele-specific proportion represents the number of reads mapped on
paternal genome divided by the total number of paternal and maternal reads mapped for each
gene.

e) Example of expression dynamics of three X-linked genes with their classification as "early
inactivated", "intermediate inactivated" or "escapee" as described in Patrat *et al*, 2009¹¹ (see
also Supplementary Figure 2). Percentage of parental origin transcripts is represented between
oocytes and blastocyst.



b

Principal Component Analysis

Figure 1- Borensztein et al

395

a

Figure 2: Kinetics of silencing of X-linked genes over the entire X chromosome during imprinted XCI in different strains.

The mean allele-specific expression ratios per embryonic stage for each informative and 399 expressed X-linked gene in 4-cell to blastocyst stage female embryos are represented as 400 heatmaps, with strictly maternal expression (ratio ≤ 0.15) in red and strictly paternal 401 expression (ratio ≥ 0.85) in blue. Color gradients are used in between these two values as 402 shown in the key. Genes are ordered by genomic position (centromere top, telomere bottom). 403 Data from CB (left) and BC (right) female embryos are shown (for thresholds see Online 404 Method) and arrows are highlighting example of early silenced and escapee genes. n= 125 405 informative X-linked genes in common for CB and BC crosses. 406



Kinetics of gene silencing along the X chromosomes

Figure 2- Borensztein et al

Figure 3: Strain of origin and chromosomal positions are associated with different kinetics of silencing.

(a) X-linked genes are clustered based on their silencing kinetics as "early" (0.15 ≤ allelic ratio 411 at 16-cell or earlier), "intermediate" (Inter: 0.15≤allelic ratio at 32-cell), "late" (0.15≤allelic 412 ratio at blastocyst), "biased" (Bias: maternally biased in blastocyst, 0.15<allelic ratio ≤ 0.3) 413 and "escapee" (Esc., not silenced). The allelic ratio of each gene represents the number of 414 reads mapped on the paternal genome divided by the total number of paternal and maternal 415 reads mapped and is represented at 4-cell, 16-cell, 32-cell and blastocyst stages from single 416 female blastomeres. Further information is provided in Supplementary Table 2 and Online 417 Methods. n= 137 X-linked genes (89 with consistent silencing kinetics between BC and CB 418 crosses and 48 BC or CB-specific). 419

(b) Parental expression ratios of X-linked genes in female blastocysts in BC and CB strains. 420 421 Each dot represents a single gene. The upper section represents data from BC embryos; lower section is the data from CB embryos (reverse cross). Xist is represented by a red dot. Green 422 423 dots represent genes that escape from early XCI in both strains; orange dots represent genes that are strain-specific escapees. Some genes with biallelic ratios are represented by black 424 dots and correspond to genes that previously underwent Xp silencing, prior to the blastocyst 425 stage, but then became re-expressed. Further information on escapees is found in 426 Supplementary Table 3. n= 125 common X-linked genes. 427

(c) Box plot representing the distribution of the genomic distances to *Xist* locus (in Mb) for
the different clusters of genes. Transcription Start Site (TSS) of each gene has been used to
measure the distance, by Dunn's test (Kruskal-Wallis); p<0.05 corresponds to *.

(d) Allelic expression of X-linked genes classified by their relative position to *Xist* "early"
binding sites (as identified during XCI induction in ESCs¹⁷): inside (TSS located in a *Xist*"entry" site), next to (TSS located less than 100 kb to an "entry" site) and outside (over 100

- 434 kb of an "entry" site). Parental expression ratio has been plotted at 32cell-stage, when paternal
- 435 XCI is initiated for most of the genes, by Dunn's test (Kruskal-Wallis); p<0.05 corresponds to
- 436 *. Consistent and strain-specific genes have been used.





Figure 3- Borensztein et al

Figure 4: Paternal knockout of *Xist* impaired XCI, dosage compensation and differentiation pathways.

440 (a) Dosage compensation between X chromosomes and autosomes is measured by 441 X:Autosome expression ratio between 8-cell to blastocyst in CB females (left panel) and 442 $Xist^{pat\Delta}$ CB females (with a paternally inherited knock-out allele) (right panel).

(b) Heatmap representing allele-specific mean expression from 8-cell to blastocyst stage of Xlinked genes (from Figure 2) in *Xist^{patA}* mutant single cells. Strictly maternally expressed genes (allelic ratio ≤ 0.15) are represented in red and strictly paternally expressed genes (allelic ratio ≥ 0.85) in blue. Color gradients are used in between and genes have been ordered by genomic position. *Tsix* data have been included in the heatmap if expressed in at least 2 single cells per stage. = 122 genes.

449 (c) Major down-regulated genes and pathways detected between CB *wt* and CB $Xist^{pat\Delta}$ 450 females extracted from Supplementary Table 4, using QIAGEN's Ingenuity Pathway 451 Analysis (IPA) software (Supplementary Table 5). Pathways were visualized using Cytoscape 452 software. Arrows color code: red: leads to inhibition; blue: leads to activation; orange: 453 findings consistent with state of downstream molecule; grey: effect not predicted.

(d) Expression data of candidate genes from *wt* CB (black) and $Xist^{pat\Delta}$ CB (red) females, extracted from scRNAseq. Expression is represented in Reads Per Retro-Transcribed length per million mapped reads (RPRT) during early development (8-cell to blastocyst stages). *Gapdh* gene is a control housekeeping gene. n= 4 to 30 cells per stage and genotype. By Kruskal-Wallis test; p<0.05 corresponds to *.



Figure 4- Borensztein et al

461 Supplementary Figure Legends

462

463 Supplementary Figure 1: Hierarchical clustering of scRNAseq

Hierarchical clustering of single cell transcriptomes, based on Spearman's correlation. Cells
were clustered by developmental stage, then by cross (BC or CB), and finally according to
their sex. ScRNAseq samples called 64C are related to the early blastocyst stage. n=184
single cell samples

468



Hierarchical clustering of the single cell transcriptomes

471 Supplementary Figure 2: Single cell RNA sequencing data corroborate RNA-FISH 472 based silencing of kinetics.

473 Candidate X-linked gene silencing kinetics has been compared to previously known kinetics 474 studied by RNA-FISH in Patrat *et al*¹¹. Maternal and paternal reads are respectively 475 represented as a red or a blue line. Percentage of parental origin transcript (top panel) and 476 level of expression in RPRT (bottom panel) of each gene are represented between oocytes and 477 blastocysts.

478



481 Supplementary Figure 3: Gene validation by RNA-FISH and chromosome-wide 482 representation of parental expression ratio.

(a) RNA FISH using probes recognizing Xist (signal in green) and Atrx (signal in red) 483 nascent RNAs on a 16C female embryo. DAPI is in blue. Right pictures are enhanced pictures 484 of two individual blastomeres. Percentage of nuclei showing pinpoints of nascent transcripts 485 by RNA FISH from Xp and Xm has been assessed and summarized as the median under the 486 picture. Normalization of the primary transcript detection frequency obtained for the paternal 487 (Xist RNA-associated) allele in female embryos has been done thanks to that obtained for the 488 maternal allele in male embryos at the same stage. Number of embryos and single cell 489 processed are indicated under each genotype. Scale bar represents 10µm. 490

(b) Heatmaps are representing the mean of allele-specific expression of X-linked genes between Oocytes to blastocyst in CB and BC crosses. Strictly maternally expressed genes (allelic ratio ≤ 0.15) are represented in red and strictly paternally expressed genes (allelic ratio ≥ 0.85) in blue. Color gradients are used in between and genes have been ordered by genomic position. Oocytes and 2C data have been included to Figure 2 heatmaps as well as strainspecific genes.

497



b Kinetics of XCI in *wt* females



Supplementary Figure 3- Borensztein et al

500 Supplementary Figure 4: Complete single cell information of X-linked gene expression

501 in pre-implantation development

Heatmap representing the allele-specific expression of informative and well expressed Xlinked genes in each single cell, from oocytes to blastocysts, in male and female embryos derived from BC and CB crosses. Strictly maternally expressed genes (allelic ratio ≤ 0.15) are represented in red and strictly paternally expressed genes (allelic ratio ≥ 0.85) in blue. Color gradients are used in between and genes have been ordered by genomic position. Genes expressed in one or both crosses have been included to the single cell heatmap. n=173 genes.

508



511 Supplementary Figure 5: Absence of X-chromosome inactivation in absence of *Xist*.

(a) Each plot represents a candidate X-linked gene that has been previously studied in
Kalantry *et al*, 2009⁵. Allele-specific expression ratio of each gene represents the number of
reads mapped on paternal genome divided by the total number of paternal and maternal reads.
Maternal and paternal reads are respectively represented as a red or a blue line. Percentage of
parental origin transcript (top panel) and level of expression in RPRT (bottom panel) of each
gene are represented between 8-cell stage and blastocysts.

(b) Table is summarizing the total differentially expressed (DE) genes between CB *wt* and CB *Xist*^{$pat\Delta$} female embryos during early development and their localization in X chromosomes or autosomes. The percentages shown correspond to the repartition of total DE genes between autosomes and X chromosomes. The blastocyst stage has been highlighted in red and candidate DE genes from this stage have been analyzed in Figure 4.

523



b

Total differentially expressed genes in Xist^{pat}CB females

Developmental	DE genes from the X chromosome				DE genes from the autosomes			
Developmental	Total Upregulated		Total Downregulated		Total Upregulated		Total Downregulated	
stages	%	n	%	n	%	n	%	n
8-cell	5.0%	27	1.9%	14	95.0%	514	98.1%	541
16-cell	11.6%	15	1.0%	1	88.4%	114	99.0%	129
32-cell	4.7%	29	2.3%	12	95.3%	591	97.7%	620
Blastocyst	29.6%	16	3.6%	2	70.4	38	96.4%	54

Supplementary Figure 5 - Borensztein et al

527 Supplementary Figure 6: Abnormal Sox17 and Rhox5 pattern in Xist^{pat∆} female
528 blastocysts.

Maximum intensity projection of wt and $Xist^{pat\Delta}$ E4.25 blastocysts analyzed by 529 immunofluorescence against Sox17 (a, c) or Rhox 5 (b, d). Staining for Sox17 or Rhox5 is in 530 red, DAPI is in grey. Scale bar represents 20µm. Percentage of positive cells have been 531 assessed and summarized as the mean+sem for Sox17 (e) and Rhox5 (f). Numbers of embryos 532 are indicated under each genotype. By Kruskal-Wallis test; p<0.05 and p<0.001 correspond 533 respectively to * and **. Quantification of immunofluorescence has been done by measuring 534 the corrected total cell fluorescence using ImageJ software (Fiji, NIH) for Sox17 (g) and 535 Rhox5 (h) and tested by Kolmogorov-Smirnov test. 536











Supplementary Figure 6- Borensztein et al

539	Supple	emental	Information	
-----	--------	---------	-------------	--

541 Supplementary Table 1: Summary of single cell RNAseq samples.

542 For each library is provided: single cell's name, stage, embryo number, gender, cross and the

- raw read number, filtered ones and percentage of mapping.
- 544

545 Supplementary Table 2: Silencing gene classes

546 Silencing classes for the 178 informative and well-expressed X-linked genes in BC and/or CB

547 crosses.

548

549 Supplementary Table 3: Summary of escapees and their status in other studies.

550 Information for escapees is provided for CB and BC crosses. Status of each escapee gene has

been evaluated in other studies using hybrid cell lines or tissues $^{13-15,33}$.

552

553 Supplementary Table 4: Misregulated genes in absence of paternal Xist

Table summarizing all the genes significantly different between CB wt and CB $Xist^{pat\Delta}$ female

embryos during early development (8-cell to blastocyst).

556

557 Supplementary Table 5: Pathways and candidate genes down-regulated in $Xist^{pat\Delta}$ female

558 blastocysts and highlighted by IPA analysis

559 Summary of all the pathway and candidate genes and relationships, which have emerged from

560 IPA analysis between CB *wt* and CB $Xist^{pat\Delta}$ female blastocysts.

562 Methods

563

564 Mouse crosses and collection of embryos

All experimental designs and procedures were in agreement with the guidelines from Frenchlegislation and institutional policies.

567 All BC and CB embryos were respectively derived from natural meetings between C57BL/6J (B6) females crossed with CAST/EiJ (Cast) males or by the reciprocal cross. The Xist^{pat\Delta} 568 mutant embryos (Xist^{+/-}) have been obtained by mating between Cast females and Xist^{-/Y} 569 males (mixed background: B6D2F1: C57BL/6J and DBA/2J, 129S1/SvlmJ and BALB/cJ). 570 Embryos have been harvested at 2-cell, 4-cell, 8-cell, 16-cell, 32-cell and blastocyst 571 (approximately 60 to 64-cell) stages, respectively at E1.5, E2.0, E2.25, E2.75, E3.25 and 572 E3.5. B6 and Cast pure oocytes have been collected at E0.5 after matings of females with 573 vasectomized males (Figure 1a). The collected embryos were only included in the analysis if 574 they showed a normal morphology and the right number of blastomeres in relation with their 575 developmental stage. 576

577

578 RNA Fluorescent In Situ Hybridization

RNA FISH on preimplantation embryos was performed as previously described³ using the
intron-spanning Fosmid probe WI1-2039P10 (BacPac Consortium at Children's Hospital
Oakland Research Institute) for *Atrx* and the intron-spanning plasmid probe p510 for *Xist*.
Images were acquired using a wide-field Deltavision core microscope (Applied Precision –
GE Healthcase) with a 60× objective (1,42 oil PL APO N) and 0.2 µm Z-sections. Images
were analyzed using ImageJ software (Fiji, NIH).

585

587 **Immunofluorescence staining**

Immunofluorescence was essentially carried out as described previously³⁴ with an additional 588 step of blocking in 3% FCS before the primary antibody incubation. Immunofluoroscence of 589 embryos either from mutant or control male progeny have been always performed in parallel 590 and in suspension. The following antibody were used: anti-goat Pem1 (Rhox5)/Santacruz sc-591 592 21650/1:50 and anti-goat Sox17/R&D Systems AF1924/1:100. Images were acquire using Inverted laser scanning confocal microscope with spectral detection (LSM700 - Zeiss) 593 equipped with a 260nm laser (RappOpto), with a 60X objective and 2 µm Z-sections. 594 Maximum projections and total corrected fluorescence measurements (=integrated density -595 (area of selected cell x mean fluorescence of background readings) were performed with 596 Image J software (Fiji, NIH) using previously described methodology³⁵. The total corrected 597 cellular fluorescence (TCCF) = integrated density – (area of selected cell × mean fluorescence 598 of background readings), was calculated. 599

600

601 Single cell dissociation from pre-implantation mouse embryos

Oocytes and embryos were collected by flushing oviducts (E0.5 to E2.75) or uterus (E3.25 602 and E3.5) with M2 medium (Sigma). The zona pellucida was removed using acid Tyrode's 603 solution (Sigma), and embryos were washed twice with M2 medium (Sigma). To isolate 604 individual cells, we then incubated embryos in Ca^{2+} , Mg^{2+} free M2 medium for 5 to 20 605 minutes, depending on the embryonic stage. For the blastocyst stage, Ca²⁺, Mg²⁺ free M2 free 606 medium was replaced by a 5-minute incubation in TrypLE (Invitrogen). After incubation, 607 each blastomere was mechanically dissociated by mouth pipetting with a thin glass capillary. 608 Single cells were then washed 3 times in PBS/acetylated BSA (Sigma) before being manually 609 picked into PCR tubes with a minimum amount of liquid. We either directly prepared the 610 cDNA amplification or kept the single cells at -80°C for future preparation. 611

612 Single cell RNA amplification:

PolyA⁺ mRNA extracted from each single cell was reverse transcribed from the 3'UTR and amplified following the *Tang et al* protocol³⁶. Care was taken to process only embryos and single blastomeres of the highest quality based on morphology, number of cells and on amplification yield. A total of 72 BC and 110 CB (or 113 *wt* and 69 *Xist^{patA}* mutant) blastomeres have been processed and passed quality controls.

618

619 Quality and sex determination

After cDNA amplification and before size selection and library preparation, quality of cDNAs 620 from each of the 182 samples were validated by studying expression level of three 621 housekeeping genes: Gapdh, Beta-Actin and Hprt. Primers used for real-time PCR were as 622 Gapdh F: ccccaacactgagcatctcc; Gapdh R: follows: attatgggggtctgggatgg; ActB F: 623 624 aagtgacgttgacatccg; ActB R: gatccacatctgctggaagg; Hprt F: cctgtggccatctgcctagt; Hprt R: gggacgcagcaactgacatt. Care was taken to process only single cells with consistent 625 amplification rate of the three housekeeping genes in the same developmental stage. 626

The sex of each embryo was accessed by expression level analysis of *Xist* (female-specific transcript) and *Eif2s3y* (male-specific transcript) by real-time PCR. Primers used were: Eif2s3y_F: aattgccaggttattttcattttc Eif2s3y_R: agttcagtggtgcacagcaa; Xist_F: ggttctctctcccagaagctaggaa and Xist_R: tggtagatggcattgtgtattatatgg.

631

632 Single cell libraries and deep-sequencing

Single-cell libraries were prepared from the 182 samples according to Illumina manufacturer
protocol and deeply sequenced on an Illumina HiSeq 2500 instruments in single-end 50bp
reads (Supplementary Table 1).

637 Quality control and filtering of raw data

Quality control was applied on raw data as previously described in (Ancelin et al, 2016)³⁴.
Sequencing reads characterized by at least one of the following criteria were discarded from
the analysis:

- 1. More than 50% of low quality bases (Phred score <5).
- 642 2. More than 5% of N bases.
- 643 3. More than 80% of AT rate.
- 4. At least 30% (15 bases) of continuous A and/or T.
- 645

646 SNP calling and allele-specific origin of the transcripts

647 <u>SNPs collection and strain-specific genome construction</u>

The VCF file (mgp.v5.merged.snps_all.dbSNP142.vcf) reporting all SNP sites from 36 mouse strains based on mm10 was downloaded from the Sanger database. Using SNPsplit tool (v0.3.0)³⁷, these SNPs were filtered based on their quality values (FI value) and used to reconstruct the Cast genome from mm10 genome assembly.

652

653 Allele-specific alignments of RNAseq reads

To study the allele-specific gene expression, reads were processed using a pipeline adapted from Gendrel *et al*, 2014^{33} . Single-end reads were first aligned to the mouse mm10 and CAST genomes using the TopHat2 software (v2.1.0)³⁴. Only random best alignments with less than 2 mismatches were reported for downstream analyses. The resulting mapping files for both parental genomes were then merged for each sample, using these following rules:

- 6591. If a read mapped at the same genomic position on the two genomes with the same660 number of mismatches, this read will be considered as a common read.
- 661 2. If a read is aligned with less mismatches on one genome, the best alignment will be

retained and this read will be considered as a specific read for the correspondingstrain.

- 664 3. If a read is aligned with the same number of mismatches on both genomes but at665 different genomic positions, this read will be discarded.
- 666

667 <u>Allelic imbalance in gene expression and gene classification</u>

SNPs between *C57BL/6J* (B6) and *CAST/EiJ* (Cast) were extracted from the VCF file used to
reconstruct the Cast genome. After removing common exonic SNPs between *Xist* and *Tsix*genes, 20,220,776 SNPs were retained.

The SAMtools mpileup utility $(v1.1)^{38}$ was then used to extract base-pair information at each 671 genomic position from the merged alignment file. At each SNP position, the number of 672 paternal and maternal allele was counted. Threshold used to call a gene informative was 5 673 reads mapped per single SNP with a minimum of 8 reads mapped on SNPs per gene to 674 minimize disparity with low polymorphic gene. The allele-specific origin of the transcripts (or 675 allelic ratio) has been measured by the total number of reads mapped on the paternal genome 676 divided by the total number of paternal and maternal reads for each gene: allelic ratio = 677 paternal reads / (paternal+maternal) reads. 678

679 Genes are thus classified into two categories:

680 1. Monoallelically expressed genes: allelic ratio value ≤ 0.15 or ≥ 0.85 .

- 681 2. Biallelically expressed genes: allelic ratio value > 0.15 or < 0.85.
- 682

683 *Estimation of gene expression levels*

Given that our RNA reverse transcription only allows sequencing on average 3 kilobases from
the 3'UTR, half of the expressed genes are partially covered (less than 50% of the gene size in
average). To estimate the transcript abundance, read counts are thus normalized based on the

amplification size of each transcript (RPRT for Reads Per Retro-Transcribed length per
million mapped reads) rather than the size of each gene (RPKM).

689

690 *Filtering of biased SNPs*

As we observed a bias for some polymorphisms in oocytes (maternal reads only) and male cells (maternal X chromosome only), oocytes (autosomes and X-chromosomes) and males (X-chromosome) were used to fix the issue. Therefore, SNPs covered by at least 5 reads and having an allelic ratio greater than 0.3 (biallelic or paternally expressed) in at least 2 of these samples were discarded. In total, 275 SNPs were filtered out, including 40 sites located on the X-chromosome.

Generation of *Xist^{patA}* mutant embryos involved the use of a *Xist^{A/Y}* stud of mixed background (B6D2F1: *C57BL/6J* and DBA/2J, 129S1/SvlmJ and BALB/cJ). We have then applied another SNP filtration to the KO samples to remove all B6 polymorphisms that could have been lost on the X chromosome by the mix background of the *Xist^{A/Y}* stud. To do so, all existing SNPs between B6 and DBA/2J, 129S1/SvlmJ and BALB/cJ, on the X chromosome, have been removed from our SNP database (34,397 SNPs, which represent 5,5% of X chromosome SNPs between B6 and Cast).

704

705 **Principal component analysis, hierarchical clustering and differentially expressed genes**

Count tables of gene expression were generated using the RefFlat annotation and the HTSeq software³⁹ (v0.6.1). Only genes with a RPRT (Reads Per Retro-Transcribed length per million mapped reads) value >1 in at least 25% of the single cells of at least one developmental stage (with a minimum of 2 cells) were kept for the downstream analysis. The TMM method from the edgeR R-package (v3.14.0)³⁷ was first used to normalize the raw counts data. Principal component analysis (PCA) and hierarchical clustering have then been used to determine how

single cells were clustered to the others though their gene expression profiles, depending of 712 713 their stage, sex and cross. PCA on normalized data was performed using FactoMineR Rpackage (v1.33). Hierarchical clustering analysis was based on Spearman correlation distance 714 and the Ward method, using the hclust function implemented in the gplots R-package 715 (v3.0.1). Limma R-package $(v3.28.4)^{38}$ was applied to identify the differentially expressed 716 genes between 8-cell stage and blastocyst in control and $Xist^{pat\Delta}$ mutant females. Using the 717 Benjamini-Hochberg correction, genes with an adjusted p-value lower than α =0.05 were 718 called as differentially expressed. 719

720

721 Functional enrichment analysis

Down-regulated genes in $Xist^{pat\Delta}$ mutant female blastocysts compared to CB female blastocysts were analyzed using QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, <u>www.qiagen.com/ingenuity</u>). The Functions and Diseases module has been used to extract the most significantly deregulated pathways and their associated genes)⁴⁰.

726

727 Heatmap generation for X-chromosome allelic gene expression

For BC and CB heatmaps, data from informative genes were analyzed at each developmental 728 stage only if the gene was expressed (RPRT>4) in at least 25% of single blastomeres (with a 729 minimum of 2 cells) at this particular stage and cross (Figures 2 and 4b and Supplemental 730 Figure 3). To follow the kinetics of expression, we decided to focus only on genes expressed 731 in at least 3 different stages between the 4-cell to blastocyst stages. Mean of the allelic ratio of 732 each gene is represented for the different stages. The same gene candidate list was used to 733 produce the $Xist^{pat\Delta}$ heatmaps (Figure 4b). A value was given only if the gene reached the 734 threshold of RPRT >4 in at least 25 % of single cells (with a minimum of 2 cells) per stage 735 and cross. 736

738 Definition of X-linked gene silencing/escape classes

We have automatically assigned X-linked genes that become strictly maternal (allelic ratio 739 ≤ 0.15) at the 16-cell stage or before to the early silenced gene class; those that become 740 maternal at the 32-cell stage to the intermediate silenced class (allelic ratio equals NA or 741 >0.15 at 16C and ≤ 0.15 at 32C) and those that are silenced only by the blastocyst stage, to the 742 late silenced gene class (allelic ratio equals NA or >0.15 at 16C and 32C and ≤0.15 at 743 blastocyst stage). At the blastocyst stage, X-linked genes showing a maternal bias of 744 expression (0.15<allelic ratio≤0.3) are categorized as maternally biased. A final group 745 746 concerns genes that escape imprinted Xp inactivation (allelic ratio >0.3 at blastocyst stage) (Figure 3a). Genes escaping XCI were separated into two classes: constitutive escapees if they 747 were classified as escapees in both CB and BC stages and as facultative escapees if they were 748 749 strain-specific escapees (Figure 3b and Supplementary Table 3).

Existence of a consistency in silencing kinetics between crosses was evaluated if no more than one developmental stage separated the same gene between BC and CB crosses. If the consistent genes belonged to two different classes, a class for all (BC+CB) has been attributed thanks to their parental ratio mean of (BC mean + CB mean) in Figure 3a and 3d.

754

755 Dosage compensation, X:A expression ratio

We measured the global X:A expression ratio in females (XX :AA ratio) and males (X :AA ratio) as the level of expression of X-linked genes divided by the global level of expression of the autosomal genes. Only genes with an expression value RPRT >4 were used for subsequent analysis (Figures 1d and 4a). Adjustment of the number of expressed genes between X and autosomes has been published to be critical for X:A expression ratio measurement⁴¹. We then added a bootstrapping step and randomly selected, for each sample, an autosomal gene set with the same number of expressed genes compared to the X to estimate the global X:A ratio.

- This step was repeated 1000 times and the X:A expression ratio was estimated as the median
- of the 1000 values.
- 765
- 766 Data access
- The Gene Expression Omnibus (GEO) accession number for the data sets reported in this
- 768 paper is GSE80810.

Supplemental References : 769

770

- 33. Gendrel, A. V. et al. Developmental dynamics and disease potential of random 771 monoallelic gene expression. Dev. Cell 28, 366-380 (2014). 772
- 34. Ancelin, K. et al. Maternal LSD1 / KDM1A is an essential regulator of chromatin and 773
- transcription landscapes during zygotic genome activation. *Elife* pii: e0885, (2016). 774
- 35. McCloy, R. A. et al. Partial inhibition of Cdk1 in G2 phase overrides the SAC and 775 decouples mitotic events. Cell Cycle 13, 1400–1412 (2014). 776
- 36. Tang, F. et al. RNA-Seq analysis to capture the transcriptome landscape of a single 777 cell. Nat. Protoc. 5, 516–35 (2010). 778
- 37. Rozowsky, J. et al. AlleleSeq: analysis of allele-specific expression and binding in a 779 network framework. Mol. Syst. Biol. 7, 522 (2011). 780
- 781 38. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009). 782
- 783 39. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-
- throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). 784
- 40. Shannon, P. et al. Cytoscape: a software environment for integrated models of 785

biomolecular interaction networks. Genome Res. 2498-2504 (2003). 786

- 41. Kharchenko, P. V, Xi, R. & Park, P. J. Evidence for dosage compensation between the 787
- X chromosome and autosomes in mammals. Nat. Genet. 43, 1167–1169 (2011). 788